Allometric scaling of RNA abundance from genes to communities

Mark Louie D. Lopez, Ya-ying Lin, Stephan Q. Schneider, and Ryuji J. Machida

Biodiversity Program, Taiwan International Graduate Program, Academia Sinica, Nankang District, Taipei 11529, Taiwan, Republic of China.

Department of Life Science, National Taiwan Normal University, Wenshan District, Taipei 11677, Taiwan, Republic of China.

Institute of Organismic and Cellular Biology, Academia Sinica, Nankang District, Taipei 11529, Taiwan, Republic of China.

Biodiversity Research Center, Academia Sinica, Nankang District, Taipei 11529, Taiwan, Republic of China

Abstract

The metabolic theory of ecology (MTE) and growth rate hypothesis (GRH) help explain the mechanistic basis of size (allometry) and temperature dependence on growth rate and whole-body-RNA content in organisms. However, testing the RNA scaling with next-generation sequencing is yet to be done. Here, we validated the assumptions of GRH and MTE on messenger RNA and ribosome abundance using mock community metatranscriptome analysis. Our findings highlight that fast-growing smaller species harbor greater RNA abundance per mass of tissue compared with species having larger body sizes and slower growth rates, where allometric slopes for genomic and gene-level RNA abundance range from –⅓ to –1. We found that genome size and body size impose significant constraints in interspecific RNA abundance scaling, while the assumed temperature dependence appeared to be weak. Lastly, allometric scaling integration in community-level models may extend the use of metatranscriptomics as a reliable tool for estimating ecosystem processes.

Codes used for the Bioinformatics Processing

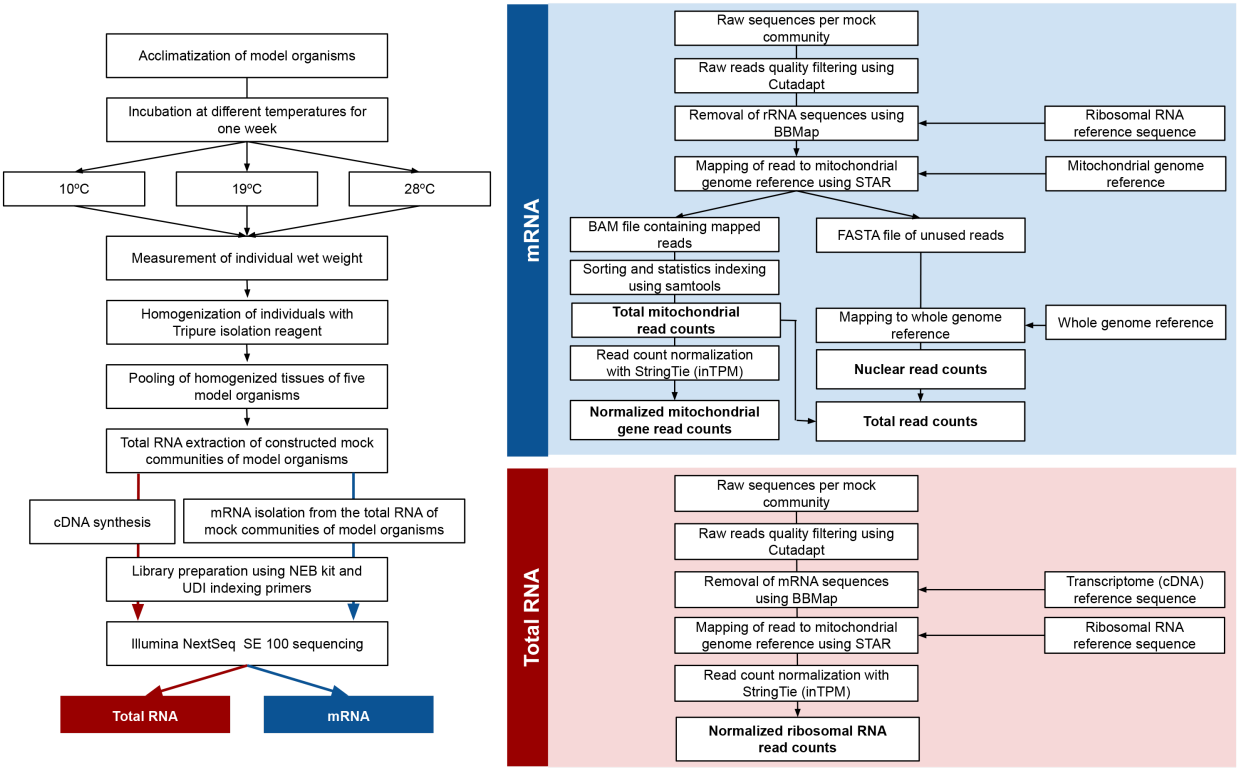

Figure 1. General workflow of the current study

Quality filtering of raw reds using cutadapt

```
cutadapt -pair-filter=any -q 15 --minimum-length 50 -o trimmed1_R1.fastq -p trimmed2_R2.fastq 'path to R1.fastq.gz' 'path to R2.fastq.gz'
```

Removal of noncoding RNA contamination from mRNA library using bbmap

```
'/home/louie/BBMap_38.86/bbmap/bbduk.sh' in1=mRNAinput.fq out1=mRNAoutput.fq outu=mRNAclean.fa ref=mRNAreference.fa k=40 hdist=1 stats=stat
```

Removal of mRNA contamination from total RNA library using bbmap

```
'/home/louie/BBMap_38.86/bbmap/bbduk.sh' in1=inputtotalRNA.fq out1=tRNAoutput.fq outu=tRNAclean.fa ref=tRNAreference.fa k=40 hdist=1 stats=
```

Normalize number of reads in each library by subsampling using seqtk

```
mRNA libraries: seqtk sample -s100 input.fa 24765099 > output.fa
tRNA libraries: seqtk sample -s100 input.fa 18306138 > output.fa
```

Separation of nuclear and mitochondrial transcripts using bbduk

```
'/home/louie/BBMap_38.86/bbmap/bbduk.sh' in1=input.fa outm=mt_sequences.fa outu=nuclear_sequences.fa ref=mt_genome.fa k=40 qhdist=1 stats=s
```

Indexing reference sequence and aligning reads on the reference sequence

```
#Index the reference genome
STAR --runMode genomeGenerate -runThreadN 10 --sjdbGTFfile input.gtf --genomeSAindexNbases 12 --genomeDir path_to_dir --genomeFastaFiles in

#Align normalized reads in reference sequence
STAR --runThreadN 20 --runMode alignReads --genomeDir path_to_dir --readFilesIn input.fa --outFileNamePrefix prefix_name --outSAMtype BAM S
```

Counting number of aligned reads using Samtools

```
for i in *.out.bam; do samtools index ${i} ; done
for i in *.out.bam; do samtools idxstats ${i} > ${i}.txt ; done
```

Normalizing the number of aligned reads based on gene length (transcript per million, TPM) using Stringtie

```
for i in *.out.bam; do '/home/louie/stringtie-2.1.4.Linux_x86_64/stringtie' ${i} > ${i}.txt ; done
```

Identification of ortholog genes from five model species

```
#Removing sequence with similar header and retaining longest one
python keep-longest.py' input_protein.fa > output.fa

#Extract just the longest transcript variant per gene
python '/home/louie/OrthoFinder-master/tools/primary_transcript.py' input.fa
```

```
#Adjust the header of each sequence
perl orthomclAdjustFasta.pl Solenopsis 1 Solenopsis.fa

#Filter the minimum length of sequence and the percentage number of codons
perl '/home/louie/OrthoFinder/Data/filter_seq.pl' '/home/louie/OrthoFinder/Data' 50 20

#Run orthofinder
./orthofinder' -f '/home/louie/OrthoFinder/Data/Data.filtered'

#Concatenate files all single copy orthologs of one species in one fasta file
cat *.fa > single_ortho_all.fa
```

#Back translate amino acid sequence into nucleotide sequence using Backtranstrate

```
backtranseq -seq pep.fa -out output.fa -cfile input.cut
```