myTAI: evolutionary transcriptomics with R

Hajk-Georg Drost1,2,*, Alexander Gabel2, Jialin Liu3, Marcel Quint4 and Ivo Grosse2,5

1Sainsbury Laboratory Cambridge, University of Cambridge, Cambridge CB2 1LR, UK, 2Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany, 3Université de Lausanne, Département d’Ecologie et d’Evolution, Quartier Sorge, 1015 Lausanne, Switzerland, 4Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany and 5German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

*To whom correspondence should be addressed.

Abstract

Motivation: Next Generation Sequencing (NGS) technologies generate a large amount of high quality transcriptome datasets enabling the investigation of molecular processes on a genomic and metagenomic scale. These transcriptomics studies aim to quantify and compare the molecular phenotypes of the biological processes at hand. Despite the vast increase of available transcriptome datasets, little is known about the evolutionary conservation of those characterized transcriptomes.

Results: The myTAI package implements exploratory analysis functions to infer transcriptome conservation patterns in any transcriptome dataset. Comprehensive documentation of myTAI functions and tutorial vignettes provide step-by-step instructions on how to use the package in an exploratory and computationally reproducible manner.


Contact: hgd23@cam.ac.uk

1 Introduction

To investigate phenotypic changes, diseases, environmental stresses, or developmental processes, transcriptome studies are often the approach of choice. Although transcriptomics is based on a solid methodology, little is known about the evolutionary conservation and dynamics of transcriptomes across species (Drost et al., 2017). Understanding the evolutionary processes that change transcriptomes over time, however, might lead to new insights on how diseases emerge or how phenotypic changes are caused by changes in transcriptomes. For this purpose, evolutionary transcriptomics studies aim to capture and quantify the evolutionary conservation of transcriptomes during specific stages of the biological process of interest (Domazet-Loso and Tautz, 2010). Here, we present the exploratory analysis package myTAI, which can combine evolutionary information of genes with their transcript levels to infer transcriptome conservation patterns. Evolutionary information is given as input to the package and can range from classical phylogenetic or orthology relationships between genes to more recent approaches such as phylogenetic comparative methods (PCMs) (Dunn et al., 2013), phylogenetic reconciliation methods (Doyon et al., 2011), or phylostratigraphy (Domazet-Loso et al., 2007). In summary, starting with a pre-computed table of gene age information and a transcriptome dataset, the R package myTAI can be used to screen for stages of high or low transcriptome conservation within a biological process of interest. If highly conserved or variable transcriptomes were found in particular stages or treatments, more specialized experimental studies could subsequently be designed to investigate the functions and mechanistic implications of these conserved or variable stages.

© The Author(s) 2017. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

Bioinformatics, 34(9), 2018, 1589–1590
doi: 10.1093/bioinformatics/btx835
Advance Access Publication Date: 22 December 2017
2 Implementation

The R package myTAI is released under the GNU General Public License within the CRAN project (R Core Team). The package can be downloaded from https://cran.r-project.org/web/packages/myTAI/index.html. The source code is publicly available at https://github.com/HajkD/myTAI. Internal myTAI functions are implemented in C++ and integrated via the Rcpp (Eddelbuettel, 2013) Application Programming Interface (API) and unit tested using Rtestthat. The myTAI package furthermore depends on R packages nortest, fitdistrplus (Delignette-Muller and Dutang, 2015), dplyr, RColorBrewer, taxize (Chamberlain and Szocs, 2013), reshape2, ggplot2 (Wickham, 2009), biomart (Drost and Paszkowski, 2017), readr, tibble, scales and gridExtra.

3 Functions and Examples

More than fifty functions are provided by the myTAI package. We recently used myTAI to investigate the developmental hourglass model of embryo development (Raff, 1996) on the transcriptomic level (Quint et al., 2012; Drost et al., 2017). Others used myTAI to investigate transcriptome conservation in plant organ development (Lei et al., 2017). To illustrate an example workflow with myTAI, we here use the developmental transcriptome of Arabidopsis thaliana embryo development (Quint et al., 2012) (see Supplementary Material for more details about data formats):

```r
# Import the myTAI package and load example dataset
library(myTAI); data(PhyloExpressionSetExample)

# Plot the Transcriptome Age Index of A. thaliana embryo development
PlotSignature(PhyloExpressionSetExample)

To quantify the transcript level of each gene age category to the overall transcriptome for each developmental stage, the gene expression level distributions for each gene age category can be visualized by:

```r
# Plot gene expression level distributions
PlotCategoryExpr(PhyloExpressionSetExample, 
legendName = "PS", 
log.expr = TRUE)
```

A linear transformation of the mean expression levels into the interval [0,1] enables the comparison of mean expression level patterns between gene age categories independent of their actual mean expression magnitude. A relative expression level of 0 denotes the minimum mean expression level compared to all other stages, and a relative expression level of 1 denotes the maximum mean expression level compared to all other stages:

```r
# Plot relative expression levels
PlotRE(PhyloExpressionSetExample, Groups = list(c(1:3), c(4:12)), 
legendName = "PS", 
adjust.range = TRUE)
```

Finally, we compare relative expression levels between groups of age categories and quantify their difference:

```r
# Compare relative expression levels between groups of age categories
PlotBarRE(PhyloExpressionSetExample, 
Groups = list(group_1 = 1:3, group_2 = 4:12), 
ylab = "Ontogeny", 
lab = "Mean Relative Expression", 
cex = 1.5)
```

In addition to these exploratory functions, myTAI provides functionality for taxonomic information retrieval, gene age enrichment analyses, differential gene expression analyses of age categories, and additional metrics for quantifying transcriptome conservation. A detailed description and interpretation of myTAI functions is available at https://github.com/HajkD/myTAI#tutorials and also in the Supplementary Material.

4 Conclusions

Evolutionary transcriptomics studies can serve as a first approach to screen in silico for the potential existence of evolutionary constraints within a biological process of interest. This is achieved by quantifying transcriptome conservation patterns and their underlying gene sets in biological processes. The exploratory analysis functions implemented in myTAI provide users with a standardized, automated and optimized framework to investigate evolutionary signatures in any transcriptome dataset of interest.

Funding

We thank ERC grant EVORREED [grant number 322621], the SKWP Research Foundation, and the German Science Foundation (grants Qu 141/5-1, Qu 141/6-1, Qu 141/7-1, GR 3526/6-1, GR 3526/7-1, GR 3526/8-1, and FZT 118) for financial support.

Conflict of Interest: none declared.

References