
aqmm Documentation

Release latest

Jul 05, 2020

Contents

1	Document for AQMM (Absolute quantification of metagenome and metatranscriptome) .	1
1.1	clone the source code to local server	1
1.2	prepare the meta_data.txt which contains all the molecular experimental data	1
1.3	Running the aqmm just by one command	2
1.4	Example to run	2
1.5	Compare with RQ methods	3

CHAPTER 1

Document for AQMM (Absolute quantification of metagenome and metatranscriptome) .

The AQMM was developed to perform absolute quantification of multiple metagenome and its parallel metatranscriptome. In order to use this algorithm, the experiment should be designed with both metagenome and metatranscriptome data. From initial stage, the molecular experimental data should be recorded to help on estimating the overall DNA or RNA of a unit (ml/gram) of sample. The AQMM was demonstrated to obtain better results of differential expression genes identification in comparative metatranscriptomic studies.

1.1 clone the source code to local server .

git clone <https://github.com/biofuture/aqmm.git> .

1.2 prepare the meta_data.txt which contains all the molecular experimental data .

The meta_data.txt should .

SID	SNAME	DNA_weight	DNA_volume	DNA_extr_eff	RNA_weight	RNA_volume	RNA_extr_eff	RNA_ratio
1	AS1	24570	0.5	0.282	6135	0.5	0.075	0.03
2	AS2	27300	0.5	0.282	6355	0.5	0.075	0.03

Each line contains information of one sample:

```
SNAME is the name of the sample, which stored under the input directory .
DNA_weight is the weight of DNA for the extraction (ng) .
DNA_volume is the total volume of sample used for the DNA extraction .
DNA_extr_eff is the extracting efficiency for the sample, which is an empirical value,
→ for a DNA extraction KIT for a sample .
```

(continues on next page)

(continued from previous page)

```

RNA_weight is the weight of RNA for the extraction .
RNA_volume is the total volume of sample used for the RNA extraction .
RNA_extr_eff is the extracting efficiency for the sample, which is an emperical value_
↪for a RNA extraction KIT for a type of sample .
mRNA ratio is the ratio of mRNA to total RNA, which is a emperical value for_
↪different environment

```

Notice: Before using this algorithm, users need to get the above meta data information as accurately as possible, although the algorithm show robustly to the variance to the extraction efficiency among different samples, users should keep all the condition as consistent as possible in order to get more accurate/reliable quantification and comparativety.

1.3 Running the aqmm just by one command .

By typing ./aqmm -h the help information will show here .

```

perl ./aqmm -a <Input_metagenome_dir> -b <Input_metatranscriptome_dir> -m <Experimental_meta_data> -n <num_threads> -o <ouput_dir> Author Jiang Xiao-Tao 2017-08 Email biofuture.jiang@gmail.com

```

where:

- a <Input_metagenome_dir> including all the fq files of your metagenomic samples after quality filtering
- b <Input_metatranscriptome_dir> including all the paralell metatranscriptomic fq files with metagenomic samples after . remvoing the rRNA with tools like SortMeRNA.
- m meta_data.txt including all the experimental extraction information and the extraction efficiency information of your sample type
- o <ouput_dir> will include all the output results for the quantification, including the esimated cell numbers of metagenomic data of each sample and the estimated sequenced cell numbers your metatranscriptomic data included. With the quantification results obtained, the abundance of gene/species could be futher normalized to per cell/volume/gram etc. al.

1.4 Example to run

There is a small demo to show how to use AQMM to generate absolute quantification of your data under the example directory.

Enter the example directory and run the following command

```

../dist/aqmm -a AS_metagenome -b AS_metatranscriptome -m meta_data.txt -n 12 -o testout

```

After runnning, check the results

```

cat dna_cell_info.txt .

```

SID	SNAME	li-brary_size	lib_cell	cell_per_ml	A_ratio	T_ratio	C_ratio	G_ratio
1	AS2	500000000	8.567425133143	3.2327522e+19	0.22371768	0.22220922	0.27649788	0.27757
2	AS1	500000000	9.93418747046	3.3731923e+19	0.21272236	0.21128754	0.28672574	0.28926

```

cat rna_cell_info.txt .

```

SID	SNAME	library_size	lib_cell	cell_per_ml	A_ratio	T_ratio	C_ratio	G_ratio
1	AS2	50000000	8.5674251	3.23275242e+19	0.22371	0.22220	0.276497	0.27757
2	AS1	50000000	9.9341874	3.37319232e+19	0.21272	0.21128	0.286725	0.28926

These numbers could be used to normalize your genes/species to per cell/volume level in the end!

1.5 Compare with RQ methods

A script was developed to process the results for both AQMM and RQ methods like RPKM/TPM/edgeR to identify the differential gene expression between groups:

```
perl compare_aqmm_rq.pl
Author: Xiao-Tao JIANG
Date:
Modified : 19-01-2019
Email: biofuture.jiang@gmail.com
perl compare_aqmm_rq.pl -a <meta_data_rna.txt> -b <unique_gene_set> -c <rna_reads_
↳count.matrix> -o <Oprefix> -h
-a input meta data file for all the metatranscriptomic data <required>
-b the gene set fasta file <required>
-c matrix of reads number mapped in RNA samples for each gene <required>
-d matrix of transcript per copy gene run AQMM required.
-o output prefix, all output result files name are prefix with this string <required>
-l the length of reads default strategies PE 150 sequencing, default 150 bps
-s this option is to select the normalization methods, RPKM, TPM, EDGER, AQMM, ALL.
↳default AQMM \
-h print this help information
```