

Order Restricted Inference for Oscillatory Systems (ORIOS) for Detecting Rhythmic Signals

Yolanda Larriba¹, Cristina Rueda¹, Miguel A. Fernández¹ and Shyamal D. Peddada²

¹Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Paseo de Belén 7, 47011, Valladolid, Spain

²Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences (NIEHS), Alexander Dr., RTP, NC 27709, USA

Programmed by: Yolanda Larriba (yolandalarriba@gmail.com)

Version: ORIOS 1.0 (May 10, 2016)

The scripts needed to perform the algorithm are: Run_ORIOSv1.1.R and ORIOSv1.1.R which correspond to the user interface (ORIOS script) and to the internal R routines (source code), respectively.

Protocol:

Formatting your data. ORIOS assumes your data are stored as a tab-delimited .txt file with data for each probe set on a different row (i.e. each row indexes a gene). Each row contains data along the ZT/CT times your samples were collected. Take care to ensure that names of rows and columns have been previously deleted from the text file you store. You will be able to merge them at the end of the analysis if you need.

Setting up R. Open 'R' and change the active directory (File>Change Dir) to a convenient folder. Make sure this folder includes your data file (described above) and the source code for ORIOS ("ORIOSv1.1.R").

Reading your data: Using a text editor (like SciTE and so on), open the "Run_ORIOS .R" file. Then, change the text file name ("ABC.txt") on line 13 to the text file you stored (by default ORIOS reads the file called "ABC"). Change the period length on line 16 according to your data (by default 24 hours long).

Handling time points, time sampling and number of periods: ORIOS assumes one sample per time point. By default, the number of time points that ORIOS considers is equal to the number of columns in the input data file (on line 24). Time sampling and number of periods (on lines 25

and 26 respectively) are derived from the period length and time points. Be careful: there must be consistency among all of them. For example, if you decide to skip odd CT/ZT (i.e. 24 time points) and maintain the period of 24 hours length, then time sampling is increased to 2 hours along the two days (periods).

Initializing and executing routines: A previous routine is required in order to estimate and prepare the filtering and classification steps described in the ORIOS algorithm. Finally, *FilterClass* function provides gene classification with the same FDR levels proposed in the paper (these level can be modified changing default parameters on *FilterClass* function).

Running ORIOS. After modifying the 'Run_ORIOS' script, select all and paste into R's graphical interface. The first half of the script (through line 31) will run quickly; the rest will take some time depending on your machine and the size of the analysis.

ORIOS Output. ORIOS outputs two different kind of files: (1) eight .txt files (two files for each ORIOS signal class, (i.e. cyclical(C), quasi cyclical (QC), non-flat and non-periodic (NFNP) and flat (F)) and, (2) a .rda file, called 'output.RData'. The .rda file contains the number of genes within each ORIOS signal classes (C, QC, NFNP and F). This output can also be shown on the R's console by typing 'output'. For each class, the two .txt files, called 'listC.txt' (resp. 'listQC.txt', 'listNFNP.txt' and 'listF.txt') and 'listPvalC.txt' (resp. 'listPvalQC.txt', 'listPvalNFNP.txt' and 'listPvalF.txt') respectively include the indexes of genes in that class and their corresponding classification p-values.

Citations.

When using ORIOS, please cite: Larriba, Y., Rueda, C., Fernández, M.A. and Peddada, S. D. (2016). Order Restricted Inference for Oscillatory Systems for detecting rhythmic signals.