FINAL SCRIPT: APPROVED FOR FILMING

Submission ID #: 68868
Scriptwriter Name: Pallavi Sharma
Project Page Link: https://review.jove.com/files_upload.php?src=21004788

Title: A User-Friendly and Powerful R Analysis of Large-Scale Datasets

Authors and Affiliations:
Jessica Allison, Chong Zhang, Riki Egoshi, Hua Lu

Department of Biological Sciences, University of Maryland Baltimore County

Corresponding Authors:
Hua Lu (hualu@umbc.edu)

Email Addresses for All Authors:
Jessica Allison (all5@umbc.edu)
Chong Zhang (woodyzc@gmail.com)
Riki Egoshi (regoshi@umaryland.edu)
Hua Lu (hualu@umbc.edu)

Author Questionnaire

1. **Microscopy**: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **No**

2. **Software:** Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes, most of them are done.**

Videographer: Please record the computer screen for the shots labeled as SCREEN

3. **Filming location:** Will the filming need to take place in multiple locations? **NO**

4. **Testimonials (optional):** Would you be open to filming two short testimonial statements **live during your JoVE shoot**? These will **not appear in your JoVE video** but may be used in JoVE's promotional materials. **No**

Current Protocol Length
Number of Steps: 27
Number of Shots: 44

Introduction

Videographer: Obtain headshots for all authors available at the filming location.

Videographer's NOTE:
Video
- Recorded at 23.98
- Recorded in Rec709

Sound Channels
- 1:internal mic, 2: internal mic, 3: Boom mic

Files:
- 52 video camera files
- 2 head shots

INTRODUCTION:

1.1. **Hua Lu:** We aim to understand the molecular basis underlying how plants defend against their pathogen and ultimately use this knowledge to improve plant health and crop yield.
   1.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

1.2. **Hua Lu:** Large datasets can be relatively easily generated in the biological field. However, analyzing such large datasets on time can be challenging.
   1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 3.3.1*

CONCLUSION:

1.3. **Hua Lu:** Our research has revealed that the circadian clock is important for plant defense against pathogens. We have routinely used the protocol described here to analyze large-scale time series datasets from both clock and defense assays.
   1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

1.4. **Hua Lu:** Our protocol using the R scripts in RStudio provides a user-friendly and convenient tool for researchers working with large-scale time series data.
   1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B roll: 2.8.1.*

1.5. **Hua Lu:** Our protocol is easy to use, has multiple statistical options, and allows a beginner who does not have prior R knowledge or programming experience to use it
   1.5.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera.

Videographer: Obtain headshots for all authors available at the filming location.

Protocol

2. Luciferase-Based Circadian Clock Analysis

Demonstrator: Jessica Allison

2.1. To begin, add the seeds to a microcentrifuge tube with three holes at the top **[1]** and place the tube inside a bell jar **[2]**. Then generate bleach vapor by adding 3 milliliters of 36 percent hydrochloric acid to 100 milliliters of household bleach in a beaker inside a fume hood **[3].** Place the beaker in the bell jar and leave it in a chemical fume hood for 3 hours **[4]**.

2.1.1. WIDE: Talent adding the seeds to a tube with holes.

2.1.2. Talent placing the beaker into a bell jar.

2.1.3. Talent adding bleach and HCl into a beaker inside a fume hood.

2.1.4. Talent placing the beaker in the bell jar in a chemical fume hood.

2.2. After sterilization, using a sterile glass Pasteur pipette, plate the seeds onto Murashige and Skoog medium plates containing 0.5 percent sucrose and 0.8 percent agar, inside a laminar flow cabinet **[1-TXT]**.

2.2.1. Talent using a sterile glass Pasteur pipette to transfer sterilized seeds onto Murashige and Skoog plates inside a laminar flow cabinet. **TXT: Put chemical waste in a sealed jar and dispose as per institutional guidelines**

2.3. Place the plated and sealed Murashige and Skoog plates in a 4-degree Celsius environment for 2 days **[1]** before moving them into a tissue culture chamber set to a 12-hour light and 12-hour dark cycle **[2-TXT]**.

2.3.1. Talent placing the plates inside a refrigerator unit.

2.3.2. Talent transferring them into a tissue culture chamber with alternating light cycles. **TXT: Allow the seeds to grow for 4 d in LD**

2.4. Then, transfer the seedlings to a 96-well plate, ensuring each well contains 180 microliters of assay media **[1]**. Cover the plate with clear film and poke 2 holes per well for aeration **[2].** Place the plates for 1 day under 12-hour light and 12-hour dark cycles, followed by 1 day under 24 hours of constant light **[2]**.

2.4.1. Talent placing individual seedlings into each well of a 96-well plate filled with assay media.

2.4.2. Talent covers the plate and pokes holes.

2.4.3. Talent positioning the plate in the tissue culture chamber.

2.5. Add treatments or mock solution to each well **[1]**. Then, record luminescence readings under constant light conditions with the emission filter lens **gain** set to **3,600** and the measurement interval time set to **1 s** *(second)* **[2-TXT]** and start the data acquisition **[3]**.

2.5.1. Talent pipetting treatment solutions into the wells of the 96-well plate. **TXT: Remove excess cover so plate stacks smoothly**

2.5.2. Shot of the computer screen showing the settings panel where the emission filter lens gain is adjusted to 3,600 and the measurement interval is set to 1 second. **TXT: Record at 1-h intervals for 5 - 7 days** *Videographer: Please film the computer screen*

Added shot 2.5.3. Shot of the computer screen showing the script for the assay and the start of the data collection.

2.6. At the end of the luminescence recording, take a photograph of the 96-well plate to document the seedling growth **[1]**.

2.6.1. Talent positioning the 96-well plate and capturing an image of the seedlings using a camera**. Videographer's NOTE**: took 4 Stills for 2.6.1

2.7. Observe the data curves from the luminescence assay to confirm treatment consistency **[1]** and save the raw luminescence data as a CSV *(C-S-V)* file for downstream R analysis, using the **Save As** function in the plate reader's data analysis software **[2]**.

Added shot: 2.7.0 2.5.2. Shot of the computer screen showing the saved data curves from the luminescence assay to confirm treatment consistency before downloading data.

2.7.1. Shot of the computer screen showing the Save As menu in the plate reader software and save the file in CSV format.

2.8. After installing RStudio software and the required algorithmic packages, select the working directory where the input file is stored, which will also serve as the output folder for results **[1-TXT]**.

2.8.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:09-00:20. **TXT: MetaCycle for the ARSER algorithm, ggplot2, dplyr, magrittr, stringr, filesstrings, circular, AICcmodavg, and broom**

2.9. Then, select the correctly formatted CSV input file. Ensure the top row contains time points and the first column lists the individual sample positions on the 96-well plate **[1]**.

2.9.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:21—00:00:23 and 00:02:55—00:02:59.

2.10. Name the samples and treatments according to the 96-well plate layout, ensuring that the design includes either 8 replicates per treatment with up to 12 treatments, or 12 replicates per treatment with up to 8 treatments **[1]**. If any wells are empty, assign names such as "Empty 1", "Empty 2", and so on to the treatment label list **[2]**.

2.10.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:25—00:00:31.

2.10.2. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:32—00:00:36.

2.11. Then, indicate the relative start time for the luciferase assay based on the time of light onset in the chamber **[1]**.

2.11.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:37—00:00:41.

2.12. Modify the User Input II section to suit specific analysis needs. Include graph generation for luminescence curves, and for period, phase, and amplitude comparisons across genotypes and treatments **[2]**.

2.12.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:43—00:00:47.

2.13. Use ANOVA test with Tukey's honest significant difference test to compare treatments based on period, phase, and amplitude **[1]**. Choose a control treatment for the analysis output or leave the control field empty to compare all treatments pairwise **[2]**. Optionally, number the analysis output files for easier reference and organization **[3]**.

2.13.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:48—00:00:52.

2.13.2. SCREEN: 68868-2.8.1-2.18.1_t2 00:00:53—00:00:56.

2.13.3. SCREEN:68868-2.8.1-2.18.1_t2 00:00:57—00:01:01.

2.14. Now, use a t-test to compare treatments based on their period, phase, and amplitude. Choose whether the t-test should be conducted as a pairwise comparison and, if so, specify whether the data are paired **[1-TXT]**.

2.14.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:01:02—00:01:07. **TXT: Select a control treatment or leave the control field empty**

2.15. Choose whether to round the time points in the dataset. If time values differ only by a few minutes, round them to the nearest hour before proceeding with the analysis **[1]**.

2.15.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:01:32—00:01:38.

2.16. Now, set the input format for well identification based on how the plate reader exports data. Choose between the standard format, where wells are listed by A1, A2, A3, and so on, or the alternative format where wells are listed by A1, B1, C1, etc. **[1]**

2.16.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:01:44—00:01:50.

2.17. Then, run the luciferase data analysis by clicking the **Source** button located in the top right corner of the RStudio console **[1]**.

2.17.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:02:04—00:02:15.

2.18. View the analysis output in the designated output folder, which contains documents and sub-folders summarizing the averaged period, phase, and amplitude statistics for each genotype and treatment **[1]**.

2.18.1. SCREEN: 68868-2.8.1-2.18.1_t2 00:04:50—00:05:00.

3. Luminol-Based ROS Assay

3.1. Using a biopsy puncher, cut 4-millimeter diameter leaf discs from the fourth to seventh leaves of 25-day-old plants **[1]**. Float the leaf discs with the hairy side facing up in 100 microliters of sterile water inside a 96-well plate **[2]**.

3.1.1. WIDE: Talent using a biopsy puncher to cut uniform discs.

3.1.2. Talent placing individual leaf discs into each well of a 96-well plate, ensuring the hairy side remains upwards.

3.2. Cover the 96-well plate with clean tin foil and place it inside a light and dark growth chamber overnight **[1]**. Then, remove the plate **[2]** and replace sterile water with 100 microliters of luminol solution **[3-TXT]**.

3.2.1. Talent covering the plate with foil and placing it carefully into the growth chamber.

Added shot: Talent taking plate with foil out of the growth chamber.

3.2.2. Talent aspirating water from the wells and pipetting the luminol solution into each well. **TXT: Use a mock solution in place of flg22 for control wells**

3.3. Immediately begin recording luminescence every minute for 40 to 60 minutes **[1]**.

3.3.1. Shot of computer screen showing the recording software interface starting a luminescence read with interval settings set to one minute for 60 cycles.

3.4. After downloading the required Rstudio packages, select the working directory **[1-TXT].** Select the input file, ensuring it is a correctly formatted CSV file. The top row should include time series data, and the first column should contain the sample positions on a 96-well plate **[2].**

3.4.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:14—00:00:23. **TXT: gplot2, dplyr, magrittr, stringr and filesstrings**

3.4.2. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:23—00:00:27 and 00:02:45—00:02:55.

3.5. Name the samples and treatments according to the plate layout as described earlier. Label empty wells explicitly as “empty 1”, “empty 2”, and so on **[1]**.

3.5.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:28—00:00:38.

3.6. Use ANOVA test with Tukey’s honest significant difference to compare total luminescence sums between treatments **[1]**. Use a two-sided t-test when comparing data from only two treatments at a time **[2]**.

3.6.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:39—00:00:44.

3.6.2. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:45—00:00:49.

3.7. Generate graphical outputs, including fluorescence curves and a bar plot showing the total luminescence sum across treatments. Add either standard deviation or standard error of the mean to the plotted bars **[1]**.

3.7.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:00:53—00:01:04.

3.8. Adjust the input reading format based on how the plate reader outputs well identifiers. Choose between standard listing by A1, A2, A3 or vertical listing by A1, B1, C1 **[1]**.

3.8.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:01:08—00:01:20.

3.9. To run the analysis, click the **Source** button located in the top right corner of the RStudio console **[1]**. View the output in the generated folder, which contains multiple documents and sub-folders summarizing the full analysis **[2]**.

3.9.1. SCREEN: 68868-3.4.1-3.9.2-t1 00:01:49—00:01:56.

3.9.2. SCREEN: 68868-3.4.1-3.9.2-t1 00:03:35—00:03:45.

Results

4. Results

4.1. The luciferase assay was performed using one transgenic line expressing the CCA1 *(C-C-A-One)* luciferase reporter and seven independently transformed transgenic lines expressing the GRP7 luciferase reporter **[1].** The luminescence traces of these plants were measured over 168 hours **[2].**

4.1.1. LAB MEDIA: Figure 3 and 4

4.1.2. LAB MEDIA: Figure 3. *Video editor: Show all colored lines curves*

4.2. Using the R method, computed clock parameters for the CCA1 *(C-C-A-One)* luciferase reporter were obtained with an amplitude of 3,000 relative luminescence units per second per seedling**,** a period of 23.5 hours, and a phase of 3.5 hours **[1].** All pGRP7 *(P-G-P-R-7)* luciferase lines displayed similar period and phase values but varied in amplitude **[2].** The period of pGRP7 luciferase is 24.2 hours, while the phase is 12 hours **[3]**.

4.2.1. LAB MEDIA: Figure 4. *Video editor: Highlight the two color bars for pGRP7wt:LUC lines (bars 2 to 9), which are visibly taller than the CCA1:LUC bar on the left*

4.2.2. LAB MEDIA: Figure 4. *Video editor: Highlight the period bars for all pGRP7wt:LUC lines (bars 2 to 9), which are higher than the first bar for CCA1:LUC*

4.2.3. LAB MEDIA: Figure 4. *Video editor: Highlight the phase bars for pGRP7wt:LUC lines (bars 2 to 9), which are taller than the CCA1:LUC bar*

4.3. To further validate the R analysis, the same dataset was re-analyzed using BioDare2, a free online platform for circadian data analysis8, and similar results were obtained.

4.3.1. LAB MEDIA: Figure 4. *Video editor: Highlight the light blue bars in all the three graphs*

4.4. The circadian data generated with U2 OS cells expressing the *Per2:dLuc* reporter were re-analyzed using the R method **[1].** The control group displayed an amplitude of 184.8 relative luminescence units, a period of 23.3 hours, and a phase of 2.8 hours **[2].** Knockdown of CRY2 *(Cry-Two)* **[3]** but not PSMD4 *(P-S-M-D-Four)* and PSMD7 *(P-S-M-D-Seven)* significantly affected circadian parameters, amplitude, phase, and period. These results were consistent with the published results **[4].**

4.4.1. LAB MEDIA: Figure 5

4.4.2. LAB MEDIA: Figure 5A and B. *Video editor: Highlight the curve and bar labeled*

"Neg si" in 5A and B, respectively

4.4.3. LAB MEDIA: Figure 5A and B. *Video editor: Highlight the curve and bar labeled CRY2 in 5A and B, respectively*

4.4.4. LAB MEDIA: Figure 5A and B. *Video editor: Highlight the PSMD4 and PSMD7 and curve bars in 5A and B*

- microcentrifuge

Pronunciation link: https://www.merriam-webster.com/dictionary/microcentrifuge
IPA: /ˌmaɪkroʊˈsɛntrəˌfjuːdʒ/
Phonetic Spelling: my-kroh-sen-truh-fyoohj

- bell jar

Pronunciation link: https://www.merriam-webster.com/dictionary/bell%20jar
IPA: /ˈbɛlˌdʒɑr/
Phonetic Spelling: bel-jar

- hydrochloric (in "hydrochloric acid")

Pronunciation link: https://www.merriam-webster.com/dictionary/hydrochloric
IPA: /ˌhaɪdroʊˈklɔrɪk/
Phonetic Spelling: hi-droh-klor-ik

- luminescence

Pronunciation link: https://www.merriam-webster.com/dictionary/luminescence
IPA: /ˌluːməˈnɛsəns/
Phonetic Spelling: loo-muh-nes-uhns

- assay

Pronunciation link: https://www.merriam-webster.com/dictionary/assay
IPA: /əˈseɪ/
Phonetic Spelling: uh-say

- ANOVA

Pronunciation link: https://www.merriam-webster.com/dictionary/anova
IPA: /əˈnoʊvə/
Phonetic Spelling: uh-noh-vuh

- Tukey's (in "Tukey's honest significant difference test")

Pronunciation link: https://www.merriam-webster.com/dictionary/tukey
IPA: /ˈtuːki/
Phonetic Spelling: too-kee

1. Murashige (from “Murashige and Skoog medium”)
   Pronunciation link: https://www.howtopronounce.com/murashige-e-skoog
   IPA: /muːrəˈʃiːgeɪ/ (approx)
   Phonetic Spelling: moo-ruh-shee-gay
2. Skoog (the surname in “Murashige and Skoog”)
   Pronunciation link: https://www.howtopronounce.com/skoog
   IPA: /skuːg/
   Phonetic Spelling: skoog
3. luciferase (as in “luciferase reporter”)
   Pronunciation link: https://www.howtopronounce.com/luciferase
   IPA: /ˌluːsɪˈfɛɹeɪz/ (American)
   Phonetic Spelling: loo-si-fer-aiz
4. BioDare2 (the online platform)
   Pronunciation link: no confirmed link found
   IPA: /ˌbaɪəʊˈdɛə tuː/ → but for American: /ˌbaɪoʊˈdɛr tuː/
   Phonetic Spelling: bye-oh-dair-too
5. AICcmodavg (the R package name)
   Pronunciation link: no confirmed link found
   IPA: /ˌeɪ-aɪ-si-siː mɒd ævɜːdʒ/ (US approx: /ˌeɪ-aɪ-siː-si-mɒd ævərɪdʒ/)
   Phonetic Spelling: ay-eye-see-see mod-av-rij
6. MetaCycle (another R package you mention)
   Pronunciation link: no confirmed link found
   IPA: /ˈmɛtəˌsaɪkəl/
   Phonetic Spelling: meh-tuh-sahy-kuhl
7. ggplot2 (R package name)
   Pronunciation link: no confirmed link found
   IPA: /ˈdʒiːˈdʒiːˌplɒt tuː/
   Phonetic Spelling: gee-gee-plot two
8. dplyr (R package name)
   Pronunciation link: no confirmed link found
   IPA: /ˈdiːˌplʌr/ (approx)
   Phonetic Spelling: dee-plur
9. magrittr (R package name)
   Pronunciation link: no confirmed link found
   IPA: /məˈgrɪtər/
   Phonetic Spelling: muh-grit-er
10. stringr (R package name)
    Pronunciation link: no confirmed link found
    IPA: /ˈstrɪŋgər/
    Phonetic Spelling: string-er
11. filesstrings (R package name)
    Pronunciation link: no confirmed link found
    IPA: /ˌfaɪlzˈstrɪŋz/
    Phonetic Spelling: files-strings

12. circular (R package name)
    Pronunciation link: https://www.merriam-webster.com/dictionary/circular
    IPA: /ˈsɜːrkjələr/ (American: /ˈsɜrkjələr/)
    Phonetic Spelling: ser-kyuh-ler
13. Metamorphosis (from your example; not in text, but just a comparable term)
    (Already provided earlier)