VANESSA – Shiny apps for accelerated time-series analysis and visualization of *Drosophila* circadian rhythm and sleep data

Arijit Ghosh, Vasu Sheeba*

Chronobiology and Behavioral Neurogenetics Laboratory, Neuroscience Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore – 560064, Karnataka, India

Running title: *VANESSA – circadian rhythm and sleep analysis apps*

*Correspondence: Vasu Sheeba, Chronobiology and Behavioral Neurogenetics Laboratories, Neuroscience Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bengaluru - 560064, Karnataka, India. Phone: +91-080-2208-2987. Email: sheeba@jncasr.ac.in

Abstract:

Chronobiologists and sleep researchers often need to estimate various rhythm and sleep parameters from locomotor activity data from different organisms. The available open-source or expensive paid tools do not offer consolidated analysis and visualization options in one bundle, are often cumbersome for users unfamiliar with coding, offer very low customization options, introduce sources of human errors by requiring users to manually pick period and power values from periodogram plots, and do not generate reproducible reports. We present VANESSA, a family of cross-platform apps written in R, which, in our opinion have several advantages compared to available tools – (a) open-source, (b) automatic period-power detection, (c) timeseries filtering and smoothing, (d) high-resolution publication-quality figures with dynamic coloring, resizing and light/dark shading, (e) reproducible code-report generation, (f) analysis and visualization of multiple monitor files, defining genotypes and replicates separately, and (g) sleep profile analysis, various sleep parameter estimations, quantification, bout analysis, and latency analysis. The current version of the app is for data acquired through *Drosophila* Activity Monitors (DAM, Trikintetics™) but can be easily extended to that from other data acquisition systems and from other organisms. We will continue to develop VANESSA with more useful features and version control will be done via archiving versions with significant changes on GitHub (https://github.com/orijitghosh/VANESSA-DAM) and Zenodo.

Keywords: *Circadian rhythm analysis, Sleep analysis, Drosophila, Time-series analysis, periodogram, DAM, open-source*

Timeseries analysis involves the evaluation of sequential data obtained over time – data collected at specific intervals, where time is an independent variable. Circadian rhythm and sleep researchers are mostly focused on events occurring at a *circa-dian* (~approximately a day) scale. Since the proliferation of circadian rhythm research in the late 1970's, multiple methods have been developed to decipher patterns in data collected to study rhythms in different organisms. Mostly these researchers wish to estimate periodicity, phase, and robustness of the rhythms; whereas sleep researchers are interested in sleep architecture, sleep parameter estimation, and latency to sleep. Rhythm parameters such as presence or absence of rhythmicity, period, and amplitude tell us about the robustness of the rhythm and whether a periodicity of circadian time scales is present or not. Sleep parameters such as total sleep time, bout numbers, bout lengths, latency tell us the amount, quality of sleep, and overall sleep architecture. Calculation of these parameters enable us to compare individuals under different conditions, genotypes, treatment groups etc., and let us uncover underlying mechanisms.

Over the past five decades, *Drosophila melanogaster* has emerged as a widely used model to study circadian rhythms and sleep. The *Drosophila* Activity Monitor (DAM) systems from Trikinetics (https://www.trikinetics.com/) are the most often used systems for automated recording of *Drosophila* locomotor activity data for circadian rhythm and sleep research. Over the years various free and paid tools have emerged to analyze DAM system outputs, some notable ones with Graphical User Interface (GUI) are - ClockLab (*ClockLab* | *Actimetrics*), El Temps (*el temps principal*), ShinyR-DAM (Cichewicz and Hirsh, 2018), ActogramJ (Schmid et al., 2011), RhythmicAlly (Abhilash and Sheeba, 2019) etc.

There are fewer publicly available open-source tools for sleep analysis, one of the major tools being pySolo (Gilestro and Cirelli, 2009), a python based program with a GUI and another tool – ShinyR-DAM, an R-based tool with a GUI and a webserver. However, these tools each have their own set of limitations, including, but not limited to – analysis of only one single DAM monitor file at a time preventing quick comparisons between monitors or genotypes/treatments/conditions/sex during analysis and visualization, inability to calculate and plot replicates separately, unavailability of multiple widely used periodogram methods within one tool, lack of access to reproducible minimal codes for publication, high probability of human errors due to manual preparation of input files without metadata, inability to produce publication-quality

figures with customization, unavailability of automatic period-power detection from periodograms, lack of features to normalize activity counts of individuals to facilitate comparison among individuals, among *identifiers* (genotypes/treatments/conditions/sex etc.), and among experiments, inability to provide estimates of important sleep features separately for light and dark phase of the day such as latency, details of sleep bouts etc.

(Figure1)

To alleviate these shortcomings in available open-source tools, we developed VANESSA (Visualization and ANalysis of timE SerieS dAta), an R-based set of tools to explore and analyze timeseries data primarily from DAM systems. Here we present the first two tools in VANESSA – (a) VANESSA-DAM-CRA (for circadian rhythm analysis) and (b) VANESSA-DAM-SA (for sleep analysis). Both of these apps have easy-to use GUIs, written completely in R (R Core Team, 2021) and GUIs deployed with Shiny (Winston et al., 2019). Both the apps are available as R packages from GitHub and are also deployed on shinyapps.io server to be used directly from a browser with an internet connection. VANESSA makes use of core R functions, packages from the TidyVerse, and some packages from the rethomics framework (Geissmann et al., 2019) – *behavr, damr, ggetho, zeitgebr* and *sleepr*. All other packages used are listed separately. A brief description of functionalities of VANESSA-DAM-CRA and VANESSA-DAM-SA follows:

VANESSA-DAM common features:

The only requirements of both the apps are the creation of a metadata file, containing details of monitors, *identifiers* (genotypes/treatments/conditions/sex etc.), replicate information (if any used, e.g., Genotype A, B, C; Treatment 1, 2, 3), experiment start and end date time, and the monitor files scanned from the DAMScan program (Trikintetics™) (*Fig. 1 A & B*). The information in the metadata file can serve as experimental records and is used to determine zeitgeber on/off time for analyses and plots, and for assigning *identifiers* for each channel of the monitors. The metadata file can be prepared manually in any spreadsheet program or created using the apps themselves. The data is then curated (data for dead flies removed – *see Documentation tab of the app*), data pertaining to the days specified by the user is extracted, day-wise normalization of activity counts is done, and sleep bouts are identified based on 5 minutes of immobility (Hendricks et al., 2000; Shaw et al., 2000). A detailed user guide is available as Supplementary Material (*Supplementary methods 1 & 2*) and from

https://github.com/orijitghosh/VANESSA-DAM. All plots made by the app can be resized, recolored and re-binned (binning of data over time) anytime dynamically.

(Figure2)

VANESSA-DAM-CRA specific features:

Periodograms:

Four popular periodogram methods have been incorporated in the app – (i) autocorrelation, (ii) chi-square (greedy chi-square method, improved method over the classical chi-square method), (iii) Lomb-Scargle, (iv) continuous wavelet transformation (CWT) (*Fig. 1 C & D*). Users can choose upper limit, lower limit and significance threshold for the periodogram analysis for all methods (*Fig. 1 C*). Arrhythmic individuals can be removed from the dataset after determining periodicity so that they do not affect further analysis and visualization of average profiles, average periodograms and batch actograms. All period and power values of each individual can be downloaded as csv files where different peaks of periodograms for all individuals will be saved (*Fig. 1 D*). Four useful visualizations are provided for period data – (i) all individual periodograms can be plotted captioned with monitor and channel number, *identifier* and replicate number with user specified colors of the particular *identifiers*; The highest peak of the periodograms will be identified and printed in these plots, (ii) periodograms can be averaged over individuals of an *identifier* and plotted, (iii) violin plots of all *identifiers* can be plotted along with individual data points and the mean of each *identifier* will be calculated and printed on the plots, (iv) a density plot of periods of *identifiers* can be plotted along with individual period values on the *x*-axis (*Fig. 1 E-G*).

Actograms and average profiles:

Activity counts can be visualized as heatmap "*ethograms*" for raw data and curated data (*Fig. 2 A*). All double-plotted actograms for all individuals can be visualized either together enabling quick and clear visualization of individuals that exhibit abnormal/unique and thus interesting patterns of activity, or separately, one-by-one along with their periodograms displayed below them enabling ease of picking representative actograms (*Fig. 2 B & C*). Double-plotted batch actograms are also available for each *identifier* separately (*Fig. 2 G*). Average profiles for individuals and all individuals of an *identifier* can be visualized either as a timeseries of

consecutive days or averaging across days (*Fig. 2 E*). Average profiles of activity can also be visualized via a circular plot for each *identifier* (*Fig. 2 D*). All these plots can be plotted either with raw activity counts or with normalized activity (*Fig. 2 F & H*).

CWT spectrogram and timeseries smoothing:

CWT spectrograms can be visualized for individuals over several days (*Supplementary methods 2 – step 18*). This function does not require the metadata file, instead the input is the monitor file itself. Wavelet powers for different periods scanned can also be visualized. The timeseries smoothing function operates as a separate module and takes one single monitor file as input (*Supplementary methods 2 – step 19*). Users can upload a monitor file, change the binning of the data, use two popular methods of timeseries smoothening (lowpass Butterworth filter and kernel smoothing). For Butterworth filter, users can change "*filter order or generic filter model*" and "*critical frequencies of the filter*" and immediately check the effect of the binning and filtering on average profile of individuals of the monitor and on the individual profiles and download the smoothened data with applied parameters. Similarly, for kernel smoothing, users can specify the "*kernel smoothing bandwidth*".

(Figure3)

VANESSA-DAM-SA specific features:

Sleep profiles:

Sleep bouts of each individual can be visualized as heatmaps across days. The other methods of sleep bout visualization are – (i) plots of sleep bouts across days for each individual (ii) sleep bouts can be averaged across chosen days for each individual and plotted, (iii) sleep bouts of individuals of an *identifier* can be averaged over chosen days and plotted, (iv) sleep bouts can be first averaged across days for each individual and then averaged across individuals of an identifier and plotted in a linear manner or in a circular scale (*Fig. 3 A-G*).

Sleep fractions, bout, and latency analysis:

Sleep analysis results such as sleep levels, bout details, latency data can be downloaded as csv files for all individuals separately for light and dark phases. Total sleep fractions and total sleep times of different *identifiers* for chosen days can be plotted as violin plots. Sleep fractions

and total times in the light and dark part of the day can also be plotted for different *identifiers* as violin plots. Bout analysis can be visualized in different ways – (i) sleep and awake bouts can be plotted for individuals of an *identifier*, (ii) violin plot of number of sleep and awake bouts for an *identifier*, (iii) violin plot of mean sleep and awake bout lengths for an *identifier*, (iv) violin plot of sleep latency for an *identifier*, (v) for all measurements, light and dark phase values can be plotted as separate violin plots (*Fig. 3 H & I, Fig. 4 A-K*).

(Figure4)

Reproducible codes:

Both apps have capability of generating HTML files with all parameters used by the user for analysis in the session and codes to reproduce the same exact analyses to generate the tables and plots (*Fig. 5 A-C*). Advanced users can tweak different codes for analyses to suit their needs and generate custom plots. For timeseries smoothing, a HTML report file is generated with all parameters used for reproducing similar smoothing and binning of data.

(Figure5)

The codes for these apps are hosted on GitHub (https://github.com/orijitghosh/VANESSA-DAM) and freely available under an MIT license along with toy/demo data. The apps are also hosted on a server and can directly be used from a browser with internet connection (https://cryptodice.shinyapps.io/vanessa-dam-cra/ and https://cryptodice.shinyapps.io/vanessa-dam-sa/). Additionally, VANESSA is available from GitHub as R packages and the instructions and usage are on the VANESSA-DAM GitHub page. VANESSA is also easily extendable to analyze locomotor activity data or any other numerical rhythmic behavioral data from other organisms recorded with systems other than DAM, and we will try our best to provide support for converting user data to DAM compatible files. Such an example is available in the Supplementary materials (*Supplementary figure 1*) where mouse locomotor activity data recorded with Trikinetics software was easily converted to VANESSA compatible files easily with a few lines of code. We plan to host an online webserver for converting other format data to VANESSA compatible files, as and when we get requests from users. For now, only mouse data conversion is in place as a function from the vanessadamcra package. VANESSA is easily extendable to analysis of ultradian or infradian rhythms, and with a few modifications, can be used for analyzing rhythms other than

locomotor activity. We welcome feature requests, suggestions, bug reports and collaborations via GitHub or email. We have provided tips regarding interpretation of different plotting options, proper usage of different periodogram methods with appropriate parameters in the VANESSA-DAM wiki (https://github.com/orijitghosh/VANESSA-DAM/wiki/Good-practices/). We believe VANESSA apps will be extremely useful to fly circadian rhythm and sleep researchers, especially those working with high throughput behavioral screenings where analyses need to be automated, and reproducibility of analysis methods is critical. Future updates will include but not be limited to – calculation of the angle-doubled center of mass for bimodal actograms, support of generating metadata file for more than 12 monitor files, in-built statistical tests, subjective and objective phase marking, support for more data acquisition systems for more organisms, sleep analysis in specific time windows etc.

Acknowledgments:

We thank two anonymous reviewers, Dr. Nishan Kannan's lab members, Dr. Pankaj Yadav's lab members, members from our lab, Dr. Quentin Geissmann and Dr. Tanya Leise for using the apps and giving feedback which were instrumental in modifying / adding various features. We also thank Aishwarya Ramakrishnan and Dr. Shahnaz R Lone for providing the demo data supplied with the apps. We would also like to acknowledge financial support from the Science and Engineering Research Board, New Delhi, to V.S. (CRG/2019/006802); intramural funding from Jawaharlal Nehru Center for Advanced Scientific Research (JNCASR), Bangalore; the Council of Scientific & Industrial Research (CSIR), New Delhi, India, for a junior research fellowship and a senior research fellowship to A.G.

Figure legends:

Figure 1: General description and period-power calculation in VANESSA-DAM. (A) The home tab of VANESSA-DAM – the Data Input tab. This tab has different parameters that can be changed according to the user's analysis needs – modulo-Tao, the LD cycle period in the experiment, summary time window, light duration, total number of monitors being analyzed, replicate declaration, subset data to specific days and indicate for how many days flies must be alive to be included in calculations and plots. This tab can also be used to customize colors for different *identifiers*. (B) Examples of periodogram calculations using Chi-square method from the "Periodogram" tab. Users can easily define upper and lower limits of periods to be scanned and

also define significance threshold for the periodogram method. (C) Example of a preview of the tabulated data file that can be downloaded with all period and power values for all individuals for all significant peaks in periodogram. (D) Example of a density plot of period values of individuals from different identifiers. All individual period values are plotted as short horizontal lines - "|" on the x-axis. (E) Average periodogram of all individuals from different identifiers. (F) Violin plot of period values of all individuals from different identifiers are plotted along with the mean period value of that identifier embedded in the plot.

Figure 2: Actograms and activity profile visualization in VANESSA-DAM-CRA. (A) Activity profiles plotted as an ethogram where darkest blue color depicts highest activity and complete white indicates zero activity. Ethograms are an easy way to look at data from all individuals at a glance and notice any abnormal or unexpected patterns. (B) Batch actograms of different *identifiers*. The light-dark shading can be quickly altered by changing the "Duration of light in hours" parameters in the "Data input" tab. (C) Actograms of all individuals of one *identifier* can be visualized together. (D) Average activity profiles can also be visualized in a polar scale. (E) Day-wise activity profiles for all *identifiers* can be plotted averaged across individuals. (F) Activity profiles of each individual averaged across days can also be visualized separately. (G) A separate tab, "Individual actograms" can help visualize and inspect each actogram individually for the purpose of choosing representative actograms along with their periodograms. (H) Average activity profile of different *identifiers* can be plotted first averaged across days for each individual, then averaged across individuals. The top panel is plotted in 15 minutes bin and the bottom panel is plotted in 30 minutes bin. This binning can be changed quickly by changing the values of the "Summary time window in minutes" parameter in the "Data input" tab.

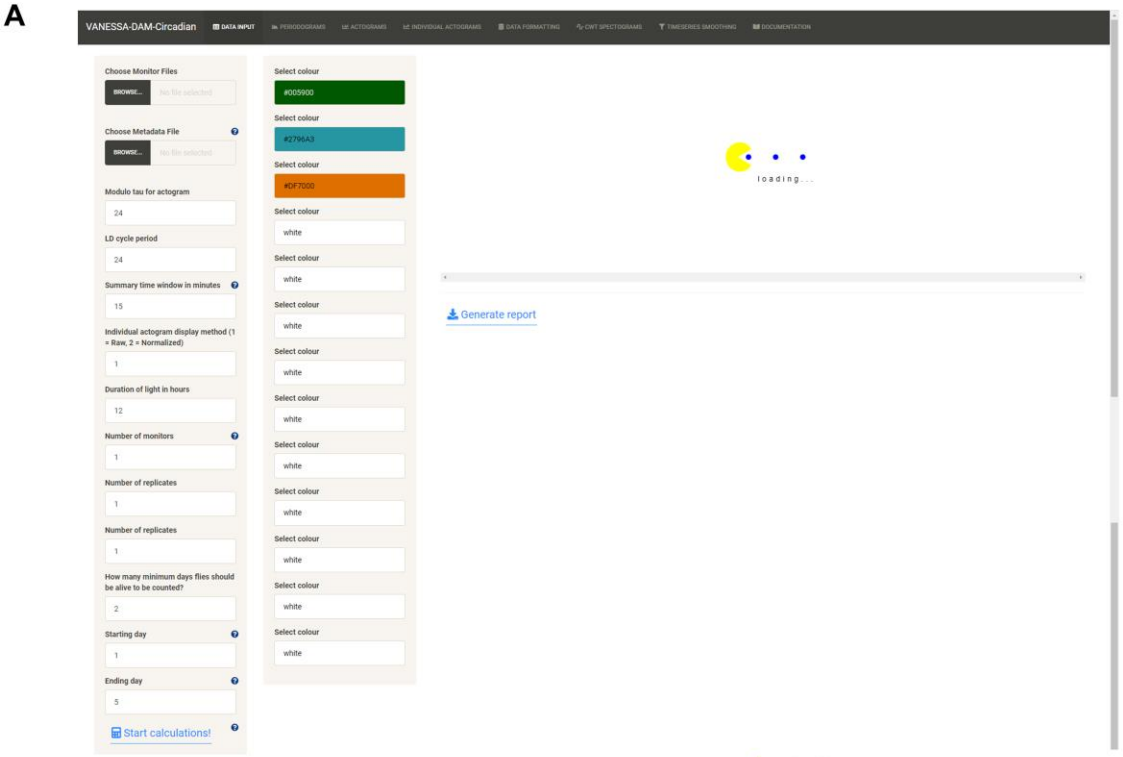
Figure 3: Somnograms and sleep fraction analyses in VANESSA-DAM-SA. (A) Somnogram of all individuals across different days selected can be visualized where clearest white color depicts zero sleep and darkest blue denotes maximum sleep. (B) Somnogram of all individuals averaged across selected days can be visualized where clearest white color depicts zero sleep and darkest blue denotes maximum sleep. (C) Day-wise sleep profiles of each individual from different *identifiers* can be plotted separately. (D) Sleep profiles of each individual averaged across selected days can be plotted. (E) Day-wise sleep profiles of each *identifier* averaged across all individuals. (F) Sleep profiles of each *identifier* first averaged across individuals then averaged across selected

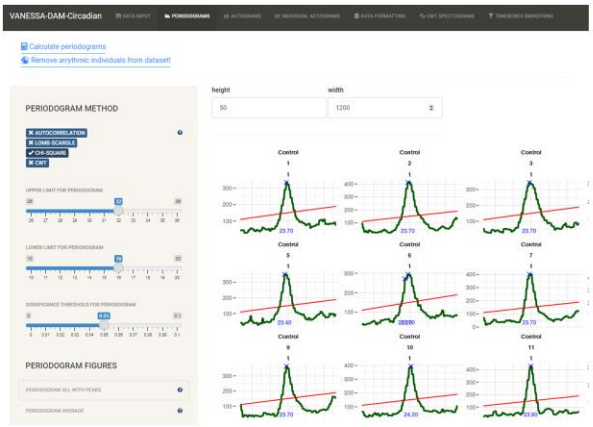
days. (G) Sleep profiles of each *identifier* first averaged across individuals then averaged across selected days in polar scale. (H) Fraction of time sleeping for each individual during the entire day for different *identifiers* as a violin plot for each day separately. (I) Fraction of time sleeping in whole day (*sleep_fraction_all*), only in the light part of the day (*sleep_fraction_l*) and only in the dark part of the day (*sleep_fraction_d*) of each individual of different *identifiers* as a violin plot for each day separately.

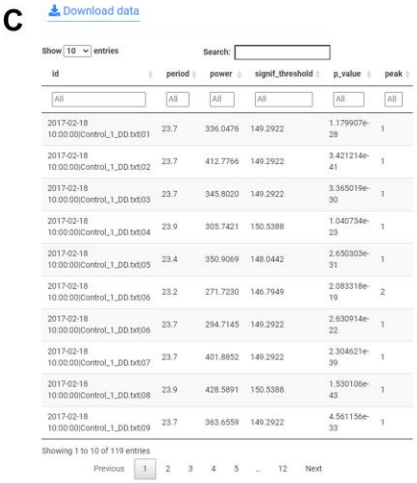
Figure 4: Sleep time, bout and latency analyses in VANESSA-DAM-SA. (A) Total time sleeping in the light part of the day (Light) and in the dark part of the day (Dark) of each individual of different *identifiers* as a violin plot for each day separately. (B) Activity index (activity count per waking minute) of each individual of different *identifiers* as a violin plot for each day separately. (C) Violin plot of total number of bouts of all individuals of each *identifier* for each day separately. (D) Starting time of bout and bout lengths in minutes are plotted averaged across all individuals of each *identifier*. (E) Violin plot of mean bout length in minutes of all individuals of each *identifier* for each day separately. (F) Violin plot of number of bouts in dark (D) and light (L) parts of the day of all individuals of each *identifier* for each day separately. (G) Density plot of mean bout lengths in minutes for each *identifier* and mean bout lengths for each individual plotted along the *x*-axis for each day separately. (H) Violin plot of latency to first bout in minutes of all individuals of each *identifier* for each day separately. (I) Violin plot of latency to first bout in minutes in dark (D) and light (L) parts of the day of all individuals of each *identifier* for each day separately. (J) Plot of average bout duration in minutes vs number of bouts for all individuals colored by *identifiers*. (K) Violin plot of total number of awake bouts of all individuals of each *identifier* for each day separately. *Note: Latency calculated as the time taken to the first sleep bout from ZT00.*

Figure 5: Examples of downloadable sleep analysis data and reproducible code reports from VANESSA-DAM. (A) Example of the tabulated sleep analysis data that can be downloaded from VANESSA-DAM-SA. (B) Snippets of reproducible code report with all used parameters for analysis to reproduce exact same analysis and plots from VANESSA-DAM-CRA. (C) Snippets of reproducible code report with all used parameters for analysis to reproduce exact same analysis and plots from VANESSA-DAM-SA.

Figure 1


A
VANESSA-DAM-Circadian
DATA INPUT
PERIODOGRAMS
ACTOGRAMS
INDIVIDUAL ACTOGRAMS
DATA FORMATTING
CWT SPECTOGRAMS
TIMESERIES SMOOTHING
DOCUMENTATION
Choose Monitor Files
BROWSE...
No file selected
Choose Metadata File
Modulo tau for actogram
24
LD cycle period
24
Summary time window in minutes
15
Individual actogram display method (1 = Raw, 2 = Normalized)
1
Duration of light in hours
12
Number of monitors
1
Number of replicates
1
Number of replicates
1
How many minimum days flies should be alive to be counted?
2
Starting day
1
Ending day
5
Start calculations!
Select colour
#005900
Select colour
#2796A3
Select colour
#DF7000
Select colour
white
loading...
Generate report


B
VANESSA-DAM-Circadian
Calculate periodograms
Remove arrythmic individuals from dataset!
PERIODOGRAM METHOD
AUTOCORRELATION
LOMB-SCARGLE
CHI-SQUARE
CWT
PERIODOGRAM FIGURES
height
width
50
1200


C
Download data
Show 10 entries
Search:
id
period
power
signif_threshold
p_value
peak
2017-02-18 10:00:00|Control_1_DD.txt|01 23.7 336.0476 149.2922 1.179907e-28 1
2017-02-18 10:00:00|Control_1_DD.txt|02 23.7 412.7766 149.2922 3.421214e-41 1
2017-02-18 10:00:00|Control_1_DD.txt|03 23.7 345.8020 149.2922 3.365019e-30 1
2017-02-18 10:00:00|Control_1_DD.txt|04 23.9 305.7421 150.5388 1.040734e-23 1
2017-02-18 10:00:00|Control_1_DD.txt|05 23.4 350.9069 148.0442 2.650303e-31 1
2017-02-18 10:00:00|Control_1_DD.txt|06 23.2 271.7230 146.7949 2.083318e-19 2
2017-02-18 10:00:00|Control_1_DD.txt|06 23.7 294.7145 149.2922 2.630914e-22 1
2017-02-18 10:00:00|Control_1_DD.txt|07 23.7 401.8852 149.2922 2.304621e-39 1
2017-02-18 10:00:00|Control_1_DD.txt|08 23.9 428.5891 150.5388 1.530106e-43 1
2017-02-18 10:00:00|Control_1_DD.txt|09 23.7 363.6559 149.2922 4.561156e-33 1
Showing 1 to 10 of 119 entries
Previous 1 2 3 4 5 ... 12 Next

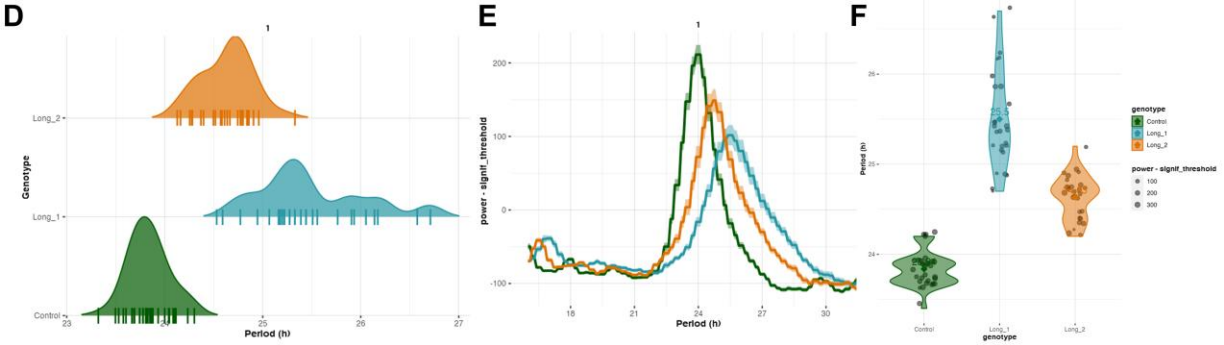
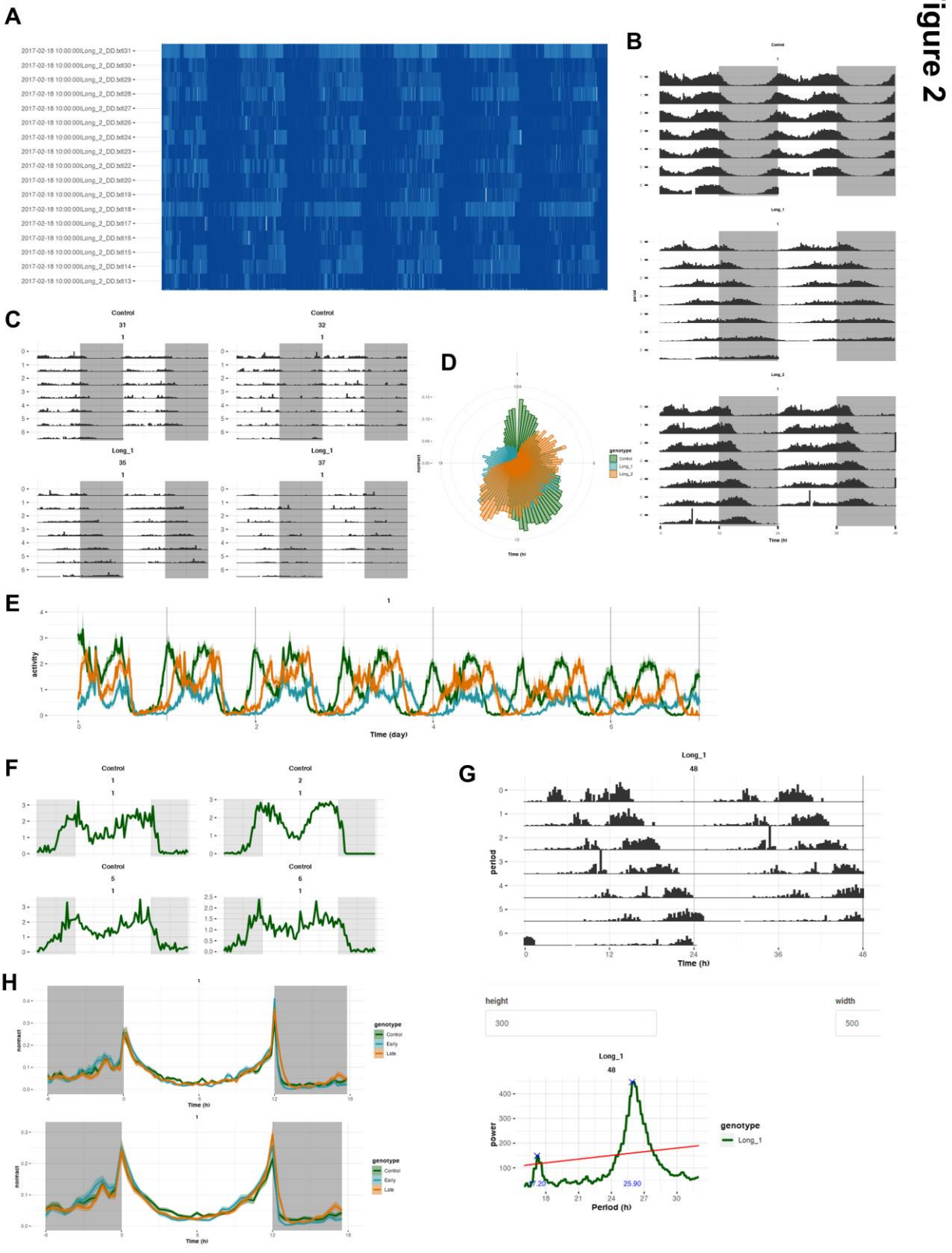

D
Genotype
Long_2
Long_1
Control
Period (h)
E
power - signif_threshold
Period (h)
F
Period (h)
genotype
Control
Long_1
Long_2
power - signif_threshold

Figure 2


A
B
C
D
E
F
G
H
Control
Long_1
Time (h)
Time (day)
activity
period
power
genotype
Control
Early
Late
Long_1
height
300
width
500
Period (h)
25.90

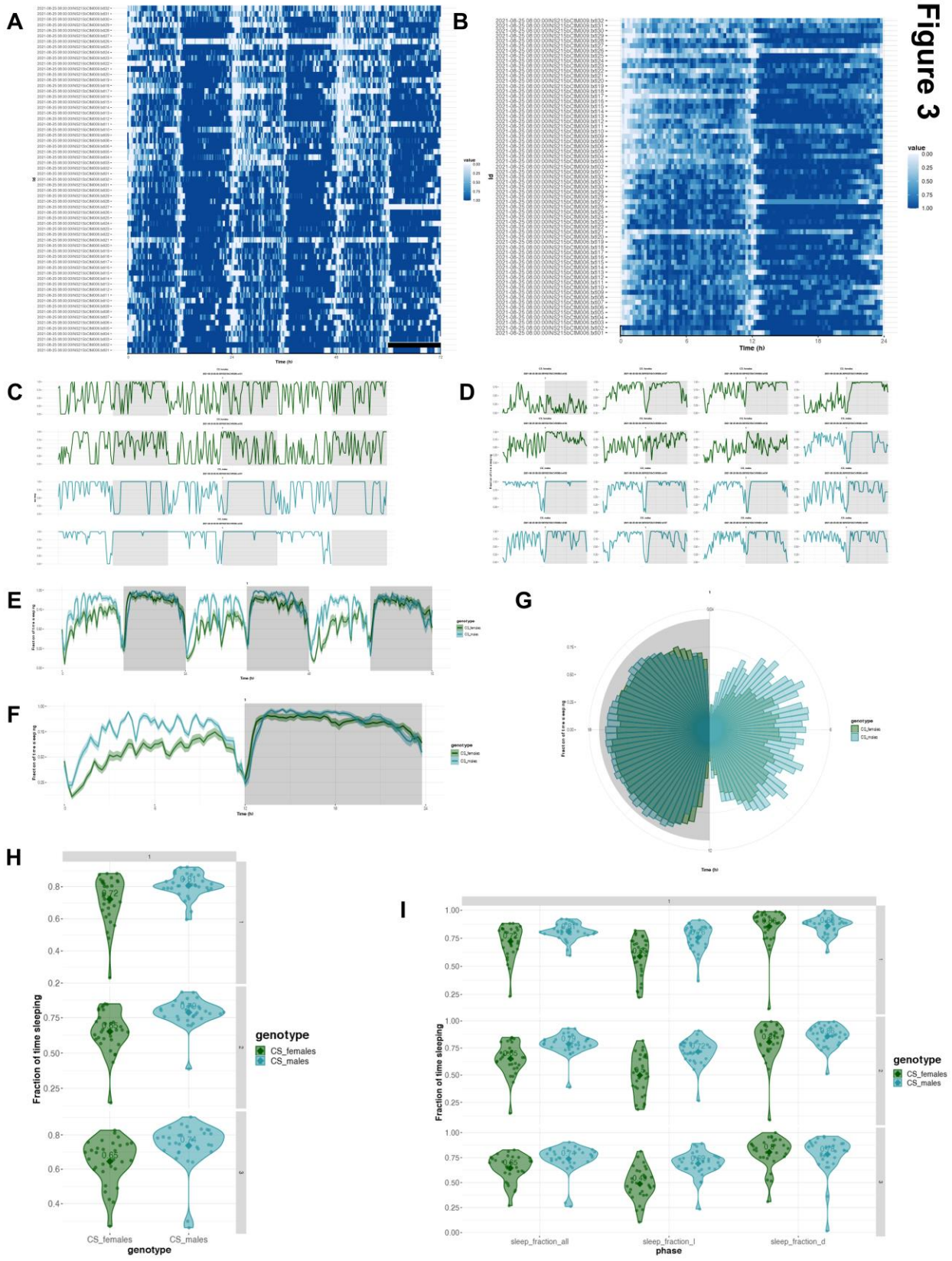

A
B
C
D
E
F
G
H
I
Time (h)
value
genotype
CS_females
CS_males
Fraction of time sleeping
phase
sleep_fraction_all
sleep_fraction_l
sleep_fraction_d

Figure 3

Figure 4

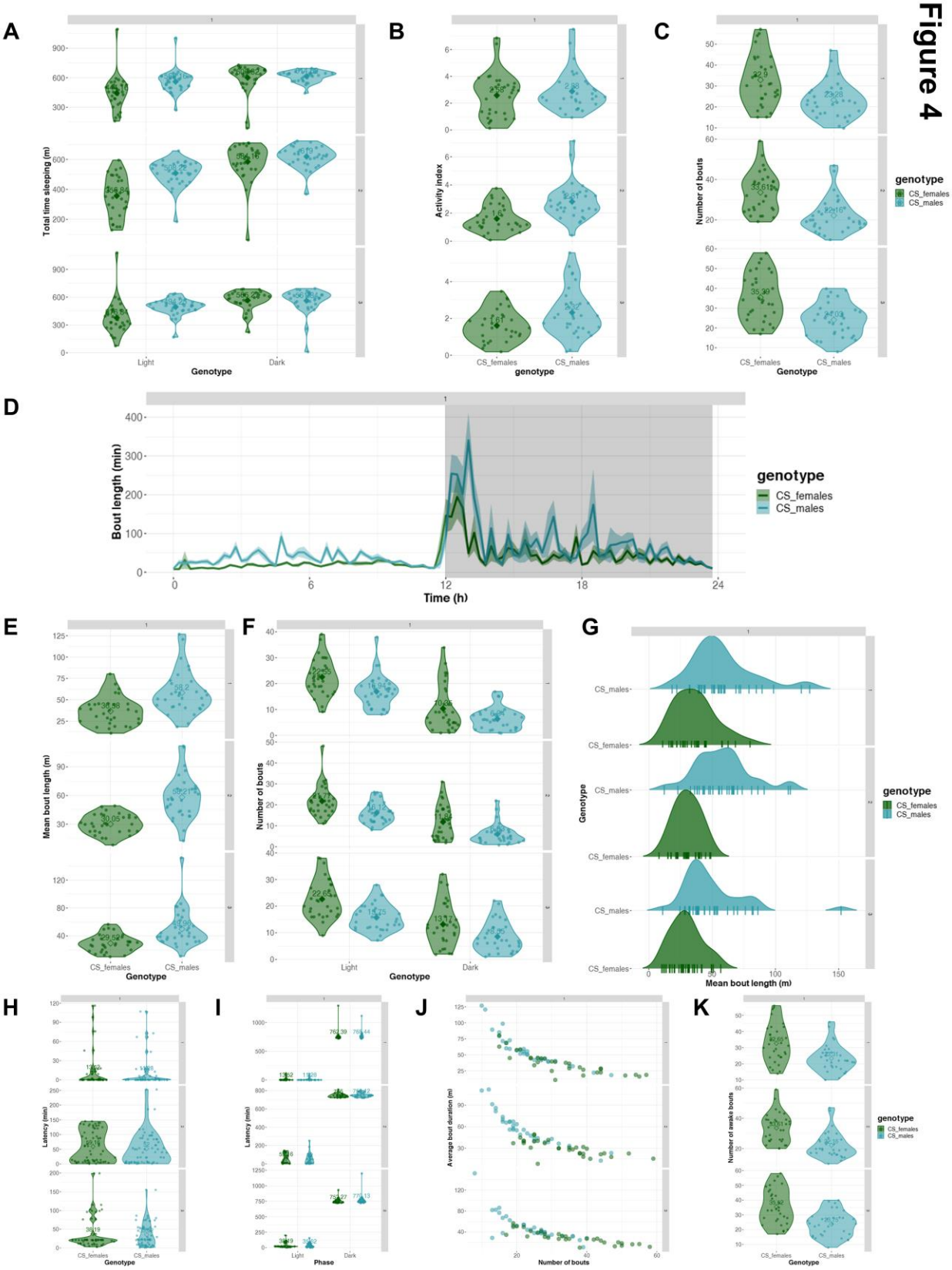

A
Total time sleeping (m)
Light
Dark
Genotype
B
Activity index
CS_females
CS_males
genotype
C
Number of bouts
CS_females
CS_males
Genotype
genotype
CS_females
CS_males
D
Bout length (min)
Time (h)
genotype
CS_females
CS_males
E
Mean bout length (m)
CS_females
CS_males
Genotype
F
Number of bouts
Light
Dark
Genotype
G
CS_males
CS_females
Genotype
Mean bout length (m)
genotype
CS_females
CS_males
H
Latency (min)
CS_females
CS_males
Genotype
I
Latency (min)
Light
Dark
Phase
J
Average bout duration (m)
Number of bouts
K
Number of awake bouts
CS_females
CS_males
Genotype
genotype
CS_females
CS_males

Figure 5

A

Download data Download data bout details

Show 10 entries Search:

id	file_info	start_datetime	stop_datetime	region_id	genotype	replicate	uid	sleep_fraction_all	sleep_fraction_l	sleep_fraction_d	latency	first_bo
All	All	All	All	All	All	All	A	All	All	All	All	All
2020-07-26 10:00:00|Monitor1.txt|01	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	1	Early	1	1	0.799028115237765	0.687716863289382	0.910416666666667	1440	4
2020-07-26 10:00:00|Monitor1.txt|02	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	2	Early	1	2	0.720897116411534	0.530555555555556	0.921110299488678	1608	5
2020-07-26 10:00:00|Monitor1.txt|03	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	3	Early	1	3	0.670600485942381	0.659264399722415	0.681944444444444	1440	1
2020-07-26 10:00:00|Monitor1.txt|04	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	4	Early	1	4	0.716417910447761	0.643997224149896	0.788888888888889	1533	12
2020-07-26 10:00:00|Monitor1.txt|06	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	6	Early	1	6	0.83963901423117	0.756419153365718	0.922916666666667	1440	3
2020-07-26 10:00:00|Monitor1.txt|07	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	7	Early	1	7	0.462339465463381	0.333795975017349	0.590972222222222	1615	6
2020-07-26 10:00:00|Monitor1.txt|08	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	8	Early	1	8	0.716417910447761	0.569049271339348	0.863888888888889	1538	6
2020-07-26 10:00:00|Monitor1.txt|09	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	9	Early	1	9	0.78132592849705	0.61970853573907	0.943055555555556	1440	4
2020-07-26 10:00:00|Monitor1.txt|10	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	10	Early	1	10	0.588337382853176	0.496877168632894	0.679861111111111	1442	9
2020-07-26 10:00:00|Monitor1.txt|12	[object Object]	2020-07-26 10:00:00	2020-07-30 10:00:00	12	Early	1	12	0.71468240194377	0.591950034698126	0.8375	1440	3

Showing 1 to 10 of 92 entries

B

Reproducible code report - VANESSA-DAM-CRA

Produced with VANESSA-DAM on 2021-12-10

All parameters used in the shiny app for analysis.

Metadata file name

```
params$meta
```

```
##            name size                type
## 1 Metadata.csv 6595 application/vnd.ms-excel
##                                                                datapath
## 1 C:\\Users\\ariji\\AppData\\Local\\Temp\\RtmpGPpbKi/a05bc879544e02c92760ca60/0.csv
```

Working directory (Use the directory where your metadata file and monitor files are saved)

```
params$wd
```

```
## [1] "D:/Work/VANESSA-DAM_github/VANESSA-DAM-main_1.3-fordeployment_github/VANESSA-DAM-CRA"
```

Modulo-tau value for actograms in hours (generally used value = 24)

```
params$modtau
```

```
## [1] 24
```

The total daylength in your experiment in hours (light + dark phase total, generally used value = 24)

```
params$ldperiod
```

```
## [1] 24
```

Summary time window in minutes (used in average profiles)

```
params$min
```

```
## [1] 15
```

Set working directory of R to where your Metadata file and Monitor files are (in this case they have to be in the same directory). Replace all params$XXXX with the values mentioned above (values you have used in the shiny app).

Load all needed libraries:

```
library(data.table)
library(ggetho)
library(zeitgebr)
library(ggplot2)
library(ggridges)
library(readr)
library(damr)
library(sleepr)
library(behavr)
library(dplyr)
```

Do data pre-processing:

```
setwd(params$wd)
metadata <- fread("params$meta")
    metadata <- na.omit(metadata)
    metadata_proc <- link_dam_metadata(metadata, result_dir = params$wd)
    dt <- load_dam(metadata_proc)
        dt[, moving := activity > 0]
        dt_curated <- curate_dead_animals(dt)
        lifespan_dt <- dt_curated[, .(lifespan = max(t)), by = id]
        valid_ids <- lifespan_dt[lifespan > days(params$remove), id]
        dt_curated <- dt_curated[id %in% valid_ids]
        dt_curated <- dt_curated[t %between% c(days(params$start), days(params$end))]
        dt_curated <- dt_curated[, Day := (t / days(1)), by = id]
        dt_curated$Day <- ceiling(dt_curated$Day)
        dt_curated[, 5][dt_curated[, 5] == params$start] <- params$start + 1
        setkey(dt_curated, Day, id)
        dt_curated[, nomact := (activity / sum(activity)) * 100, by = .(id, Day)]
        setkey(dt_curated, id)
        setbehavr(dt_curated, metadata_proc)
        dt_curated[, uid := 1:.N, meta = T]
        dt_curated[, .(id, uid), meta = T]
```

Theme configuration for figures:

C

Reproducible code report - VANESSA-DAM-SA

Produced with VANESSA-DAM on 2021-12-10

All parameters used in the shiny app for analysis.

Metadata file name

```
params$meta
```

```
##            name size                type
## 1 Metadata.csv 4718 application/vnd.ms-excel
##                                                                datapath
## 1 C:\\Users\\ariji\\AppData\\Local\\Temp\\RtmpGPpbKi/f2f2877986iaf2a00e437321/0.csv
```

Working directory (Use the directory where your metadata file and monitor files are saved)

```
params$wd
```

```
## [1] "D:/Work/VANESSA-DAM_github/VANESSA-DAM-main_1.3-fordeployment_github/VANESSA-DAM-SA"
```

Modulo-tau value for actograms in hours (generally used value = 24)

```
params$modtau
```

```
## [1] 24
```

The total daylength in your experiment in hours (light + dark phase total, generally used value = 24)

```
params$ldperiod
```

```
## [1] 24
```

Summary time window in minutes (used in average profiles)

```
params$min
```

```
## [1] 15
```

Set working directory of R to where your Metadata file and Monitor files are (in this case they have to be in the same directory). Replace all params$XXXX with the values mentioned above (values you have used in the shiny app).

Load all needed libraries:

```
library(data.table)
library(ggetho)
library(zeitgebr)
library(ggplot2)
library(ggridges)
library(readr)
library(damr)
library(sleepr)
library(behavr)
library(dplyr)
```

Do data pre-processing:

```
setwd(params$wd)
metadata <- fread("params$meta")
    metadata <- na.omit(metadata)
    metadata_proc <- link_dam_metadata(metadata, result_dir = params$wd)
    dt <- load_dam(metadata_proc, FUN = sleepr::sleep_dam_annotation)
        dt[, moving := activity > 0]
        dt_curated <- curate_dead_animals(dt)
        lifespan_dt <- dt_curated[, .(lifespan = max(t)), by = id]
        valid_ids <- lifespan_dt[lifespan > days(params$remove), id]
        dt_curated <- dt_curated[id %in% valid_ids]
        dt_curated <- dt_curated[t %between% c(days(params$start), days(params$end))]
        setbehavr(dt_curated, metadata_proc)
        dt_curated[, uid := 1:.N, meta = T]
        dt_curated[, .(id, uid), meta = T]
        summary_dt <- rejoin(dt_curated[, .(sleep_fraction = mean(asleep)), by = id])
        dt_curated[, phase := ifelse(t %% hours(params$ldperiod) < hours(params$light), "L", "D")]
        summary_dt <- rejoin(dt_curated[, .(
          sleep_fraction_all = mean(asleep),
          sleep_fraction_l = mean(asleep[phase == "L"]),
          sleep_fraction_d = mean(asleep[phase == "D"])
        ), by = id])
        summary_dt_melted <- melt(summary_dt,
          measure.vars = patterns("^sleep_fraction_"),
```

References:

Abhilash L and Sheeba V (2019) RhythmicAlly: Your R and Shiny–Based Open-Source Ally for the Analysis of Biological Rhythms. J Biol Rhythms 34: 551–561.

Cichewicz K and Hirsh J (2018) ShinyR-DAM: a program analyzing Drosophila activity, sleep and circadian rhythms. Communications Biology 1.

ClockLab | Actimetrics. Available at: https://actimetrics.com/products/clocklab/ (accessed 12 August 2021).

el temps principal. Available at: http://www.el-temps.com/principal.html (accessed 12 August 2021).

Geissmann Q, Rodriguez LG, Beckwith EJ, et al. (2019) Rethomics: An R framework to analyse high-throughput behavioural data. PLoS ONE 14.

Gilestro GF and Cirelli C (2009) PySolo: A complete suite for sleep analysis in Drosophila. Bioinformatics 25: 1466–1467.

Hendricks JC, Finn SM, Panckeri KA, et al. (2000) Rest in Drosophila Is a Sleep-like State. Neuron 25: 129–138.

R Core Team (2021) R core team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www. R-project. org.

Schmid B, Helfrich-Förster C, Yoshii T, et al. (2011) A New ImageJ Plug-in 'ActogramJ' for Chronobiological Analyses. J Biol Rhythms (26): 464–467.

Shaw PJ, Cirelli C, Greenspan RJ, et al. (2000) Correlates of Sleep and Waking in Drosophila melanogaster. Science 287: 1834–1837.

Winston C, Cheng J, JJ A, et al. (2019) shiny: Web Application Framework for R. R package version 1.4.0.

Supplementary online material for -

VANESSA – Shiny apps for accelerated time-series analysis and visualization of *Drosophila* circadian rhythm and sleep data

Arijit Ghosh, Vasu Sheeba*

Chronobiology and Behavioral Neurogenetics Laboratory, Neuroscience Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore – 560064, Karnataka, India

Running title: *VANESSA – circadian rhythm and sleep analysis apps*

*Correspondence: Vasu Sheeba, Chronobiology and Behavioral Neurogenetics Laboratories, Neuroscience Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bengaluru - 560064, Karnataka, India. Phone: +91-080-2208-2987. Email: sheeba@jncasr.ac.in

A

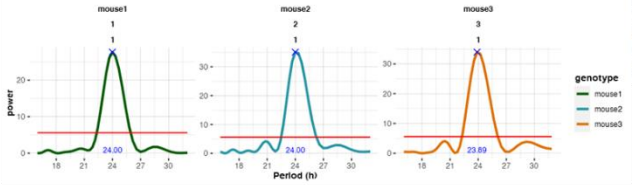
```r
rm(list = ls())
library(hms)
library(lubridate)
library(tidyverse)
library(tidyr)
setwd("C:/Users/ariji/Desktop/OSM_data")
bin = 5   ############binning in minutes
df_raw <- read.delim("T-Timer 100 M_Counts.csv", sep = ",", header = F)
df_raw_cut <- df_raw[-c(1:4),-6]
df_raw_cut <- df_raw_cut[!apply(df_raw_cut == "", 1, any),]

# df1 <- data.frame()
df1 <- matrix(0, nrow = nrow(df_raw_cut), ncol = 42)
df1[,1] <- seq(1, (nrow(df_raw_cut)), by = 1)
df1[,11] <- df_raw_cut[,4]
date_time_start <- dmy_hm(paste(df_raw[3,1], df_raw[3,2]))
# total_time_min = nrow(df_raw_cut) * bin
date_time_end <- date_time_start + days(as.numeric((df_raw_cut[nrow(df_raw_cut),1]))-1) +
hours(df_raw_cut[nrow(df_raw_cut),2]) +
  minutes(df_raw_cut[nrow(df_raw_cut),3])
ag <- data.frame(timestamp = seq(as.POSIXct(date_time_start), as.POSIXct(date_time_end), by = bin*60))
# df1[,2] <- seq(as.POSIXct(date_time_start), as.POSIXct(date_time_end), by = (bin*60))
df1 <- cbind(df1, ag)
df1[,2] <- (df1$timestamp)
df1 <- df1[,-43]
# df1[,2] <- as.character(df1[,2])
# df1[,2] <- as_datetime(df1[,2])
df1 <- df1 %>% separate(2, c("a", "b"), sep = " ")
df1 <- df1[,-4]
df1$a <- ymd(df1$a)
df1$a <- format(df1$a, format = "%d %b %y")
# df1 <- na.omit(df1)
df1 <- df1[complete.cases(df1),]
df1[,4] <- 1
df1[,8] <- "Ct"
# df1[,c(12,42)] <- 0
write.table(df1, "testmouse1.txt", col.names = F, row.names = F, sep = "\t", append = T, quote = F)
```

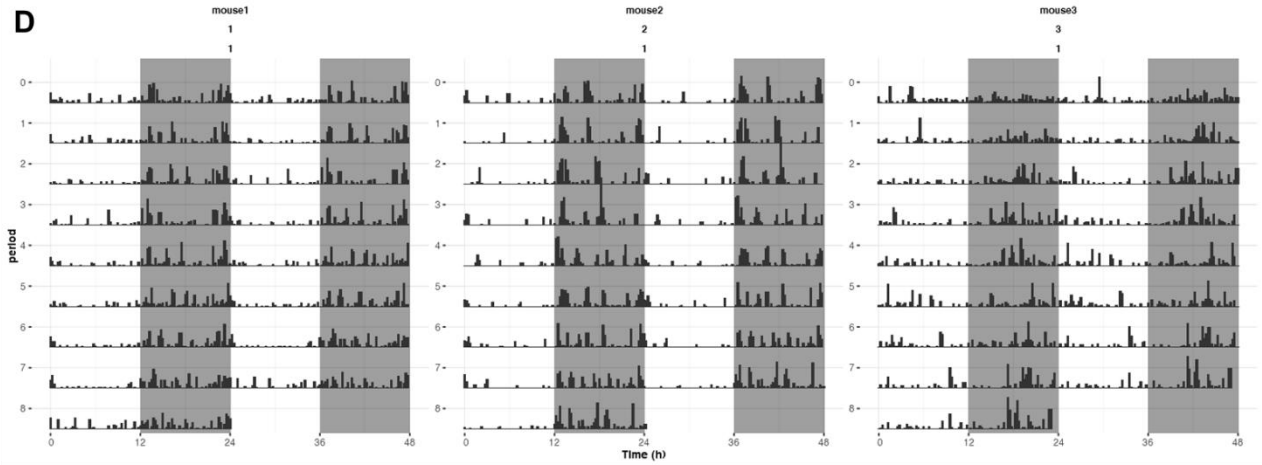
B

	A	B	C	D	E
1	ClockLab				
2	T-Timer 100 Motion				
3	29-Aug-16	0:00			
4	Day	Hr	Min	Cnts/min	Lights
5	1	0	0	NaN	NaN
6	1	0	5	NaN	NaN
7	1	0	10	NaN	NaN
8	1	0	15	NaN	NaN
9	1	0	20	NaN	NaN
10	1	0	25	NaN	NaN
11	1	0	30	NaN	NaN
12	1	0	35	NaN	NaN
13	1	0	40	NaN	NaN
14	1	0	45	NaN	NaN
15	1	0	50	NaN	NaN
16	1	0	55	NaN	NaN
17	1	1	0	NaN	NaN
18	1	1	5	NaN	NaN

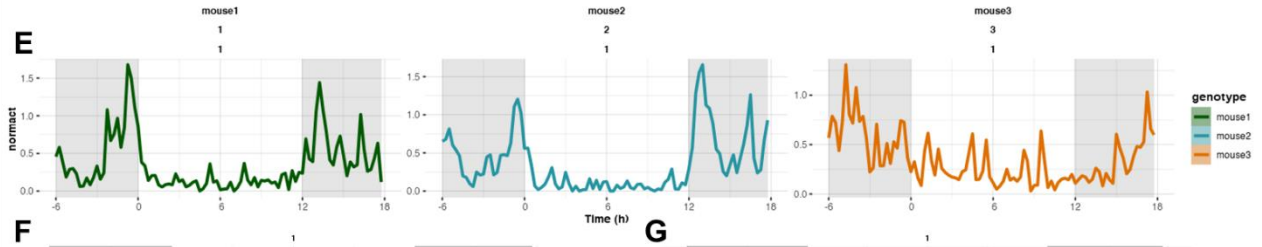
C



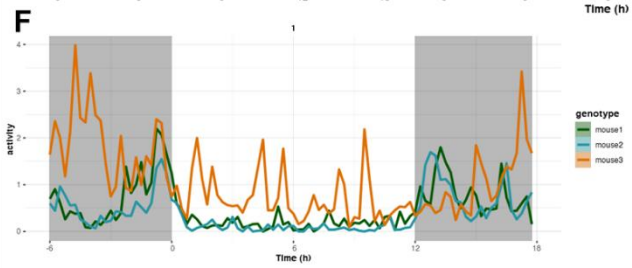
D



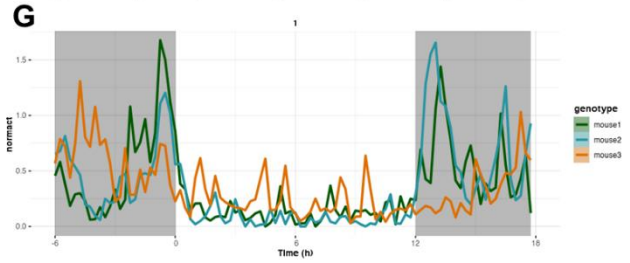
E



F



G



Supplementary Figure 1: Example of Mouse circadian rhythm analysis recorded with Trikinetics ClockLab software in VANESSA-DAM. (A) Example of a few lines of code that can convert mouse activity data acquired through ClockLab software into VANESSA compatible DAM files. This code is available as a function from the vanessadamcra package. (B) Example of mouse activity data acquired through ClockLab software which is converted to VANESSA compatible DAM files. (C) Lomb-Scargle periodograms of three mice in LD12:12 for 9 days. (D) Normalized actograms of these three mice. (E) Average normalized activity profiles of three mice. (F) Average activity profiles of three mice plotted together for comparison. (G) Average normalized activity profiles of three mice plotted together for comparison. All data taken from *Leise et al., 2018* (https://osf.io/seyhp/ - Harrington Lab Motion Data).

Supplementary Methods 1: Easy tutorial to start using VANESSA-DAM-CRA

Visualization and ANalysis of timE SerieS dAta - *Drosophila* Activity Monitors (VANESSA-DAM) is a collection of useful tools to visualize and analyze Time series data obtained from *Drosophila* Activity Monitors (https://www.trikinetics.com/). The first in the series of tools is a shiny app for circadian rhythm analysis and visualization - VANESSA-DAM for circadian rhythm analysis (VANESSA-DAM-CRA). For any suggestions, questions, troubleshooting or customization, please contact arijitghosh2009@gmail.com.

VANESSA-DAM-CRA

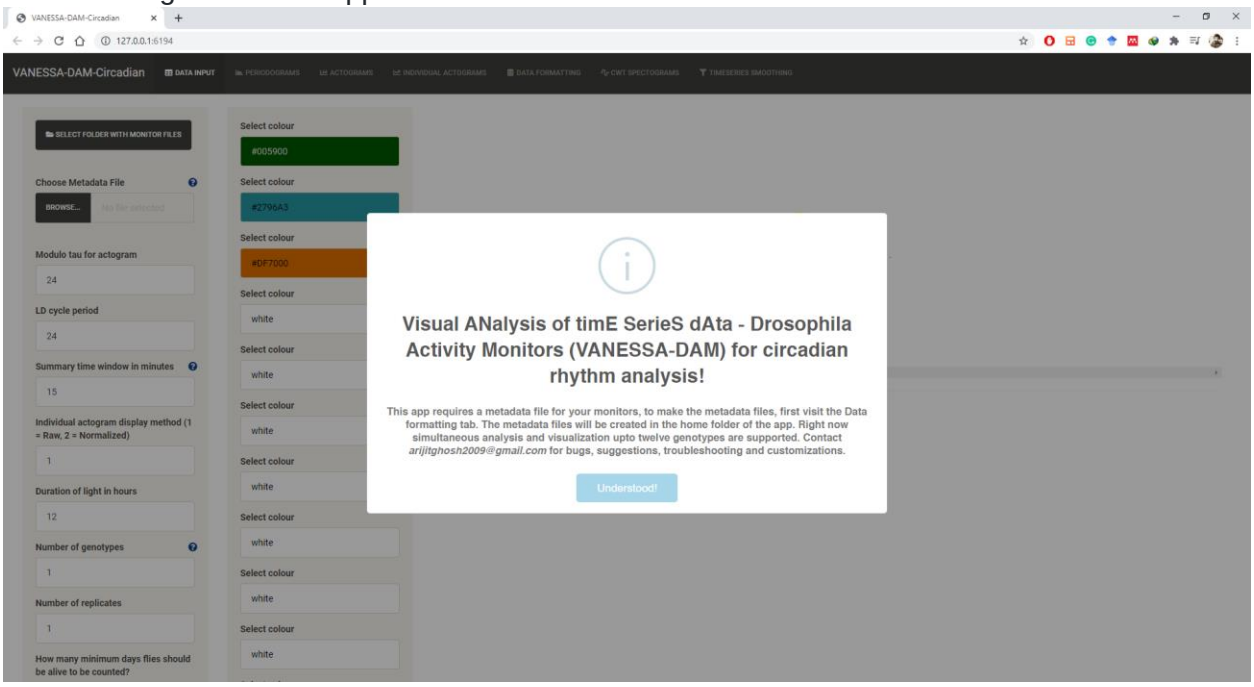
VANESSA-DAM-CRA is dependent on Quentin Geissmann's rethomics family of packages - behavr, damr, ggetho, zeitgebr, for some analysis and visualization options. It offers several advantages over existing tools for circadian rhythm analysis from DAM systems, some mentionable ones are -

1. Analysis and visualization of multiple monitor files, genotypes, replicates together in a high-throughput manner.
2. Automatic period power detection through a range of periodogram methods.
3. Producing high-resolution publication-quality figures with a plethora of customization.
4. Data curation - automatic user-defined parameter-based removal of dead and arrhythmic individuals.
5. Individual wise CWT spectrograms and wavelet power plot.
6. Visual comparison among genotypes, replicates.
7. Timeseries filtering with kernel smoothing and Butterworth filters.
8. Reproducible code report so that you can generate the figures and analysis without the shiny app from RStudio directly.

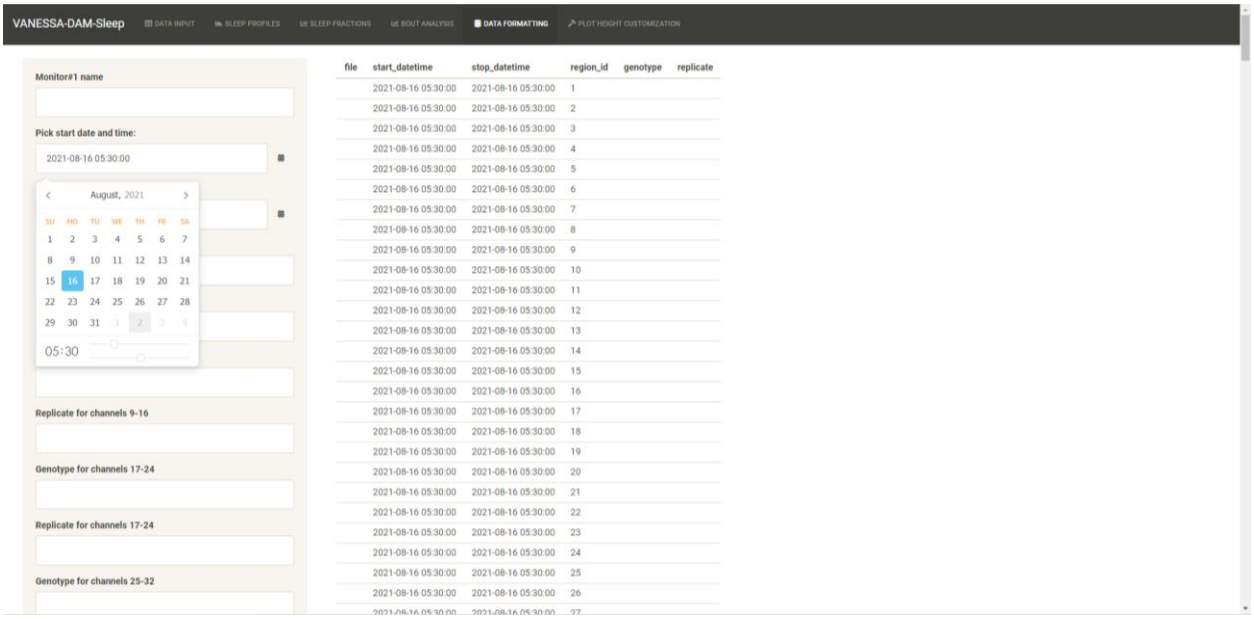
This short tutorial provided (*Easy tutorial to start using VANESSA-DAM-CRA.pdf*) here should be self-explanatory and should help beginners start using the app right away.

Easy tutorial to start using VANESSA-DAM-CRA

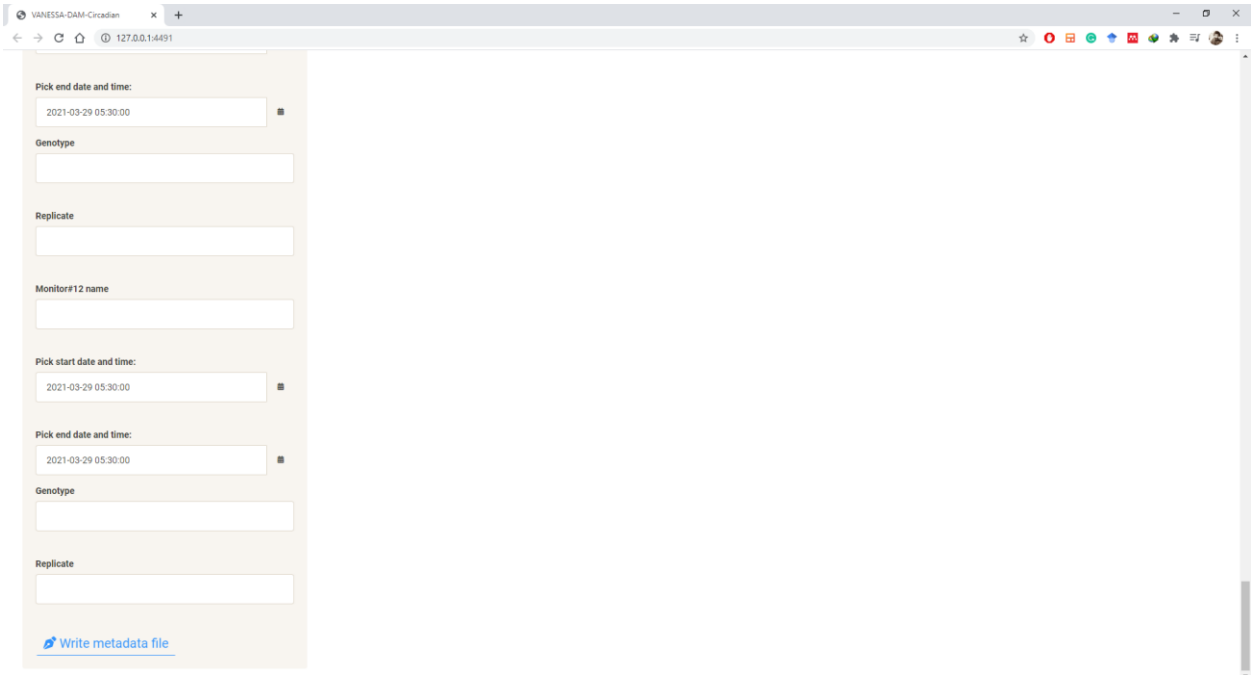
1. Install R (version ≥ 3.6.3) and RStudio (version ≥ 1.2) on your computer.
2. Download the zipped folder containing the app and unzip to a folder.
3. Open RStudio and Run the **startup.r** file to install all necessary packages to run the app.
   Note: If you have trouble installing the *damr* package, the source files are provided. Please go to RStudio > Tools > Install packages > Install from : Package Archive file and select the supplied *damr_0.3.7.tar.gz* file to install the package.
4. Restart R by pressing **ctrl+shift+F10** or from **Session > Restart R** from RStudio menu. This step is useful to clean the global environment so that previously loaded packages do not interrupt, or mask packages/functions needed for VANESSA-DAM-CRA.
5. Set working directory as the home folder of the app (by pressing **ctrl+shift+h** OR by using the `setwd()` command) and then run the app by typing `shiny::runApp(launch.browser = T)` in the console. OR load the **server.r** or **ui.r** file in RStudio by double-clicking on any of them and click on the "**Run App"** button in RStudio panel (for better experience select "*Run External*" from the dropdown menu on the right side of the "**Run App"** button beforehand).
6. The following screen will appear:



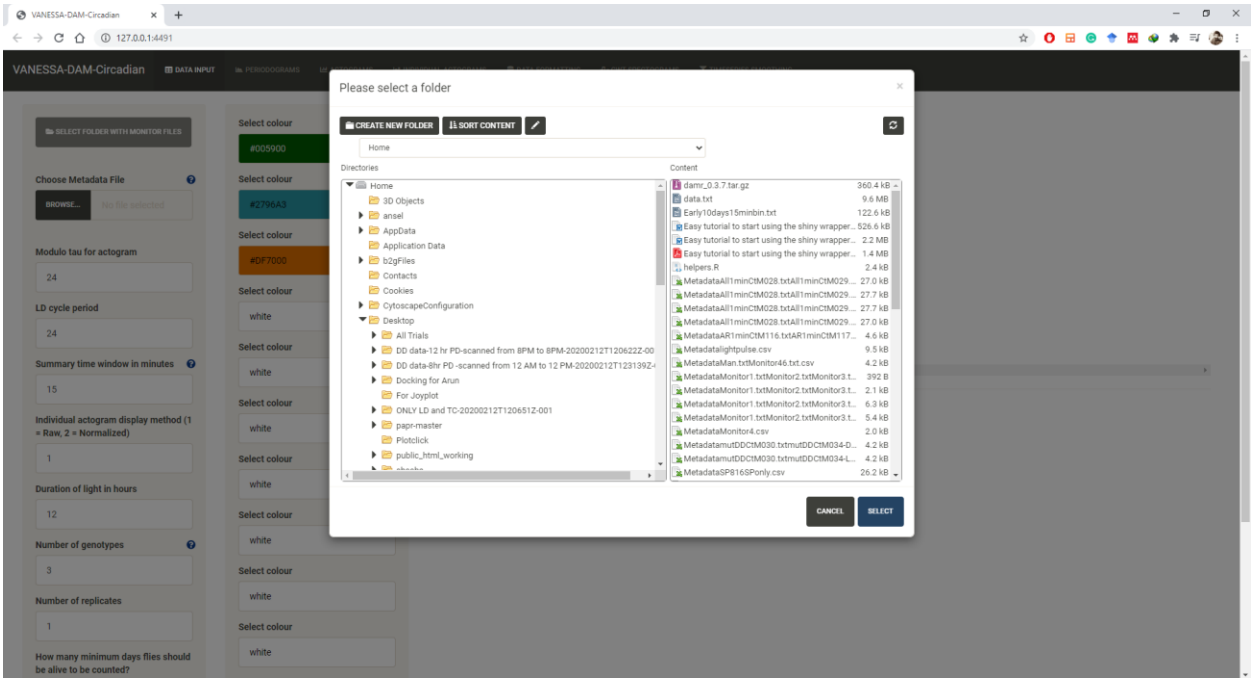
7. Click on **Understood!** button or press **Esc** to begin.
8. Assuming you have your Monitor files (*Control_1_DD.txt, Long_1_DD.txt, Long_2_DD.txt* provided with Genotypes *Control, Long period 1 and Long period 2* respectively – check the ***Notes*** section for details of the data provided) from DAMScan in the home folder of the app, first job is to make a Metadata file (Provided as *Metadata.csv*).
9. In this case, we have 3 genotypes in 3 Monitor files, so enter 3 at the **Number of Genotypes** box. Then go to the **DATA FORMATTING** tab.



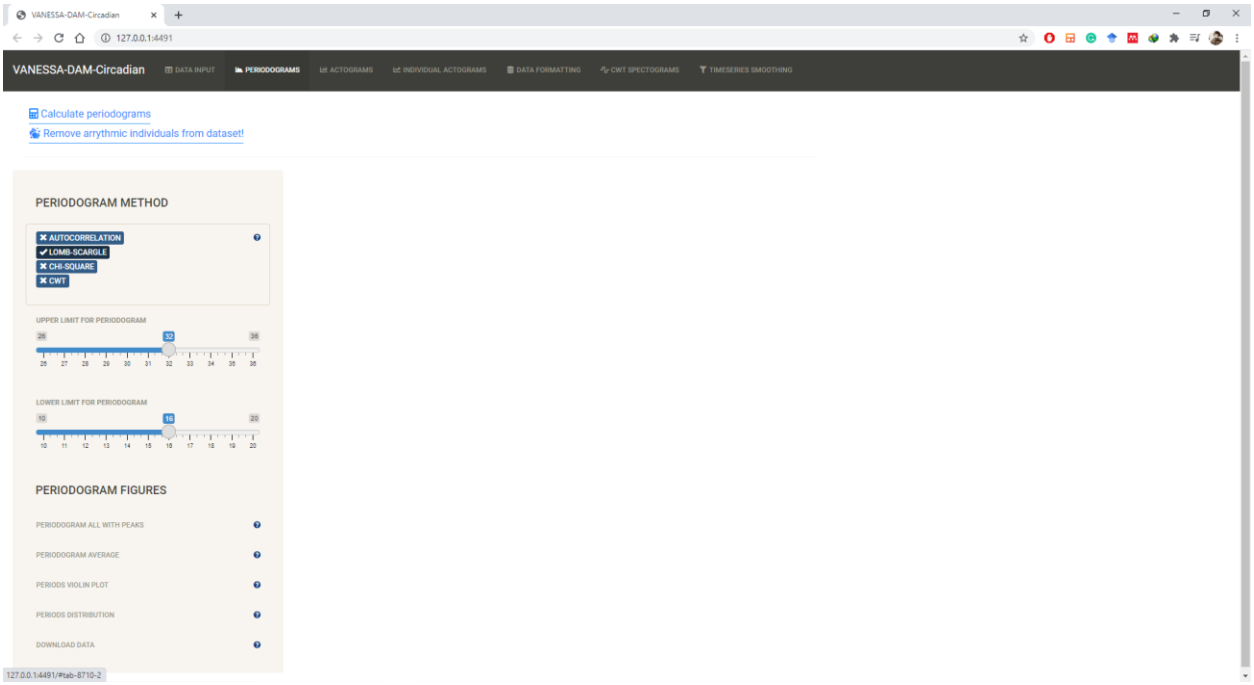
10. Enter name of your first monitor file in the **Monitor#1** name box, include extensions also (e.g., if the name of the monitor file Monitor1.txt, then enter the FULL name, including ".txt"). Fill the next boxes - **Start date and time (If your start date is 1st August 2020 and Lights-on time is 10 AM, then you write 2020-08-01 10:00:00), End date (If your end date is 10th August 2020 – 10 AM, then you write 2020-08-10 10:00:00)** with your experimental details as shown above. Enter **Genotype** values and **Replicate number** (only if your experimental design has multiple replicates for all/some genotypes, else enter any random number for all monitors, it will not affect any calculation) in proper boxes for proper channels (marked as **Genotype for channels** 1-8 and **Replicate for channels** 1-8 etc.). A date and time picker has been added to facilitate easy entering of start and end date times without error in formatting. Each monitor can have maximum of 4 genotypes with 8 individuals.
    Note: Creating the metadata file is the single most important step for using this app, any wrong information in the metadata file will wreak havoc on the results and the app may crash. DO NOT change information in the DATA FORMATTING tab once you have entered. For some unknown problem in shiny reactivity, this section does not handle deletion and change of values nicely. Alternatively, you can just create your metadata file by changing data in the already provided metadata file in a good text editor like Notepad++ (DO NOT USE EXCEL, as excel may change the date-time format). The metadata file serves as a base of information for your analysis too and is good for record keeping of your experiments.



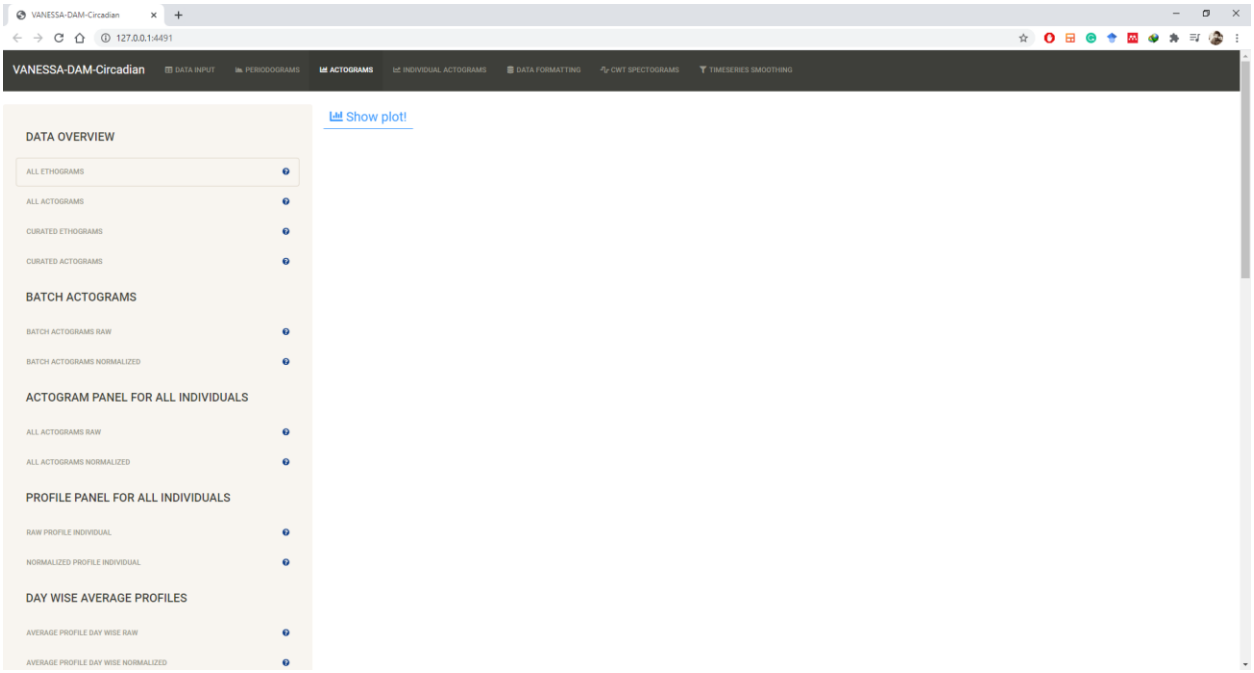
11. After filling in experimental details for all 3 monitors, press the **Update metadata** button, and then go to the bottom of the page and click on the **Download metadata** button, this will download the Metadata as a CSV file in your desired folder and is ready to be used.
12. Go back to the **DATA INPUT** tab. Enter choices one by one- **Duration of light in hours** in your experiment, **How many minimum days flies should be alive to be counted**. Also chose your **starting day** and **ending day** for analysis (**starting day** = 1 means first day in your data, so if you want to select first five days, **starting day** should be 1 and **ending day** should be 5). Enter **LD cycle period** in your experiment, **Modulo tau** for actogram can be changed for visualization purpose later. The same goes for **Summary time window in minutes** and **Individual actogram display method**. You can change any of these parameters later while looking at plots, the plots will immediately update accordingly. For example, if you're looking at your actograms in 15 minutes bin, and you want to visualize them with 5 minutes bin, you just have to change the value in Summary time window in minutes box, similarly LD shading can also be changed. You are all set to start analysis now. **ONLY PROCEED FORWARD WHEN YOU HAVE FIXED ALL ANALYSIS PARAMETERS.** Select the monitor files to be analyzed by clicking the **Choose Monitor Files** button and upload your Metadata file by clicking on the **BROWSE** button in the **Choose Metadata File** box. Press the **Start Calculations!** button, when the calculations are done, you will be notified by a sound. Depending on how long your data is and how many monitors you are analyzing, this step (which involves curation also) will take anything between 5 seconds to 1 minute typically. To follow this tutorial, enter **Number of monitors** as 3 and **Number of replicates** as 1.
    Note: Most of the parameters have a small question mark symbol in blue near them, clicking on the symbol will show a modal with information about the parameter. This help is available throughout the app for different parameters, analyses and plots.



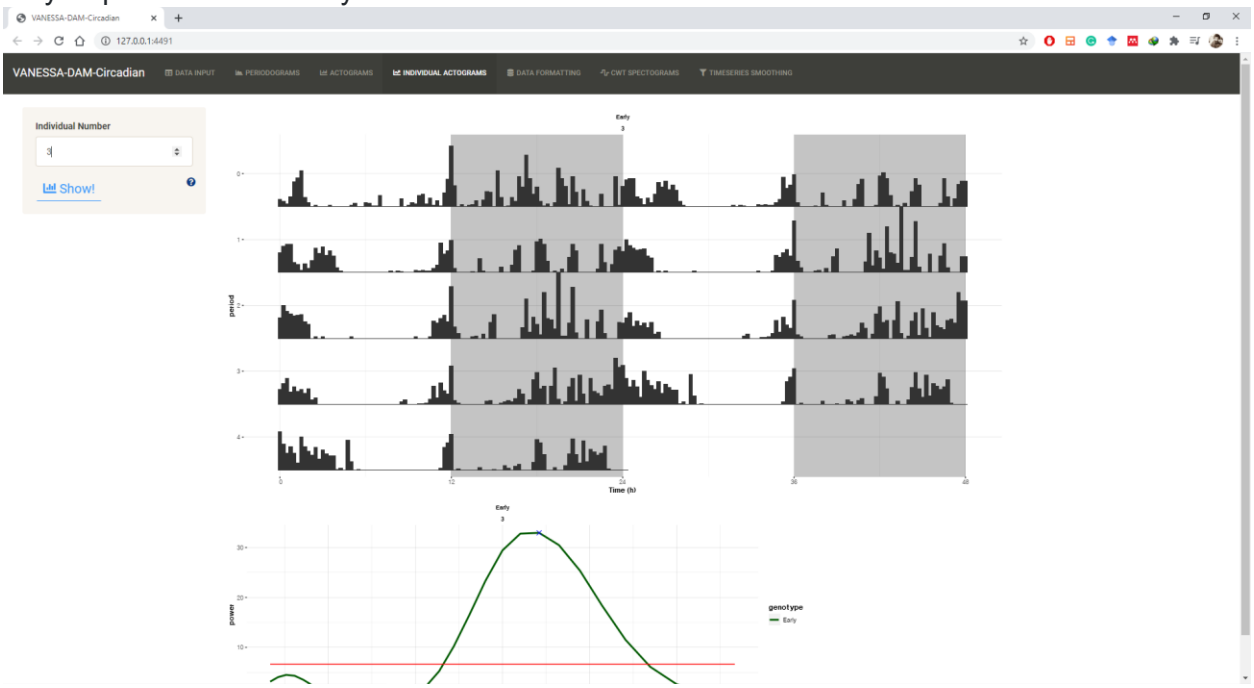
13. Go to the **PERIODOGRAMS** tab next. You will see the screen below. Choose **PERIODOGRAM METHOD** from four available methods and choose **Upper limit for periodogram** and **Lower limit for periodogram**. Click on **Calculate periodogram** button to start calculating period of the individuals with the method you have chosen previously. If you click on the **Remove arrythmic individuals** from dataset! button next, all arrhythmic individuals will be removed from all calculations and plots. **Once you remove the arrhythmic individuals, ALL calculations and plots will be done with only rhythmic individual on curated data.** If you want to visualize the arrhythmic individuals also, do the analysis without clicking on the **Rhtymic arrhythmic individuals** button.



14. When you click on different buttons under **PERIODOGRAM FIGURES** , you will start seeing plots as mentioned in different side panels after clicking on the **Plot** button.
15. Go to the **Download data** tab to download a csv file for all period and power values. Different peaks of periodograms for all individuals are noted in the file. If you open the file in Excel or LibreOffice or any spreadsheet program, you can put a filter on the "peak" column which is the last column of the file. When you filter the data by "peak" and select "peak" as "1", it will show you the highest peak of the periodogram of all rhythmic individuals. You'll notice that this file will now show you these columns - a) id: start_datetime|Monitor name|Indivudual number, b) period: the period defined by the "peak" (if you select 1, it will show you the period of the highest peak in the respective periodogram), c) power: the peak value from the periodogram (in case of autocorrelation it's the peak autocorrelation value, i.e., rhythmicity index, in case of chi-sq, it's the power of the chi-sq periodogram, etc.), d) signif_threshold: significance threshold for the particular peak, e) p_value: the p_value is reported here for the particular peak, f) peak: different peaks in the periodogram (filter based on this).
16. Go to the **ACTOGRAMS** tab to start visualizing Actograms and Profiles as mentioned in different side tabs, feel free to explore.

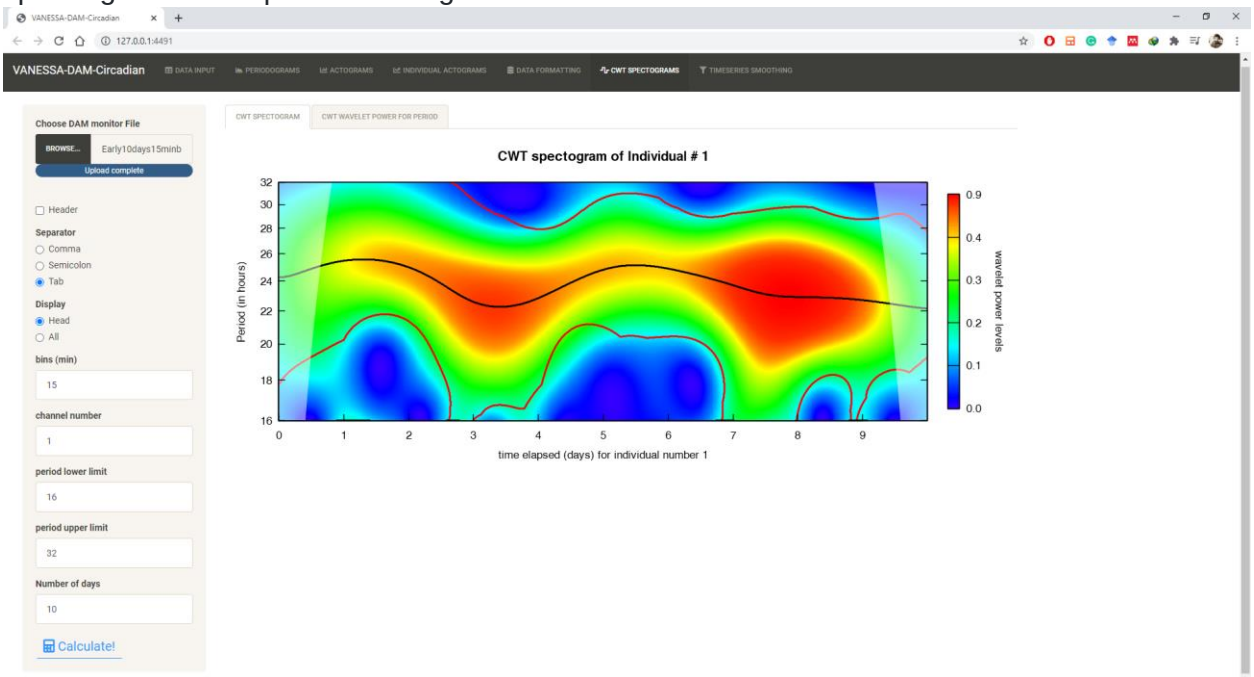


17. Go to the **INDIVIDUAL ACTOGRAMS** tab to see individual wise actograms (raw or normalized) for closer inspection of your data. Change **Individual Number** by pressing up or down arrow or use the buttons in the box to change individuals to look at and they'll update automatically.

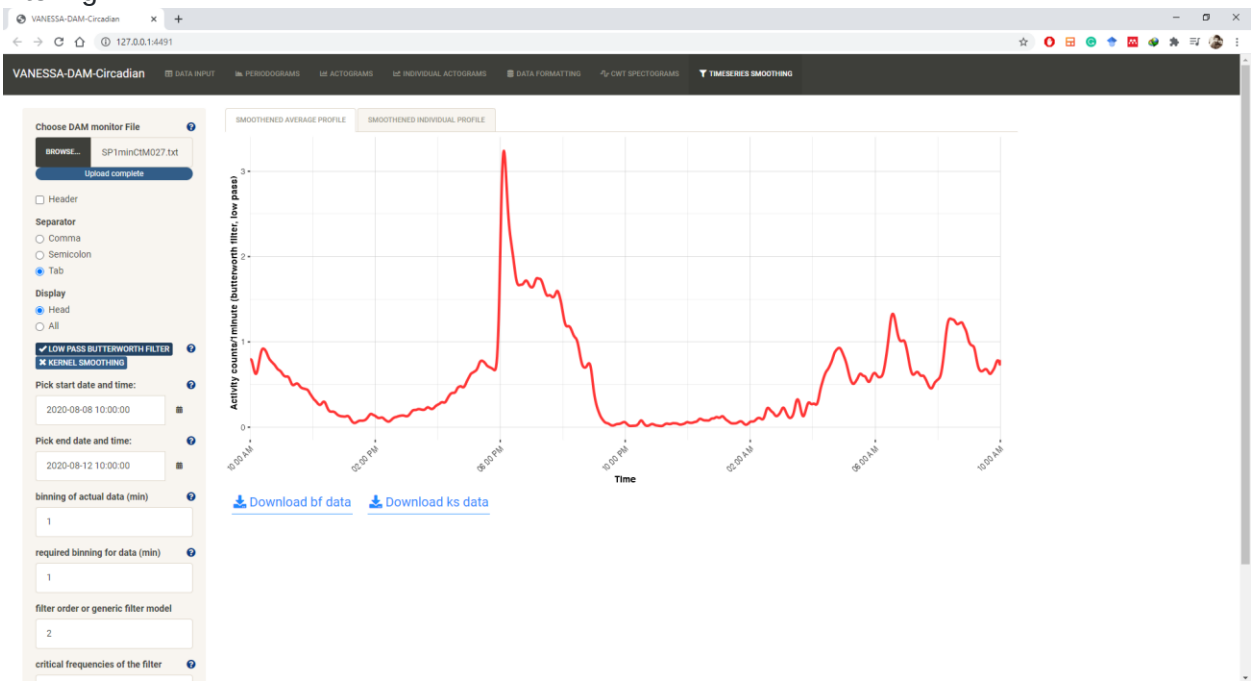


18. The next tab is **CWT SPECTOGRAMS**. This helps in constructing individual CWT spectrograms and plot ridge values and provides 95% CI. Useful for visualizing changing periods in case of long recordings and regime changes. After entering appropriate details in the **bins (min), period lower limit, period upper limit** and **Number of days** boxes, upload your Monitor file in the **Choose DAM monitor file** box and press the

Calculate! Button. After calculations are done for the individual chosen, a CWT spectrogram will be plotted. Change channel number to visualize different individual.



19. The last tab is **Timeseries Smoothing**. This tab has functions which are for day-to-day use while wrangling your data. You can re-bin your data, extract data from different dates, filter and smoothen them, and best of all, download the filtered/smoothened/extracted data. Choose appropriate options, and then visualize average profiles or individual profiles until you're satisfied with the smoothening or filtering.

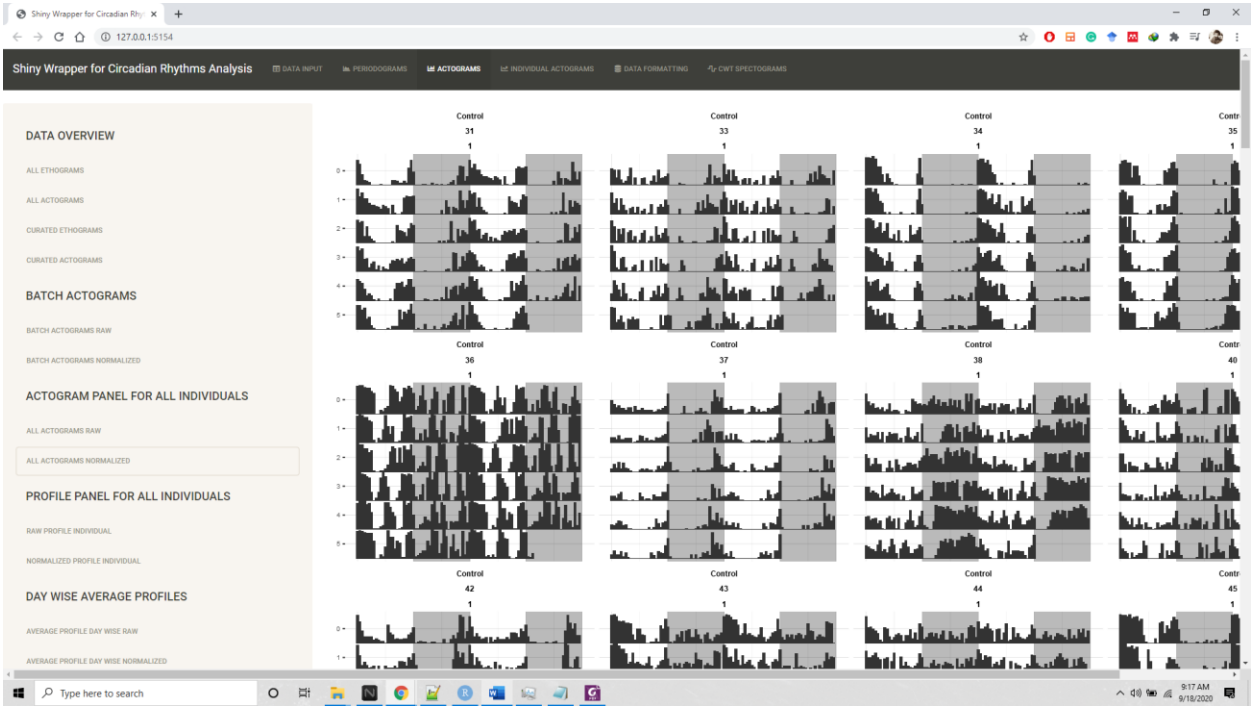


NOTES:

1. If you have replicates, they will be plotted separately in **DAY WISE AVERAGE PROFILES, AVERAGE PROFILES** and **CIRCULAR AVERAGE PROFILES** tabs.

2. **RAW PROFILE ALL REPLICATE AVERAGE** and **NORMALIZED PROFILE ALL REPLICATE AVERAGE** tabs will average over ALL individuals in all replicates in a Genotype.

3. In faceted panels, like **ALL ACTOGRAMS RAW, ALL ACTOGRAMS NORMALIZED** etc., the values on top of each panel is as following: 1-Genotype name (Control in the following example), 2-Arbitrary serial ID assigned on raw data (31,33,34 etc. in example), 3-Replicate number (1 in example).



4. All images produced are high-resolution, can be copied onto clipboard, saved as png files and directly used.

5. For best results, use Monitor files with <5 minutes bin data. Also, cleaning up your Monitor file before using is desired, you don't necessarily have to subset your data by date, all data can be in the Monitor file, only the dates you specify in your Metadata file will be used, thus reducing hassle for the user.

6. If your run was in DD, please put **Duration of light in hours** as 23.99, as it does not take 0 as input, it does not affect any calculations. Also, after you know your average period value in DD, you can change **Modulo tau for actogram** and **LD cycle period** accordingly, and plots will be updated accordingly.

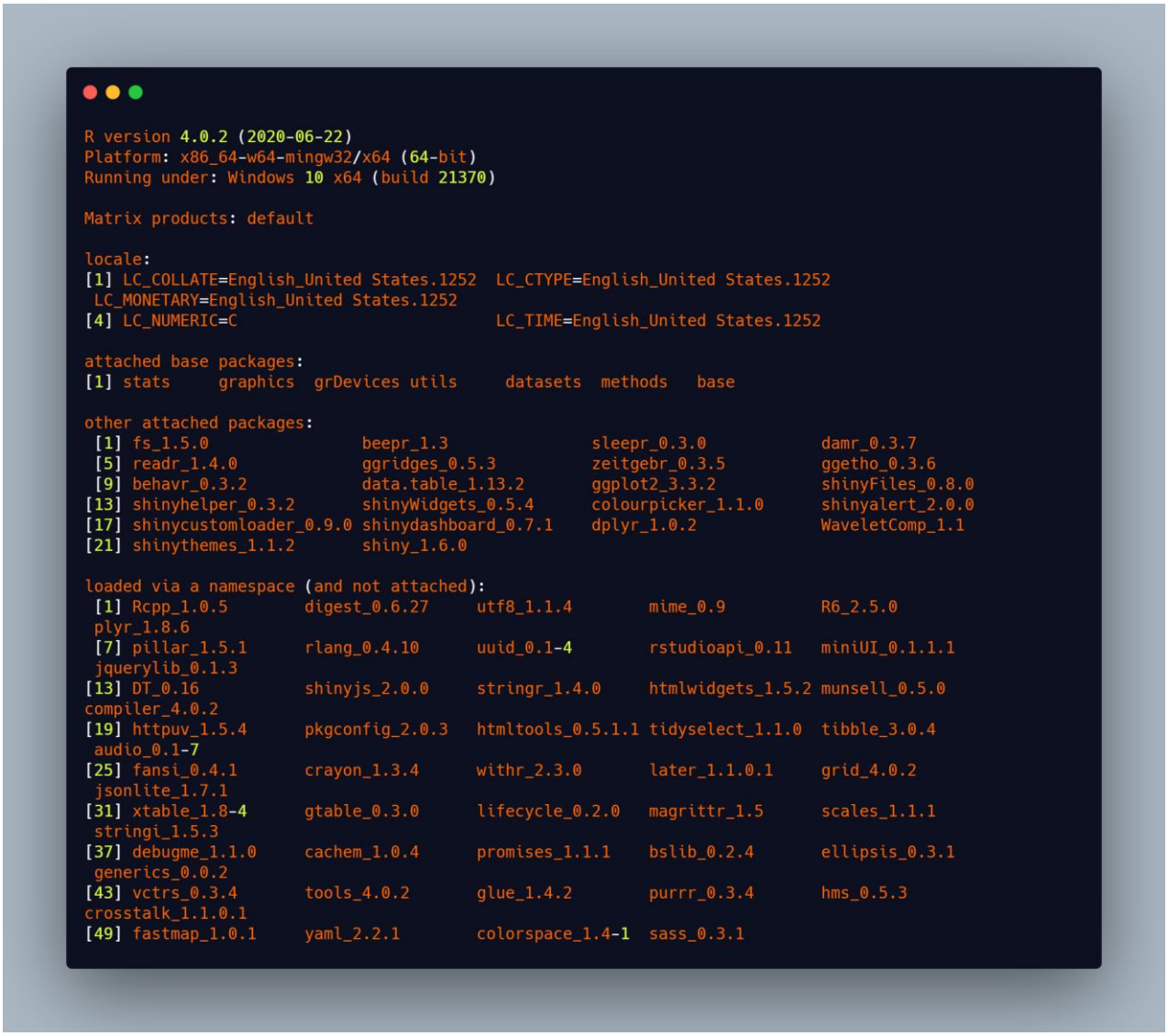
7. The demo data provided is from a locomotor activity run with DAM2 system under DD condition (constant darkness) for 8 days of three groups of flies with different free-running-periods. Each monitor file has 32 flies loaded onto it (standard format for DAM2 systems).

8. All plots can be resized from the “**height**” and “**width**” parameters on top of them for best visualization according to the user. To show the plot, you have to click on the **Plot** button, and each time you change dimension of the plot, you again have to click the **Plot** button.

9. In case you just want to re-bin your data and extract average profiles for specified days using the **Timeseries Smoothing** tab, just select *Kernel smoothing* as your method, and use kernel bandwidth as ***1***, then select start and end date-time and download your average profiles from the **“Download ks data”** button.

10. Use the “*Early10days15minbin.txt*” file provided to use in the “**CWT Spectograms**” tab as an example and the “*SP1minCtM027.txt*” file for use in the “**Timeseries smoothing**” tab.

11. Following is the **`sessionInfo()`** output after using VANESSA-DAM-CRA successfully. Please verify you have similar versions of the packages used in case you run into any problem running the app.



```
R version 4.0.2 (2020-06-22)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 21370)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252
 LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C                           LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] fs_1.5.0                beepr_1.3               sleepr_0.3.0            damr_0.3.7
 [5] readr_1.4.0             ggridges_0.5.3          zeitgebr_0.3.5          ggetho_0.3.6
 [9] behavr_0.3.2            data.table_1.13.2       ggplot2_3.3.2           shinyFiles_0.8.0
[13] shinyhelper_0.3.2       shinyWidgets_0.5.4      colourpicker_1.1.0      shinyalert_2.0.0
[17] shinycustomloader_0.9.0 shinydashboard_0.7.1    dplyr_1.0.2             WaveletComp_1.1
[21] shinythemes_1.1.2       shiny_1.6.0

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.5        digest_0.6.27     utf8_1.1.4        mime_0.9          R6_2.5.0
 plyr_1.8.6
 [7] pillar_1.5.1      rlang_0.4.10      uuid_0.1-4        rstudioapi_0.11   miniUI_0.1.1.1
 jquerylib_0.1.3
[13] DT_0.16           shinyjs_2.0.0     stringr_1.4.0     htmlwidgets_1.5.2 munsell_0.5.0
compiler_4.0.2
[19] httpuv_1.5.4      pkgconfig_2.0.3   htmltools_0.5.1.1 tidyselect_1.1.0  tibble_3.0.4
 audio_0.1-7
[25] fansi_0.4.1       crayon_1.3.4      withr_2.3.0       later_1.1.0.1     grid_4.0.2
 jsonlite_1.7.1
[31] xtable_1.8-4      gtable_0.3.0      lifecycle_0.2.0   magrittr_1.5      scales_1.1.1
 stringi_1.5.3
[37] debugme_1.1.0     cachem_1.0.4      promises_1.1.1    bslib_0.2.4       ellipsis_0.3.1
 generics_0.0.2
[43] vctrs_0.3.4       tools_4.0.2       glue_1.4.2        purrr_0.3.4       hms_0.5.3
crosstalk_1.1.0.1
[49] fastmap_1.0.1     yaml_2.2.1        colorspace_1.4-1  sass_0.3.1
```

Supplementary Methods 2: Easy tutorial to start using VANESSA-DAM-SA

Visualization and ANalysis of timE SerieS dAta - *Drosophila* Activity Monitors (VANESSA-DAM) is a collection of useful tools to visualize and analyze Time series data obtained from *Drosophila* Activity Monitors (https://www.trikinetics.com/). The first in the series of tools is a shiny app for sleep analysis and visualization - VANESSA-DAM for sleep analysis (VANESSA-DAM-SA). For any suggestions, questions, troubleshooting or customization, please contact arijitghosh2009@gmail.com.

VANESSA-DAM-SA

VANESSA-DAM-SA is dependent on Quentin Geissmann's rethomics family of packages - behavr, damr, ggetho, sleepr, for some analysis and visualization options. It offers several advantages over existing tools for sleep analysis from DAM systems, some mentionable ones are -

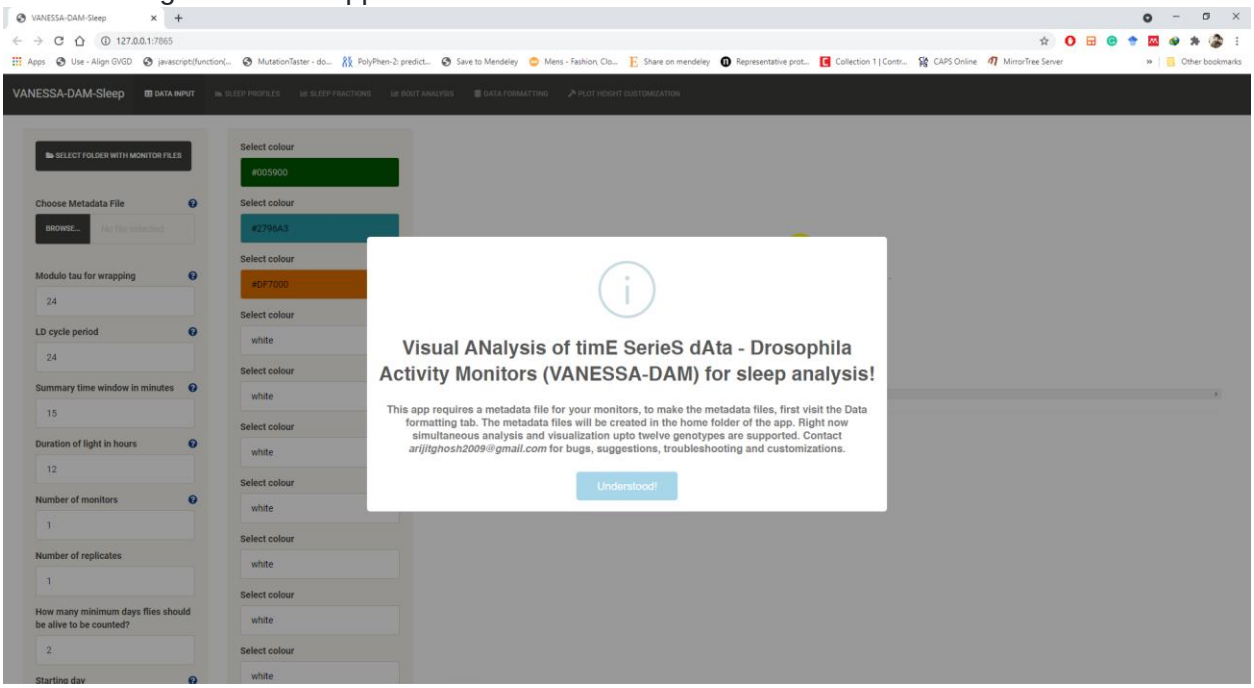
1. Analysis and visualization of multiple monitor files, genotypes, replicates together in a high-throughput manner.
2. Creating metadata files for information about experiment and better reproducibility.
3. Producing high-resolution publication-quality figures with a plethora of customization.
4. Data curation - automatic user-defined parameter-based removal of dead and arrhythmic individuals.
5. Sleep profile analysis, various sleep parameter estimation and quantification, bout analysis, latency analysis.
6. Visual comparison among genotypes, replicates.
7. Dynamic plot resizing and recoloring.
8. Reproducible code report so that you can generate the figures and analysis without the shiny app from RStudio directly.
9. Minimizing human errors – no need to tinker with raw data to accommodate analysis tools.

This short tutorial provided (*Easy tutorial to start using VANESSA-DAM-SA.pdf*) here should be self-explanatory and should help beginners start using the app right away.

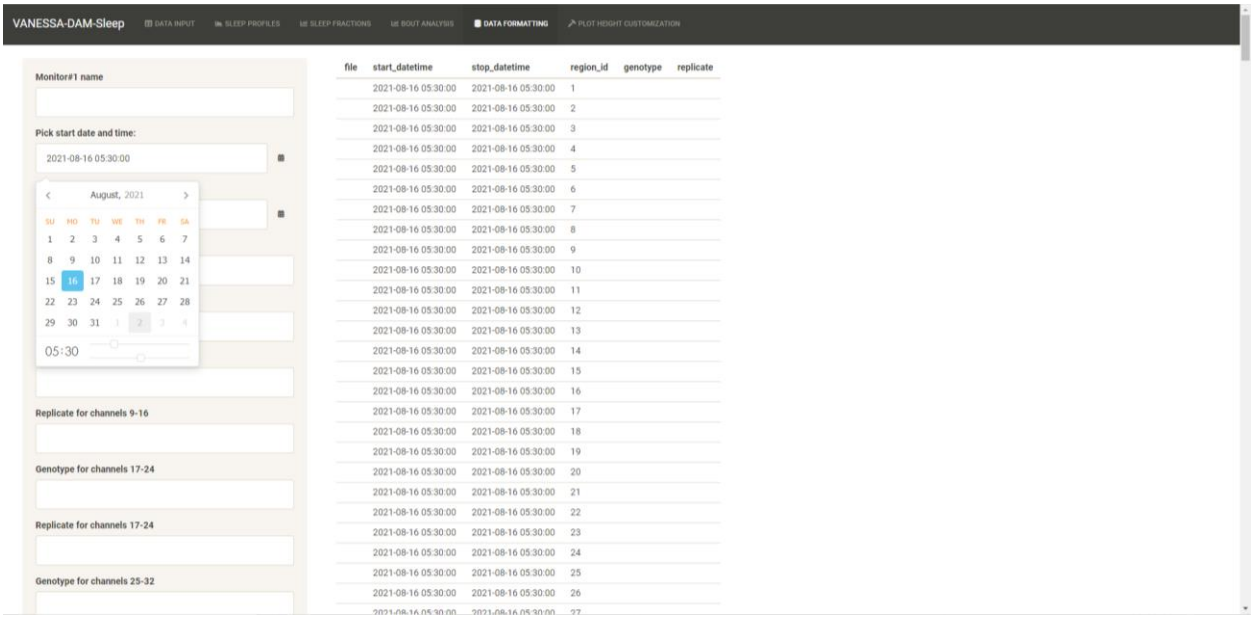
Easy tutorial to start using VANESSA-DAM-SA

1. Install R (version ≥ 3.6.3) and RStudio (version ≥ 1.2) on your computer.
2. Download the zipped folder containing the app and unzip to a folder.
3. Open RStudio and Run the **startup.r** file to install all necessary packages to run the app.
   Note: If you have trouble installing the *damr* package, the source files are provided. Please go to RStudio > Tools > Install packages > Install from : Package Archive file and select the supplied *damr_0.3.7.tar.gz* file to install the package.

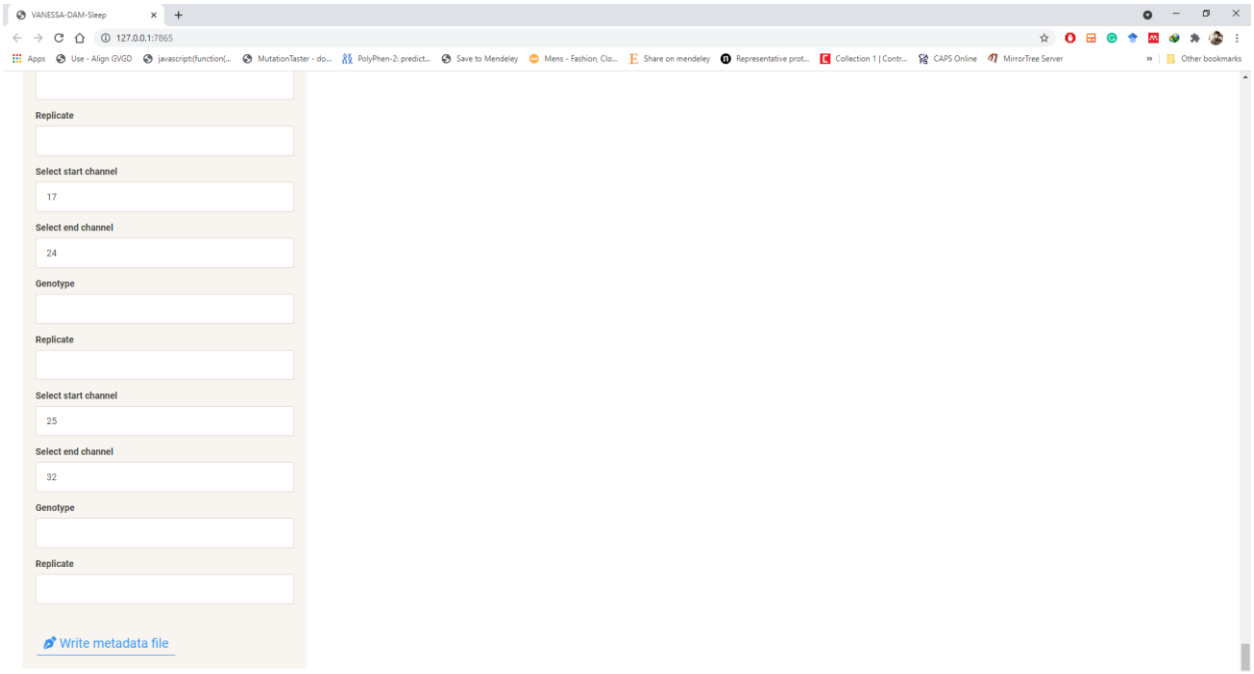
4. Restart R by pressing **ctrl+shift+F10** or from **Session > Restart R** from RStudio menu. This step is useful to clean the global environment so that previously loaded packages do not interrupt, or mask packages/functions needed for VANESSA-DAM-SA.
5. Set working directory as the home folder of the app (by pressing **ctrl+shift+h** OR by using the **`setwd()`** command) and then run the app by typing **`shiny::runApp(launch.browser = T)`** in the console. OR load the **server.r** or **ui.r** file in RStudio by double-clicking on any of them and click on the "**Run App"** button in RStudio panel (for better experience select "*Run External*" from the dropdown menu on the right side of the "**Run App"** button beforehand).
6. The following screen will appear:



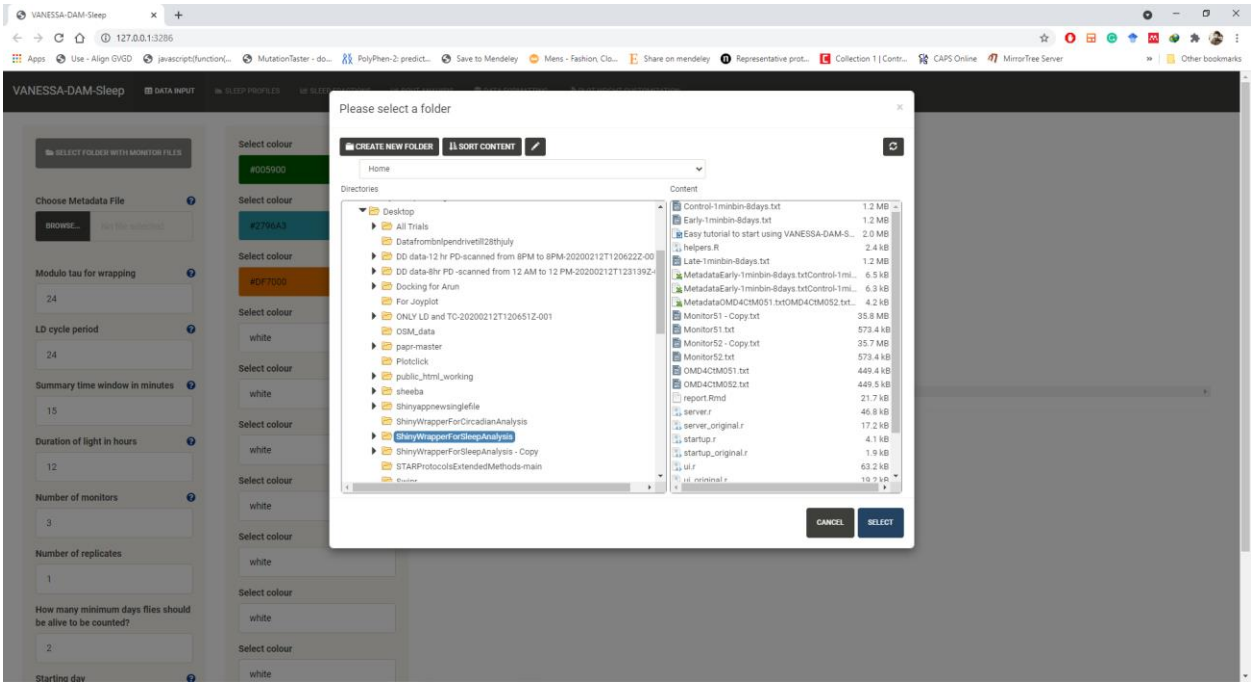
7. Click on **Understood!** button or press **Esc** to begin.
8. *NS215bCtM006.txt, NS215bCtM009.txt* provided with Genotypes *CS-males and CS-females* respectively - check the General Notes section for details of the data provided) from DAMScan in the home folder of the app, first job is to make a Metadata file (Provided as *Metadata.csv*).
   Note: For sleep analysis, raw data files will have to have data in 1 minute bin.
9. In this case, we have 2 genotypes in 2 Monitor files, so enter 2 at the **Number of Genotypes** box. Then go to the **DATA FORMATTING** tab.



10. Enter name of your first monitor file in the **Monitor#1** name box, include extensions also (e.g., if the name of the monitor file Monitor1.txt, then enter the FULL name, including ".txt"). Fill the next boxes - **Start date and time (If your start date is 1st August 2020 and Lights-on time is 10 AM, then you write 2020-08-01 10:00:00), End date (If your end date is 10th August 2020 – 10 AM, then you write 2020-08-10 10:00:00)** with your experimental details as shown above. Enter **Genotype** values and **Replicate number** (only if your experimental design has multiple replicates for all/some genotypes, else keep random number, it won't affect any calculation) in proper boxes for proper channels (marked as **Genotype for channels** 1-8 and **Replicate for channels** 1-8 etc.). A date and time picker has been added to facilitate easy entering of start and end date times without error in formatting. Each monitor can have maximum of 4 genotypes with 8 individuals.
    Note: Creating the metadata file is the single most important step for using this app, any wrong information in the metadata file will wreak havoc on the results and the app may crash. DO NOT change information in the DATA FORMATTING tab once you have entered. For some unknown problem in shiny reactivity, this section does not handle deletion and change of values nicely. Alternatively, you can just create your metadata file by changing data in the already provided metadata file in a good text editor (DO NOT USE EXCEL, as excel may change the date-time format), like Notepad++. The metadata file serves as a base of information for your analysis too and is good for record keeping of your experiments.



11. After filling in experimental details for all 3 monitors, press the **Update metadata** button, and then go to the bottom of the page and click on the **Download metadata** button, this will download the Metadata as a CSV file in your desired folder and is ready to be used.
20. Go back to the **DATA INPUT** tab. Enter choices one by one- **Duration of light in hours** in your experiment, **How many minimum days flies should be alive to be counted**. Also chose your **starting day** and **ending day** for analysis (**starting day** = 1 means first day in your data, so if you want to select first five days, **starting day** should be 1 and **ending day** should be 5). Enter **LD cycle period** in your experiment, **Modulo tau** for actogram can be changed for visualization purpose later, same goes for **Summary time window in minutes** and **Replicates**. You can change any of these parameters later while looking at plots, the plots will immediately update accordingly. For example, if you're looking at your actograms in 15 minutes bin, and you want to visualize them with 5 minutes bin, you just have to change the value in Summary time window in minutes box, similarly LD shading can also be changed. You are all set to start analysis now. **ONLY PROCEED FORWARD WHEN YOU HAVE FIXED ALL ANALYSIS PARAMETERS.** Select the monitor files to be analyzed by clicking the **Choose Monitor Files** button and upload your Metadata file by clicking on the **BROWSE** button in the **Choose Metadata File** box. Press the **Start Calculations!** button, when the calculations are done, you will be notified by a sound. Depending on how long your data is and how many monitors you are analyzing, this step (which involves curation also) will take anything between 5 seconds to 30 seconds typically. To follow this tutorial, enter **Number of monitors** as 2 and **Number of replicates** as 1.
12. *Note: Most of the parameters have a small question mark symbol in blue near them, clicking on the symbol will show a modal with information about the parameter. This help is available is available throughout the app for different parameters, analyses and plots.*



13. Go to the **SLEEP PROFILES** tab next. Explore the various analysis and visualization options by clicking on the left panel tabs. Each tab has description associated with the question mark button.
14. Go to the **Download data** tab to download a csv file for all sleep parameters extracted from your data. There are two different files to be downloaded, one with all sleep parameters, and one with specific bout analysis details. The ***total_bout_length*** parameter in the downloaded files is total sleep time either on whole day or in light and dark phases. **ALL time values are in minutes.**
15. Go to the **SLEEP FRACTIONS** tab to start visualizing sleep fractions as mentioned in different side tabs, feel free to explore.
    Note: sleep_fraction_l denotes sleep in the light part of the day, sleep_fraction_d denotes sleep in the dark part, and sleep_fraction_all denotes total sleep in a day.
16. Go to the **BOUT ANALYSIS** tab to see different plots from sleep architecture analysis.

GENERAL NOTES:

1. If you have different replicates, they will be plotted separately.
2. In faceted panels, the values on top of each panel are as following: 1-Genotype name, 2- serial ID assigned on raw data, 3-Replicate number.
3. All images produced are high-resolution, can be copied onto clipboard, saved as png files and directly used.
4. Cleaning up your Monitor file before using is desired, you don't necessarily have to subset your data by date, all data can be in the Monitor file, only the dates you specify in your Metadata file will be used, thus reducing hassle for the user.
5. If your run was in DD, please put **Duration of light in hours** as 23.99, as it does not 0 as input, it does not affect any calculations. Also, after you know your average period value in DD, you can change **Modulo tau for actogram** and **LD cycle period** accordingly, and plots will be updated accordingly.

6. The data provided is from a locomotor activity run with DAM2 system under LD12:12 for 8 days of CantonS flies (one monitor for males and another for females). Each monitor file has 32 flies loaded onto it (standard format for DAM2 systems).
7. All plots can be resized from the **“height”** and **“width”** parameters on top of them for best visualization according to the user. To show the plot, you have to click on the **Plot** button, and each time you change dimension of the plot, you again have to click the **Plot** button.
8. Most of the violin plots will be faceted into different violin plots over days, day number will be printed on the right side of each plot.
9. Following is the **`sessionInfo()`** output after using VANESSA-DAM-SA successfully. Please verify you have similar versions of the packages used in case you run into any problem running the app.

```
R version 4.0.2 (2020-06-22)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 21370)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252
 LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C                           LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] ggforce_0.3.2           fs_1.5.0                beepr_1.3               sleepr_0.3.0
 [5] damr_0.3.7              readr_1.4.0             ggridges_0.5.3          zeitgebr_0.3.5
 [9] ggetho_0.3.6            behavr_0.3.2            data.table_1.13.2       ggplot2_3.3.2
[13] shinyFiles_0.8.0        shinyhelper_0.3.2       shinyWidgets_0.5.4      colourpicker_1.1.0
[17] shinyalert_2.0.0        shinycustomloader_0.9.0 shinydashboard_0.7.1    dplyr_1.0.2
[21] WaveletComp_1.1         shinythemes_1.1.2       shiny_1.6.0

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.5         digest_0.6.27      utf8_1.1.4         mime_0.9           R6_2.5.0
plyr_1.8.6
 [7] pillar_1.5.1       rlang_0.4.10       uuid_0.1-4         rstudioapi_0.11    miniUI_0.1.1.1
jquerylib_0.1.3
[13] DT_0.16            labeling_0.4.2     shinyjs_2.0.0      stringr_1.4.0      htmlwidgets_1.5.2
 polyclip_1.10-0
[19] munsell_0.5.0      compiler_4.0.2     httpuv_1.5.4       pkgconfig_2.0.3    htmltools_0.5.1.1
 tidyselect_1.1.0
[25] tibble_3.0.4       audio_0.1-7        fansi_0.4.1        crayon_1.3.4       withr_2.3.0
 later_1.1.0.1
[31] MASS_7.3-51.6      grid_4.0.2         jsonlite_1.7.1     xtable_1.8-4       gtable_0.3.0
lifecycle_0.2.0
[37] magrittr_1.5       scales_1.1.1       stringi_1.5.3      debugme_1.1.0      cachem_1.0.4
farver_2.0.3
[43] promises_1.1.1     bslib_0.2.4        ellipsis_0.3.1     generics_0.0.2     vctrs_0.3.4
 RColorBrewer_1.1-2
[49] Cairo_1.5-12.2     tools_4.0.2        glue_1.4.2         tweenr_1.0.1       purrr_0.3.4
 hms_0.5.3
[55] crosstalk_1.1.0.1  fastmap_1.0.1      yaml_2.2.1         colorspace_1.4-1   sass_0.3.1
```