RhythmicAlly: Your R and Shiny based open source ally for the analysis of biological rhythms

Abhilash Lakshman and Sheeba Vasu

Chronobiology and Behavioural Neurogenetics Laboratories

Jawaharlal Nehru Centre for Advanced Scientific Research

Jakkur, Bangalore – 560064

User Manual

Table of Contents

About RhythmicAlly

Research on circadian rhythms often requires researchers to estimate period, robustness and phase of the rhythm. These are important to estimate owing to the fact that they act as readouts of different features of the underlying clock. The commonly used tools, to this end, suffer from either being very expensive or have very limited interactivity or are very cumbersome to use, or a combination of these. To overcome these issues and to ease the analysis of biological time-series data, we have written RhythmicAlly, an open source program using R and Shiny that has the following advantages; a) it is free, b) it allows subjective marking of phases on actograms, c) provides high interactivity with graphs, d) allows visualisation and storing of data for a batch of individuals simultaneously, and e) does what other free programs do but with fewer mouse clicks, thereby being more efficient and user-friendly. Moreover, in many cases, due to experimental constraints time series data for estimating periodicity can only be obtained over shorter intervals of time (relative to what is required in traditional time-series analyses methods). Such data can be analysed using COSINOR, but given that most biological time-series, even if rhythmic are not sinusoidal, COSINOR is not appropriate in such situations. Autocorrelation based method must be employed and RhythmicAlly allows you to do so (in addition to Enright and Lomb-Scargle periodograms). The first version of RhythmicAlly is available on Github and we aim to maintain the program with subsequent versions having updated methods of analysing and visualising time-series data. Although the goal of RhythmicAlly was to cater to biological rhythms in the circadian range, changing parameters can also allow one to use the program for analysing ultradian or infradian rhythms. Lastly, RhythmicAlly will appeal to researchers who are uncomfortable with coding due to the availability of a user-friendly GUI and also to researchers who are familiar with coding who can very easily modify the codes to tailor the program specifically to their data.

Link to download RhythmicAlly: https://github.com/abhilashlakshman/RhythmicAlly

RhythmicAlly was created and is maintained by Abhilash Lakshman. Any queries regarding usage or bugs in the code must be addressed to abhilashl@jncasr.ac.in.

Last update on: 28th February 2019.

We are planning to submit RhythmicAlly as a manuscript, so that people who use it can cite the program. But until then, it would be great if you acknowledged us in case you use this software to analyse your datasets.

R, RStudio and Initialising RhythmicAlly

RhythmicAlly is written on the R platform with the help of Shiny to make it interactive and user-friendly. It must be acknowledged here, that RhythmicAlly has also used packages written by other authors and will be cited as and when required. Following is a step-by-step approach to initialise RhythmicAlly.

1. Download and install the latest version of R (at the time of the writing of this program the latest version of R was 3.5.2 or "Eggshell Igloo"). The latest version can be downloaded from this link: https://www.r-project.org/. It is important to note here that RhythmicAlly may not work as expected in older versions of R.
2. Download and install RStudio. RStudio is a graphical user interface (GUI) for the R platform and is easier for people to visualise and use. RStudio can be downloaded from here: https://www.rstudio.com/.
3. Download all the contents of the RhythmicAlly folder from Github ().
4. Inside the folder there will two subfolders, and initialisation.R file and this user manual.
5. One of the subfolders which is called 'For_DAM' contains the application for running when your data is acquired using the DAM system; the other which is called 'Others' contains the application for running when your data is acquired outside of the DAM system. Both these subfolders have a few example data sets in a tab delimited text file. Make sure that your data is arranged in the same format as that of the data sets in the example files.
6. First, open the initialisation.R file in RStudio and run it. You must select all the code (Ctrl+A) and then press (Ctrl+Enter). This will install all the required packages for running RhythmicAlly.
7. Now you ARE ready to use RhythmicAlly. The above steps need to be performed only the first time you install all these programs. The next steps are what you will do every time you want to analyse a new data set.
8. Open RStudio and type the following in your console and press the Enter key: "**shiny::runApp("***Location_of_the_folder_where_you_have_saved_RhythmicAlly***", launch.browser = TRUE**". If you are copying and pasting the command, do not copy the double quotes. The "launch.browser = TRUE" will make sure that the GUI opens in your system's default browser. It is advisable to avoid Firefox and Edge.
9. The analysis interface will open in your default browser, and will look like this:
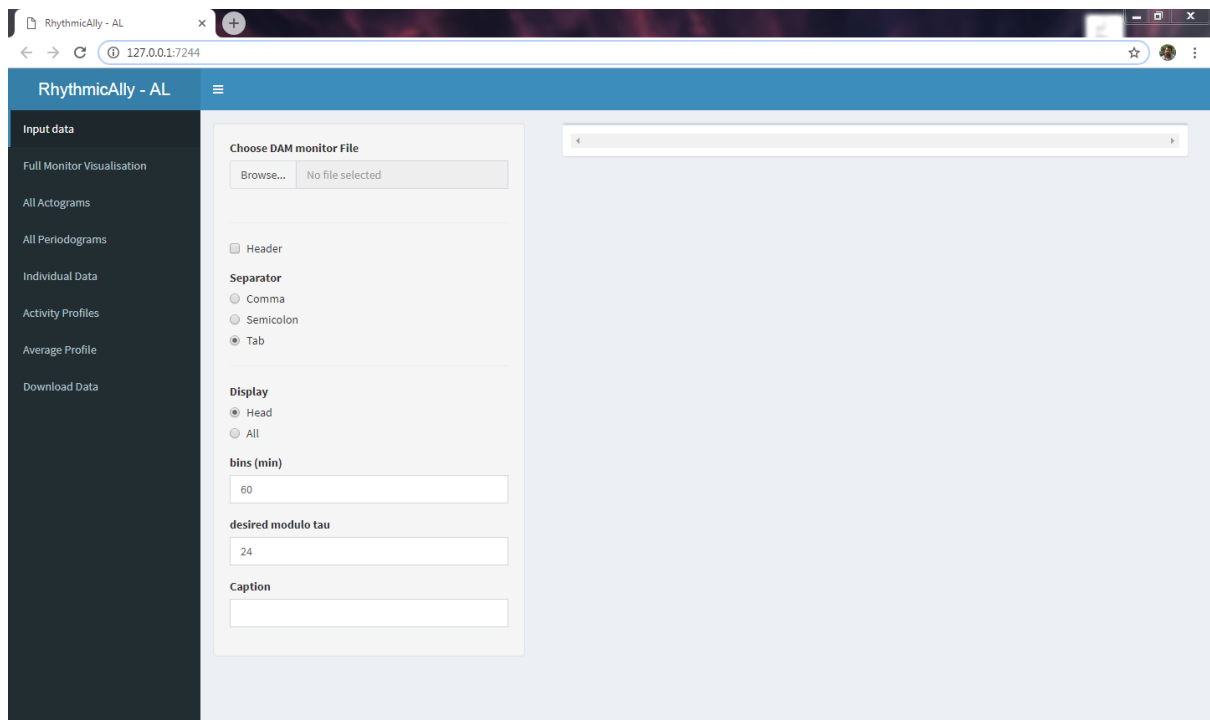
10. This is an example of the home screen of the analysis program from the application in the 'For_DAM' folder.

Input Data

1. The dark side pane of the application will be referred to as the side panel and the light blue part of the application will be referred to as the main panel.
2. On the top left corner of the main panel you can use the Browse button to browse and upload your raw time-series data.
3. If you are running the application from the files in the 'For_DAM' subfolder, then your input file must be a scanned monitor file acquired from a DAM (Drosophila Activity Monitor by TriKinetics) recording system. Otherwise the input file can be acquired from anywhere else as long as it conforms in terms of format to the example data sets.
4. The Header option must be checked only if there are column titles for your data.
5. The Seperator option determines how data are separated in your text file.
6. The Display option tells the program whether or not to display the entire data set or just the first few rows.
7. The bins (mins) option requires you to input in what bin your data was collected.
8. The desired modulo tau function tells the program how to treat your data while plotting actograms and computing average profiles.
9. The Caption input must be a character string, something that will uniquely identify your data set when it is downloaded.
10. Once you upload the input dataset, the right side of the main panel will display the summary of your data set. Make sure the data is getting displayed correctly before proceeding.

11. The bin (mins), desired modulo tau input values will be used in all the tabs in the side panel except the Download Data tab. The Caption input value will be used in the Full Monitor Visualisation and Download Data tabs only.

Full Monitor Visualisation

1. When you click on the Full Monitor Visualisation tab, you will see on a heatmap interactive visualisation of the entire data set that you have uploaded will appear. The x-axis of this plot is time index and the y-axis is the individual data (could be individual fly or anything else or vials). Each ribbon is depicting activity counts or any other biological variable that you are measuring as a time-series. The colour scale represents the value of the variable. The lower most ribbon depicts the average time-series across all individuals.
2. On the top of the main panel you will see a slider input. This places a cap on the maximum value of activity/variable that is being measured that will correspond to the highest colour value provided in the colour bar.
3. This heatmap is interactive. This means that when your cursor is hovering on any point on the plot, it gives you information of how much activity there is and what is the phase at the point at which your cursor is located. Also, you can drag and zoom into specific regions of interest to visualise cycle-to-cycle and across individual variation in the time series data. You can double click to restore back to original scale.
4. Moreover, this will also allow you to see immediately which individuals have died or have no recording thereby allowing you to remove them from your downstream analyses.

All Actograms

1. This tab will show you a heatmap representation of the time series for all individuals in the batch that you are analysing.
2. There are two additional inputs here that you can modulate. On the top left corner of the main panel, you can choose how many plots you want your actogram to be, i.e., they could be either double-plotted or triple-plotted etc.
3. The second input is the same as that in the Full Monitor Visualisation tab.
4. Similar to the plot in the Full Monitor Visualisation tab, these plots are also interactive.
5. The extra inputs that will be used in this tab will be utilised in this tab only.

The last actogram is always of the time-series averaged over all individuals. So if your batch has 32 flies, you will be able to see 33 actograms, 33rd being the batch actogram.

All Periodograms

This tab is useful for analysing periodicity and robustness of the time-series data that you have collected. There are two sections to this tab.

Periodogram plots

1. Based on the input that you choose on the left side of the main panel, periodogram plots for all individuals in your DAM data set or other non-DAM data sets will show up.
2. The available choices for time series analysis are Enright periodogram or autocorrelation or Lomb-Scargle. The time resolution for analysis of time-series using Enright periodogram is set at the bin value (in other words, maximum possible resolution).
3. You are free to choose the shortest and longest period values in your analysis and also set p-value cut-off for your data set.
4. These plots are displayed on the top right hand side of the main panel.
5. Input values provided in this tab will be used in this and the Individual Data tabs.

Data storage

1. Data storage is facilitated by a 'click to store' method in a table in the bottom right side of the main panel.
2. Given that these plots are interactive, when you hover your cursor around the periodogram it will display the period value at that point and its corresponding power. Both these values get stored if you click on the point ONCE. If you click twice, then the same reading is recorded twice. If you zoom in to one of the periodogram plots then in this case you must not double click to restore original scale. This will record some random value twice in the table.
3. In such cases there is a button on the top right side of the plot which can be clicked once to restore original scales.
4. All the displayed values are stored and can be accessed again if you wish to download them in the Download Data tab.

The last periodogram is always of the time-series averaged over all individuals. So if your batch has 32 flies, you will be able to see 33 periodograms, 33rd being the batch periodogram.

Individual Data

This tab has its own set of input values, all of which will again be exclusively used in this tab only. The channel input is the individual ID i.e., 1 will indicate individual 1, 2 will indicate individual 2 and so on. The nplots input refers to the number of plots you want while visualising your actograms. The threshold tab puts a cap on the amount of activity/variable that will get displayed in each row of the actogram. The starting time input refers to the first time-point in your time-series input data file. Phases for storage are calculated from that time-point.

Individual Actogram

1. This is displayed in the middle of the main panel. The actogram displayed will be of the channel that has been mentioned in the left hand side input pane in the main panel.

Corresponding Periodogram

1. The periodogram analysis will be performed for the channel that is entered in the channel input and the analyses will be based on parameters chosen in the previous tab.

Data Storage

1. Data storage is facilitated by a 'click to store' method in a table in the bottom of the main panel.
2. Given that these plots are interactive, when you hover your cursor around the actogram it will display the time-index, activity count and the trace number (this corresponds to the cycle that you are recording data from if you use the first plot in the actogram). The time index value and trace number are recorded when you click on the point in the actogram. The time-index value is converted to the hours after the starting time and that phase is recorded in the table below. This way one can subjectively mark phases, something that, to the best of my knowledge, only ClockLab allows. Both these values get stored if you click on the point ONCE. If you click twice, then the same reading is recorded twice.
3. Every time you change the channel input, the actogram and periodogram plots will update to give the new computed plots. The table will not change and you can continue recording phases of the new actogram.
4. All the displayed values are stored and can be accessed again if you wish to download them in the Download Data tab.

The last actogram is always of the time-series averaged over all individuals. So if your batch has 32 flies, you will be able to see 33 actograms, 33rd being the batch actogram.

Activity Profiles

1. This computes activity profiles for each individual averaged over cycles based on the input values provided in the Input Data tab.
2. The top part of the main panel shows activity of all individuals averaged over all cycles.
3. The bottom part of the main panel just shows the data for all individuals that are plotted in the above panel.
4. Given that these are interactive plots, each profile can like the other plots, show activity level and phases of peak or anticipation etc for all individuals simultaneously.

Average Profile

1. This is merely the profile averaged over all the individuals in the previous tab.

Download Data

1. The left side of the main panel shows a drop down menu. You can use this to choose a data set that you want to download.
2. This data set shows up on the right side of the main panel. Once you confirm that this is the data set of your interest that you want to download, you can click on the Download data button on the left side and the file will be downloaded as a comma separated file in your default downloads folder.

Additional tabs in the RhymicAlly program in the 'Others' subfolder

Raw Ind Profiles

1. The top part of the main panel shows the profile of your behaviour on a modulo-τ scale after averaging raw values over all cycles.
2. The bottom part of the main panel shows the raw data that has been plotted and will be available for download in the Download Data tab.

Prop Ind Profiles

In many cases due to high cycle-to-cycle variation in behaviour, it is crucial to use the observed value as a fraction of total value over that cycle in order to avoid misinterpretation of data. So analyses in this tab first calculates the proportion of total value of variable that is exhibited in each time-point and then saves those values for downloading and plotting.

1. The top part of the main panel shows the profile of your behaviour on a modulo-τ scale after averaging proportion values over all cycles.
2. The bottom part of the main panel shows the proportion data that has been plotted and will be available for download in the Download Data tab.

Average Profiles

The average profiles tab in this case has two profiles, one averaged over all individuals using the raw data and the other is averaged over all individuals using proportion data.

Rose plots

These are circular analogs of a histogram plot. In other words, you can think of them as circular distributions of your variable. 0° corresponds to 12am (00-h or 24-h).

1. The top left side of the main panel has input for individual id. Every time you change this value, the new rose plot will get updated. Note that the radial axis scale will change with individuals. So it is important to be careful during interpretation of the data.
2. The data used for these rose plots are raw values of the variable averaged over all cycles for each individual.

CoM

Phases, by definition, are circular random variables and therefore must be treated as such whenever possible. Phase of Centre of Mass (CoM) is the mean phase of the oscillation in polar coordinates, and is an unbiased and non-subjective measure of phase of an oscillation. Like in an X-Y Cartesian plane, points on a polar coordinate plane can be defined using a) the distance from the centre (referred to as r) and b) the angle the point makes from an arbitrary 0 (referred to as ϑ). While ϑ represents the phase of the oscillation, r represents the consolidation of your variable across time-of-day. For instance, the r value can be thought of as the gate-width of eclosion rhythm. These parameters are computed using the raw values of the variable averaged over all cycles for each individual.

1. The top left side of the main panel shows the ϑ and r values for each individual. This allows the user to view across individual variation in phase and consolidation of the rhythm.
2. The table in the bottom of the figure in the main panel has the values stored for each individual (these are the values that have been plotted). These results will be available for download in the Download Data tab.

Things to note

1. The first time you click a tab, the program will first perform computations and then plot, so when you have a lot of data it may be slow.
2. When this computation is happening or the data is getting updated from the input that you have provided the screen will plot region will grey out. Please wait for the new figure to get displayed.
3. Once the computation is done and then you go to another tab and then come back to the previous tab, all the figures will immediately get displayed without any delay. This is because the computations have already been done and stored. This however will not be the case, if you update the input variable. Every time you update you must wait a brief moment until the new plots are remade.
4. Every time you input a numeric input in any tab: If you use the scroll bar to move across values then for each value that you crossed, the plots will be remade. This is okay when you are intending to look at data in sequence. However, if you are interested in moving to individual 5 from individual 1, it will be more efficient to type 5 instead of using the scroll bar to navigate from 1 to 2 to 3 to 4 to 5.
5. All the plots have a capture image button on the top right hand side that is inbuilt with Plotly (the program that was used to make all the graphs). This will capture and save the image in '.png' format on your local hard drive.
6. In the RhythmicAlly program that will be run from the 'Others' subfolder can only handle a maximum number of 32 individuals. So if you have data where there are more, please split it such that you process only 32 at a time.
7. When you run RhythmicAlly from the 'Others' subfolder, make sure your input text file has time in decimal format and not in HH:MM:SS format.

Acknowledgements

It is imperative to acknowledge all the authors of several earlier packages that I have used to make this program.

Firstly, R for providing a free platform to code and make analysis easier for everyone. Secondly, Shiny and Shiny Dashboard, that made it possible to make the environment user friendly and interactive. Plotly is an amazing interactive graph builder and has been extensively used in this program. The interactivity brought about by Plotly has been crucial for the development of this program. Additionally, our program uses the following packages, zoo, RColorBrewer, pracma and zeitgebr.

I also thank all the people who have provided me with example data sets to test this program.

References

Achim Zeileis and Gabor Grothendieck (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06

Carson Sievert (2018) plotly for R. https://plotly-book.cpsievert.me

Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

Hans W. Borchers (2018). pracma: Practical Numerical Math Functions. R package version 2.2.2. https://CRAN.R-project.org/package=pracma

Quentin Geissmann and Luis Garcia (2018). zeitgebr: Analysis of Circadian Behaviours. R package version 0.3.3. https://CRAN.R-project.org/package=zeitgebr

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Winston Chang and Barbara Borges Ribeiro (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.0. https://CRAN.R-project.org/package=shinydashboard

Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2018). shiny: Web Application Framework for R. R package version 1.1.0. https://CRAN.R-project.org/package=shiny