

R Markdown
Multiqc HTML
geneBodyCoverage
countsDistribution
sampleCluster
principalComponents
MA-Plot
citations

2008D_Report

Compiled: April 27, 2020



(<https://www.rnaseqcore.vet.cornell.edu/>)

R Markdown

This R Markdown aims to provide a brief summary of all diagnostic plots for your RNA-seq experiment. The fastq files are available for download upon request.

Users receiving files from RSCshare are advised to delete all files once they have securely copied them to their own respective drives.

a copy of your data will be securely archived on our end

Multiqc HTML

The multiqc html summarises the alignment statistics along with the summary of raw counts generated via STAR (<https://academic.oup.com/bioinformatics/article/29/1/15/272537>);

The raw fastq reads were first processed with **trim-galore** (Barbraham Institute) package to:

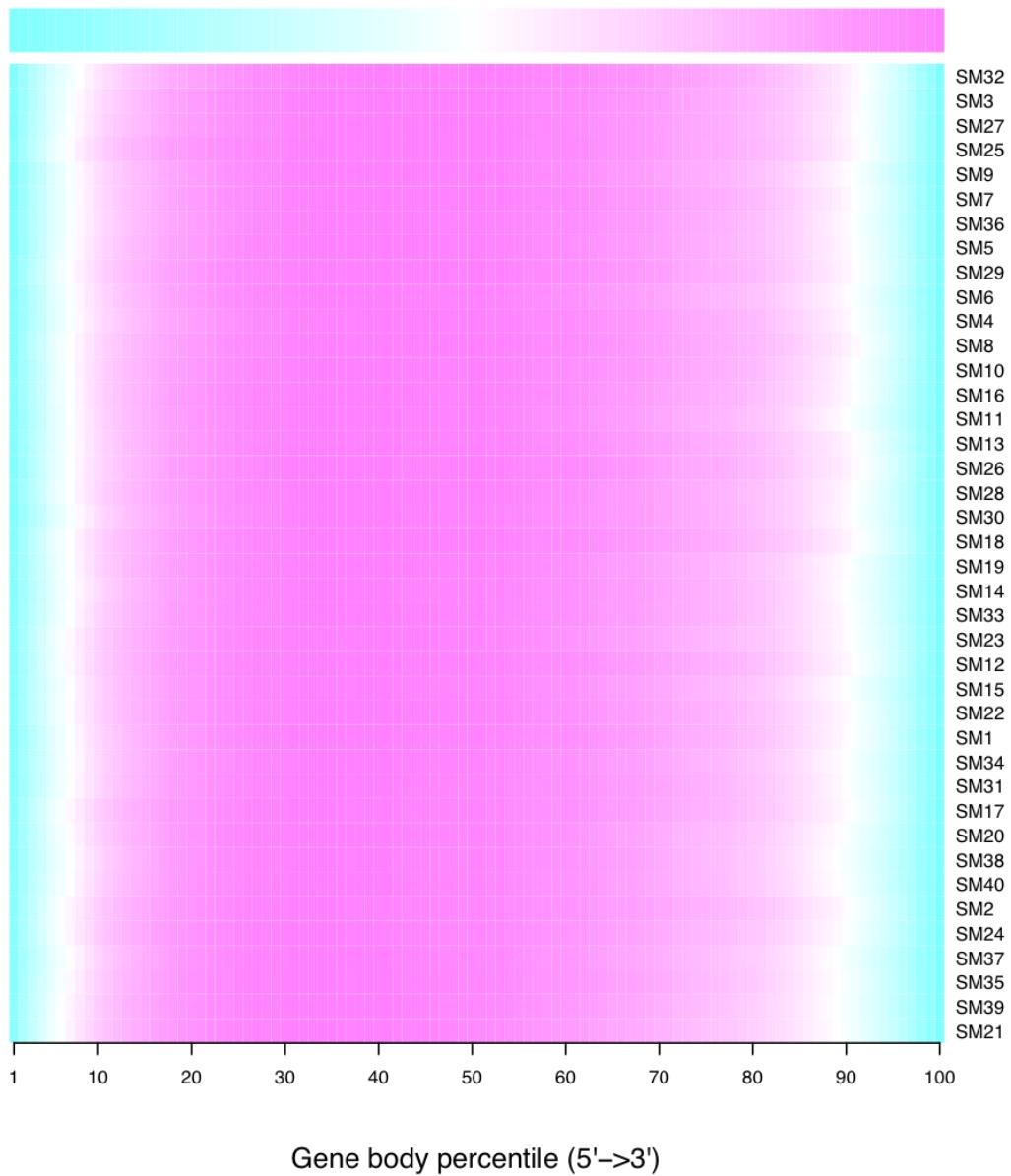
- Trim for low quality reads;
- 2 color chemistry bias (next-seq);
- Trim for noisy short fragments;
- Trim for adapter sequence;

```
trim_galore --nextseq 20 --gzip --length 50 --paired --fastqc
```

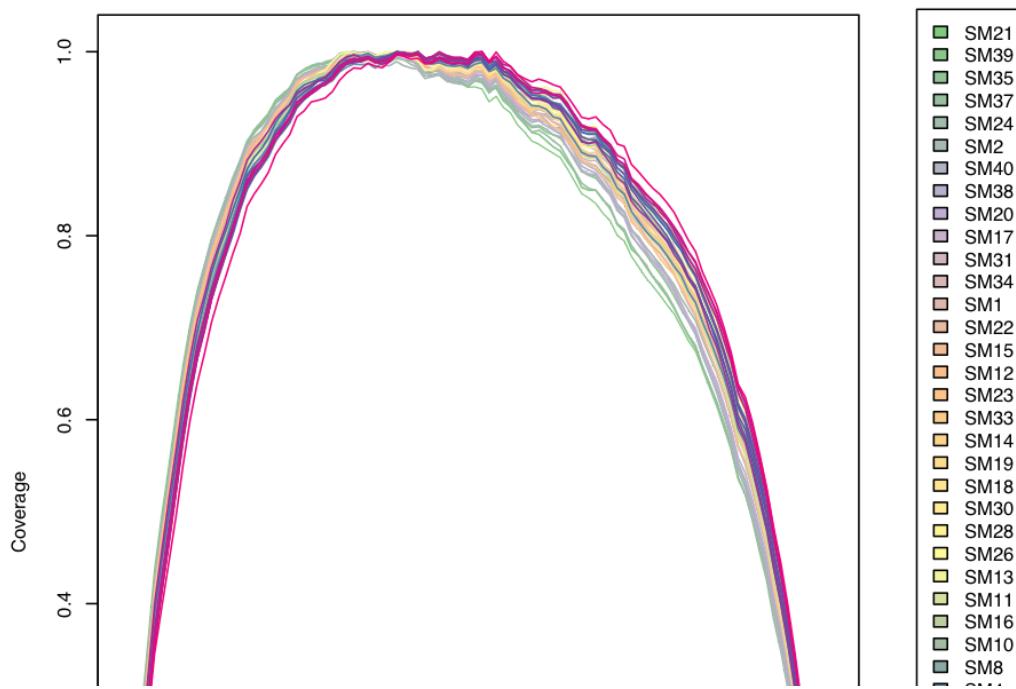
The filtered reads were then aligned to GeoFor_1.0 reference genome with ENSEMBL annotations;

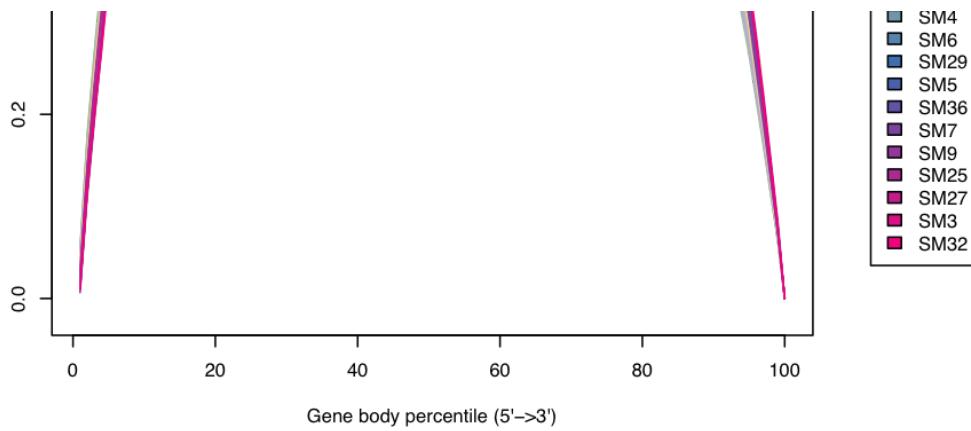
* --paired only applicable for paired-end data; --length 10 for smRNA-seq libraries

geneBodyCoverage

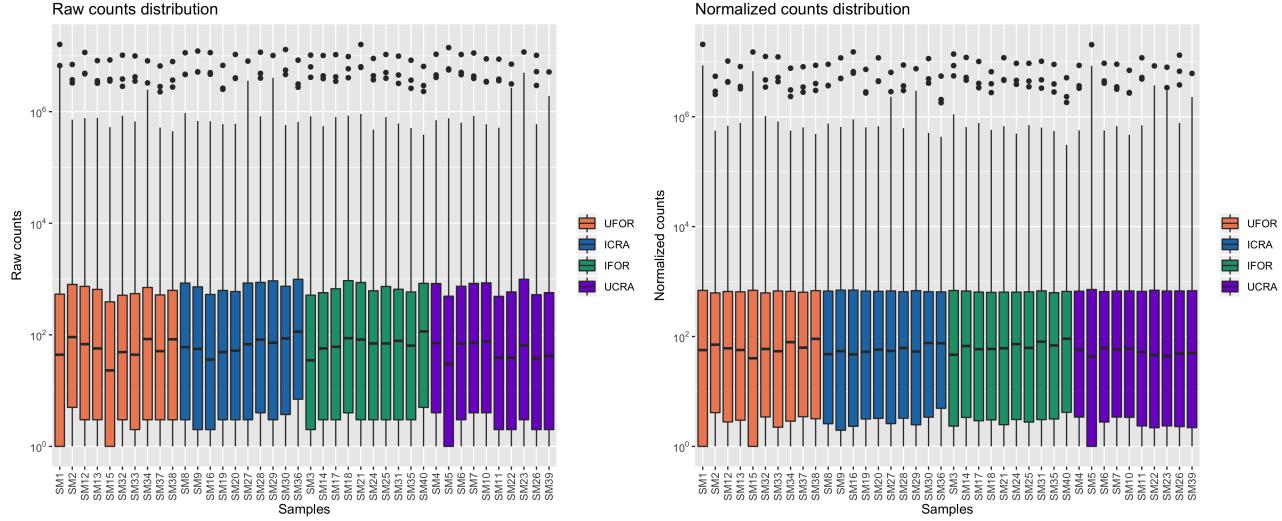


Gene body percentile (5'→3')



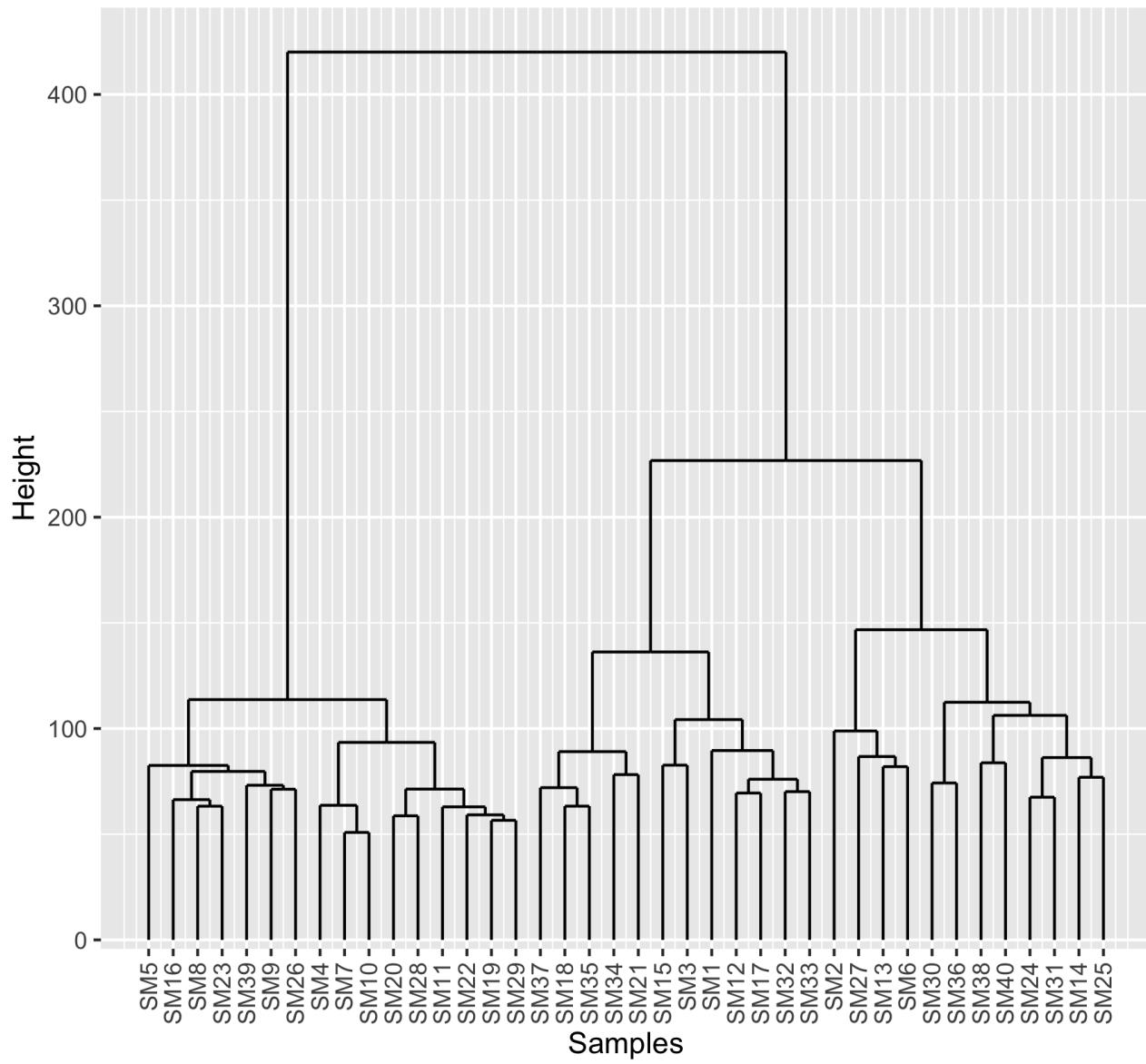


countsDistribution



sampleCluster

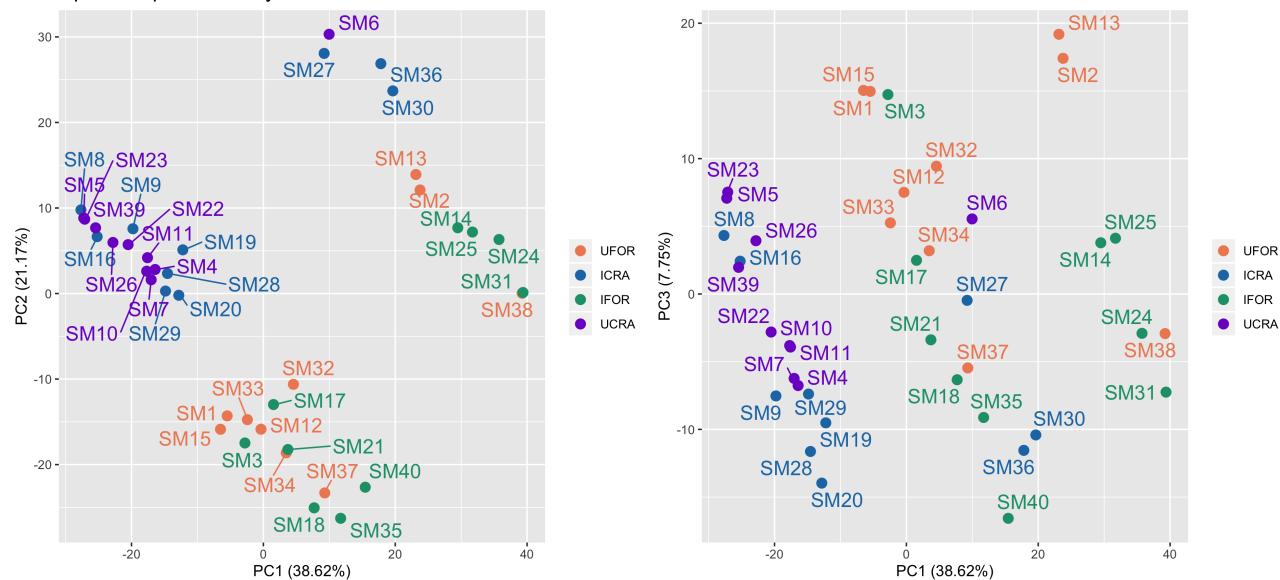
Cluster dendrogram
Euclidean distance, Ward criterion



An euclidean distance is computed between samples, and the dendrogram is built upon the Ward criterion. We expect this dendrogram to group replicates and separate biological conditions.

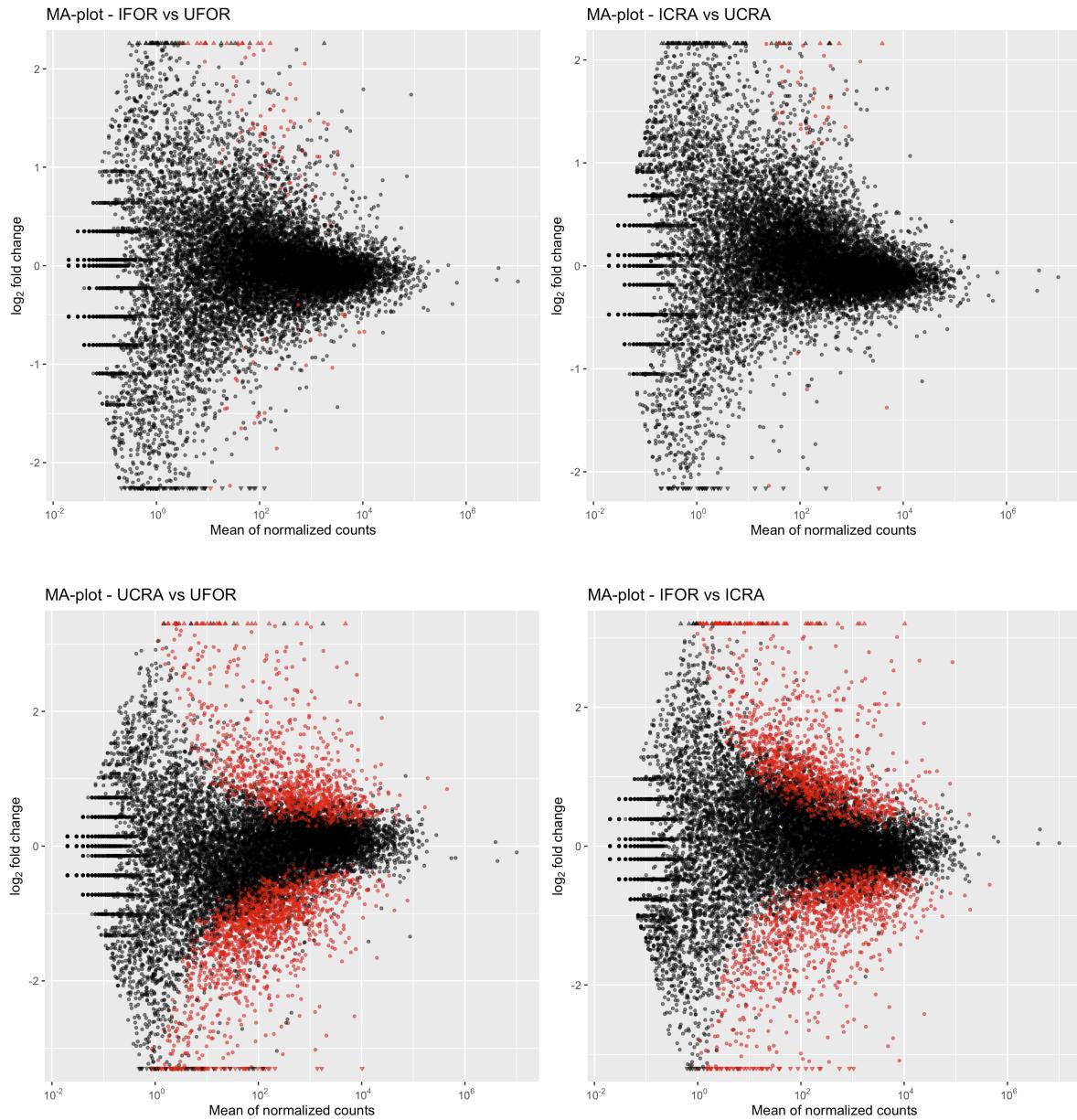
principalComponents

Principal Component Analysis



Another way of visualizing the experiment variability is to look at the first principal components of the PCA. On this figure, the first principal component (PC1) is expected to separate samples from the different biological conditions, meaning that the biological variability is the main source of variance in the data.

MA-Plot



The above figure represents the MA-plot of the data for the comparisons done, where differentially expressed features are highlighted in red. A MA-plot represents the log ratio of differential expression as a function of the mean intensity for each feature. Triangles correspond to features having a too low/high $\log_2(\text{FC})$ to be displayed on the plot.

citations

```
citation("DESeq2")
```

```
##  
##   Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change  
##   and dispersion for RNA-seq data with DESeq2 Genome Biology 15(12):550  
##   (2014)  
##  
## A BibTeX entry for LaTeX users is  
##  
##   @Article{,  
##     title = {Moderated estimation of fold change and dispersion for RNA-seq  
##              data with DESeq2},  
##     author = {Michael I. Love and Wolfgang Huber and Simon Anders},  
##     year = {2014},  
##     journal = {Genome Biology},  
##     doi = {10.1186/s13059-014-0550-8},  
##     volume = {15},  
##     issue = {12},  
##     pages = {550},  
##   }
```

```
citation("SARTools")
```

```
##  
##   Hugo Varet, Loraine Brillet-Gu guen, Jean-Yves Copp e and  
##   Marie-Agn s Dillies (2016): SARTools: A DESeq2- and EdgeR-Based R  
##   Pipeline for Comprehensive Differential Analysis of RNA-Seq Data.  
##   PLoS One, 2016, doi: http://dx.doi.org/10.1371/journal.pone.0157022  
##  
## A BibTeX entry for LaTeX users is  
##  
##   @Article{,  
##     title = {SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive  
##              Differential Analysis of RNA-Seq Data},  
##     author = {Hugo Varet and Loraine Brillet-Gu guen and Jean-Yves Copp e  
##              and Marie-Agn s Dillies},  
##     year = {2016},  
##     journal = {PLoS One},  
##     doi = {10.1371/journal.pone.0157022},  
##     url = {http://dx.doi.org/10.1371/journal.pone.0157022},  
##   }
```