

EEB 603 – Reproducible Science

Syllabus

Sven Buerki - Boise State University

2023-08-18

Contents

1	Instructor	2
2	Our class ethos	2
3	Class details	3
3.1	Teaching mode and class location	3
3.2	Meeting days and times	3
3.3	Timetable	3
3.4	Health and safety protocols	3
4	Course goal & description	3
5	Structure of the course	4
5.1	Content of the course	5
6	Shared Google Drive	6
7	Publications & Textbooks	6
8	Examples of reproducible science	6
9	The computing tools of reproducible science	7
9.1	The bioinformatic tools	7
9.2	Installing the main software	7
9.3	Installing markup languages	8
9.4	Installing R packages	8
10	RStudio Cheat Sheets	8
11	R tutorials	9

12 Assessments	9
12.1 Tests conducted during the course & grading	9
12.2 Bioinformatic tutorial (150 points)	10
12.3 Teaching your bioinformatic tutorial (100 points)	11
12.4 Individual report on thesis project/publication (200 points)	11
12.5 Oral presentation on thesis project/publication (100 points)	12
13 Late work policy	12
14 Engagement expectations	12
15 What you can expect of me	12
16 This course was designed with you in mind	13
16.1 This class’s community is inclusive	13
16.2 We use each other’s preferred names and pronouns	13
16.3 This course is accessible to students with disabilities	14
16.4 For students responsible for children	14
17 Academic integrity	14
18 Student well-being	14
18.1 What do you need?	14
19 References	15
20 Appendix 1	17

[Download pdf version](#)

[Raw data on GitHub](#)

1 Instructor

- Name: Sven Buerki (he/his)
- Office: Science building, office 114 (ground floor).
- Email: svenbuerki@boisestate.edu
- Office hours: By appointment.

2 Our class ethos

Everyone here is smart; distinguish yourself by being kind.

Kindness in Science is an inclusive approach that fosters diversity, respect, wellbeing & openness leading to better science outcomes.

The overarching aim of this course is therefore to provide students with the theoretical knowledge and bioinformatic tools necessary to improve transparency, reproducibility and efficiency in scientific research. Across the course of the course, students will be taught how to use open source software for their research such as [R](#), [RStudio](#) and [R Markdown](#) (incl. [knitr](#)). To further master these latter skills and improve communication and teaching skills, students will *i*) design bioinformatic tutorials and teach those to their peers and *ii*) design and implement individual projects aiming at developing a reproducible workflow specific to their research interest.

Overall, this course **provides students with key knowledge to gather, store, share, prepare and analyze data as well as communicate results to the scientific community and various stakeholders** (see [Figure 1](#)).

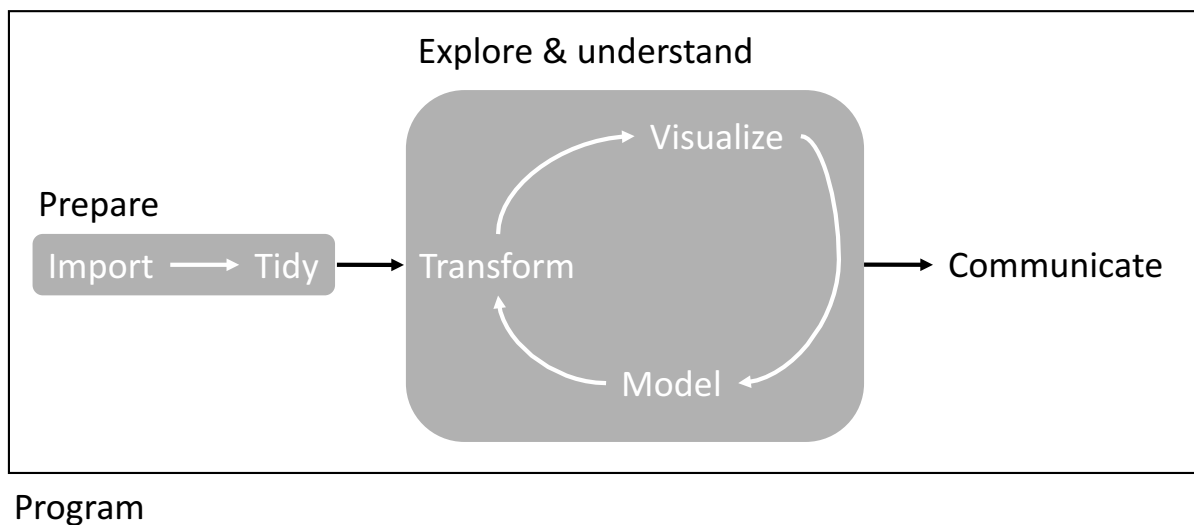


Figure 1: Overview of workflow studied in PART2.

5 Structure of the course

The course is subdivided into three parts:

- *PART 1: The Big Picture*
- *PART 2: Bioinformatics for Reproducible Science*
- *PART 3: Apply a Reproducible Approach to your Data*

Part 1 provides students with key theoretical knowledge on reproducible science allowing them to successfully design and implement a reproducible approach tailored to Ecology, Evolution & Behavior. This part will also cover material associated with open science and data management and how these practices intertwine with scientific publications.

Part 2 provides students with opportunities to further learn and apply the coding and bioinformatic tools required to implement a reproducible workflow ([Figure 1](#)). Here, students

(with the support of the instructor) will develop and teach a tutorial on a specific bioinformatic subject, which will be taught over two full classes. Tutorials will be written in RMarkdown and distributed to the class one week in advance. Depending on class size, this assignment will be conducted in group or individually.

Finally, part 3 was designed to provide students with an opportunity to develop individual reproducible workflows tailored to their research project (or to data presented in a publication if they don't have a thesis subject yet). This assignment will be done by applying knowledge gained during the previous parts and working in collaboration with the instructor (in some cases, we might seek support from thesis advisers).

5.1 Content of the course

PART 1: The Big Picture

1. **Introduction: The reproducibility crisis & prospects to tackle it!**
 - Chapter 1: Introduction to R, RStudio, Markdown (incl. referencing) & User-defined functions.
 - Chapter 2: The reproducibility crisis.
 - Chapter 3: A roadmap to implement reproducible science in Ecology, Evolution & Behavior.
 - Chapter 4: Open science and CARE principles.
2. **Getting started: Overview of data workflow & used software**
 - Chapter 5: Data management, Reproducible code.
 - Chapter 6: Getting published & Peer review.

PART 2: Bioinformatics for Reproducible Science

3. **Prepare your data**
 - Chapter 7: Organize and import data with R.
 - Chapter 8: Prepare/tidy data for analyses in R.
4. **Explore & understand your data**
 - Chapter 9: Statistical modelling and *knitr*.
 - Chapter 10: Visualize results with Tables.
 - Chapter 11: Visualize results with Figures (incl. handling phylogenetic trees).
5. **Communicate/Disseminate your results**
 - Chapter 12: Git & GitHub: What are those and how can they help you with your code and communicating your research?

PART 3: Apply a Reproducible Approach to your Data

6. **Students develop and present reproducible workflows applied to their project**
 - Produce individual reports showcasing reproducible workflows tailored to thesis projects.
 - Oral presentations of individual students' projects.

6 Shared Google Drive

A shared Google drive is set up to facilitate data sharing and allow students uploading their [bioinformatic tutorials](#).

- URL: <https://drive.google.com/drive/folders/1eZyjGQ478vaKiGD01hcryyNtdGp9t1J7?usp=sharing>

7 Publications & Textbooks

The reading material at the basis of this course is composed of a mixture of publications and chapters mostly from two textbooks ([Gandrud, 2015](#); [Wickham and Golemund, 2017](#)). We will also study the “*Guides to*” published by the [British Ecological Society](#). Please find below the references used in each **chapter**. **This list is not exhaustive and additional literature will be provided in class.**

Chapter	Reference(s)
Chap. 1	Chapter 3 of Gandrud (2015)
Chap. 2	Baker (2016); Freedman et al. (2015); Munafo et al. (2017); Peng and Hicks (2021); Sarewitz (2016)
Chap. 3	Bone et al. (2015); Markowetz (2015); Smith et al. (2016)
Chap. 4	Carroll et al. (2021); Creative Commons ; Wagner et al. (2022); Williams et al. (2023)
Chap. 5	British Ecological Society (2014a) & Chapter 4 of Gandrud (2015); British Ecological Society (2014d) & Chapter 2 of Gandrud (2015); Trisovic et al. (2022)
Chap. 6	British Ecological Society (2014b); British Ecological Society (2014c)
Chap. 7	Chapter 6 of Gandrud (2015)
Chap. 8	Chapter 7 of Gandrud (2015) & Chapters 9-10 of Wickham and Golemund (2017)
Chap. 9	Chapter 8 of Gandrud (2015)
Chap. 10	Chapter 9 of Gandrud (2015)
Chap. 11	Chapter 10 of Gandrud (2015), Chapters 1 and 22 of Wickham and Golemund (2017) & Guangchuang et al. (2017)
Chap. 12	Chapter 5 of Gandrud (2015)

8 Examples of reproducible science

To further exemplify how a reproducible approach can be implemented in your research, the instructor provides here examples of publications and software produced by students that have attended this course (alphabetically sorted):

- Cole et al. ([2022](#)) – **EcoCountHelper: an R package and analytical pipeline for the analysis of ecological count data using GLMMs, and a case study of**

bats in Grand Teton National Park.

- Ellestad et al. (2022) – **Genomic Insights into Cultivated Mexican Vanilla planifolia Reveal High Levels of Heterozygosity Stemming from Hybridization.**
- Wojahn et al. (2021) – **G2PMineR: A Genome to Phenome Literature Review Approach.**

9 The computing tools of reproducible science

Research is often presented in the form of slideshows, articles or books. These presentation documents announce a project’s findings, but they are not the research, they are the advertisement part of the research project!

The research is the full software environment, code, and data that produced the results (Donoho, 2010).

When we separate the research from its advertisement, we are making it difficult for others to verify the findings by reproducing them.

This course will give you the tools to dynamically combine your research with the presentation of your findings. The first tool will be a workflow for reproducible research weaving the principles of reproducibility throughout your entire research project, from data gathering to the statistical analysis, and the presentation of results. To reach this goal, you will learn how to use a number of computer tools that make this workflow possible.

9.1 The bioinformatic tools

The main bioinformatic tools covered in this course are:

- The **R** statistical language that will allow you to gather data and analyze it.
- The **LaTeX** and **Markdown** markup languages that you can use to create documents (slideshows, articles, books, webpages) for presenting your findings.
- The *knitr* and *rmarkdown* **packages** for R and other tools, including **command-line shell programs** like GNU Make and Git version control, for dynamically tiding your data gathering, analysis, and presentation documents together so that they can be easily reproduced.
- **RStudio**, a program that brings all of these tools together in one place.

9.2 Installing the main software

As shown above, **R** and **RStudio** are at the core of this course and will have to be installed on your computers. This can be easily done by downloading the software from the following websites:

- **R**: <https://www.r-project.org>
- **RStudio**: <https://www.rstudio.com/products/rstudio/download/>

The download webpages for these software have comprehensive information on how to install them, so please refer to those pages for more information.

9.3 Installing markup languages

If you are planning to create LaTeX documents, you will need to install a Tex distribution. Please refer to this website for more details: <https://www.latex-project.org/get/>

If you want to create Markdown documents you can separately install the *rmarkdown* package in R (see [below](#) for more details).

9.4 Installing R packages

We will be using a number of R packages especially designed to support reproducible research. Many of those packages are not included in the default R installation and will need to be installed separately. To install key packages used in class, copy the following code and paste it into your R console:

```
install.packages(c("brew", "countrycode", "devtools", "dplyr", "ggplot2", "googleVis",  
"knitr", "rmarkdown", "tidyr", "xtable"))
```

Once you enter this code, you may be asked to select a CRAN “mirror” to download the packages from. Simply select the mirror closest to you.

Finally, it is highly likely that we will have to install additional packages. In this case, you can simply install it by using the same R function `install.packages()` or by using RStudio as follows: Select “Tools” -> “Install Packages ...” and then type the name of the package in the window (make sure to tick the “Install dependencies” box).

10 RStudio Cheat Sheets

RStudio provides a suite of cheat sheets that can be accessed by going to the “Help” menu and selecting “Cheatsheets”.

Five cheat sheets are especially relevant to chapters taught in this course:

- *RStudio IDE : Cheat sheet*
- *Data Manipulation with dplyr, tidyr*
- *Data Visualization with ggplot2*
- *R Markdown Cheat Sheet*
- *R Markdown Reference Guide*

These documents together with the material presented in [publications & textbooks](#) will provide the basis to design your [bioinformatic tutorials](#).

Table 2: Grading scale applied in this course.

Percentage	Grade
100-98	A+
97.9-93	A
92.9-90	A-
89.9-88	B+
87.9-83	B
82.9-80	B-
79.9-78	C+
77.9-73	C
72.9-70	C-
69.9-68	D+
67.9-60	D
59.9-0	F

11 R tutorials

Please find below two documents providing a comprehensive introduction to R:

- R for beginners (a tutorial by Emmanuel Paradis): https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf
- An introduction to R: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

12 Assessments

There will not be any classical exams in this course, but we will rather focus on developing theoretical and bioinformatic skills and applying those to your research. In this context, each student will be asked to produce a bioinformatic tutorial and teach it to their peers (see **PART 2**). Each student will also be tasked to produce a report (tailored to their thesis project or a publication) and present their results and conclusions in class.

12.1 Tests conducted during the course & grading

Students will be graded based on the following four tasks:

- Produce a bioinformatic tutorial focusing on a chapter from **PART 2**. Depending on enrollment students may be working in pairs (150 points).
- Teach a bioinformatic lab (spread across 2 sessions; 100 points).
- Produce an individual report on thesis project/publication (200 points).
- One oral presentation on thesis project/publication (100 points).

Exams are summing to a total of **550 points** and Table 2 exhibits the grading scale applied in this course.

12.2 Bioinformatic tutorial (150 points)

During week 1, students will be assigned a chapter of **PART 2** to study and produce a bioinformatic tutorial. Based on enrollment, students might work individually or in pairs (please see below for more information).

Tutorials will have to be written in the *knitr/rmarkdown* language as implemented in RStudio. Tutorials should be focused on developing a suite of exercises aiming at gaining key bioinformatic skills specific to each chapter (see **PART 2**). Students will be welcomed to use material presented in Gandrud (2015) and Wickham and Grolemund (2017) to develop their tutorials, but they can also use other sources as long as they are properly cited in their documents. See **Publications & Textbooks** and **RStudio Cheat Sheets** sections for more details.

Students should design their tutorials to be completed within 2 laboratory sessions (see **below**). **Tutorials should be submitted to the instructor 1 week in advance for correction and to be uploaded onto the shared Google drive.**

12.2.1 What do you need to keep in mind?

While designing your tutorials, please think about the following points:

1. What are your major objectives and how do they fit within the scientific process and data life-cycle? Please also focus on the theoretical background.
2. What are the specific learning outcomes?
3. What data are used to teach this tutorial?
4. What are the bioinformatic dependencies and how to install them?
5. When you teach each concept, make sure to clearly explain the rationale, steps to get there (pseudo-code) and focus on the input/output and data class/format used for each function.

12.2.2 Content of the tutorial

Based on the information provided above, your tutorial should include:

- A short introduction highlighting the theory and aims of the tutorial.
- A section on R package requirements with instructions on how to install those packages and their dependencies.
- A section introducing the data (dataset) used to support these exercises (and how to download those).
- A references section and links to manuals of R packages.
- Commented R code necessary to guide users through the exercises as well as some knowledge on expected outputs.

12.2.3 Google sheet with groups for bioinformatic tutorials

The instructor is asking students to sign up for designing and teaching a bioinformatic tutorial by accessing the following Google sheet (the number of students per chapter is provided,

please respect this guideline):

- https://docs.google.com/spreadsheets/d/1gUwbqJ4gbhhSY1DfsZ7HsoDLzOKB-57xAAC1_IJ3qmk/edit?usp=sharing

12.3 Teaching your bioinformatic tutorial (100 points)

Students are expected to prepare a 10 minutes presentation providing general guidelines to complete the tutorial. Presentations will be uploaded onto the shared Google drive and made accessible to all students. Students are expected to support their peers in completing tutorials by answering questions. The instructor will also be answering questions, but students are leading the teaching of the bioinformatic laboratories.

Students will be graded according to their abilities to teach their tutorials and answer questions. The instructor might also use student's feedback to grade this test.

12.4 Individual report on thesis project/publication (200 points)

Students will work alongside the instructor to develop a reproducible workflow specific to their thesis project. In cases where students do not yet have a clear idea on their thesis project, they will work with instructor to identify a publication that can serve as basis for their individual project.

Reports will be written using the *knitr/rmarkdown* markup language as implemented in RStudio. The instructor expects students to provide a list of references supporting their reports. References will have to be cited in the text: it is not enough to through a bunch of references at the end of the report. This exercise aims at supporting methodological decisions taken in the report and increasing transparency.

12.4.1 Students gather prior key elements on their project

- What kind of data are already published/available (and which of those are relevant to your topic)?
- Where are those published data deposited?
- Can you reproduce the analyses based on published data?
- What types of data are or will be produced during your thesis project?
- What are the specificity of your data (in term of storage, sharing, etc.)?
- What are the publication standards in your field (see e.g. [Donoho, 2010](#); [Smith et al., 2016](#))?
- etc...

12.4.2 Students develop the following core processes for their project

- A data management workflow specific to your research, which will cover the following stages of the data life cycle (see e.g. [British Ecological Society, 2014a](#)):
 - Create
 - Process

- Document
- Preserve
- Share
- Reuse
- A reproducible code to perform the following tasks to your data (see Fig. 1):
 - Import
 - Tidy/clean
 - Transform
 - Visualize
 - Model
 - Communicate

12.5 Oral presentation on thesis project/publication (100 points)

Each student will have to present their report during final week. The presentation should follow the same structure as the report and not exceed 15 minutes. There will be 5 minutes at the end of the presentation allocated for questions.

13 Late work policy

The instructor expects students to deliver their assignments on time and set enough time aside to work on their projects (see above for more details). However, if you have any issues preventing completion of your work on time, please contact the instructor as soon as possible to find common solutions.

14 Engagement expectations

The instructor is expecting students to attend classes (please join on time and for the full duration of the class) and actively engage by asking questions and giving feedback on teaching material and course content. This course was designed to help students implementing a reproducible approach to their research projects. If you are judging that additional content should be covered in class, please contact the instructor. The instructor will do his very best to obtain information or seek support from colleagues to cover the requested material. In the case that you have any issues attending class, please contact the instructor by email (svenbuerki@boisestate.edu) asap and see below for more details.

15 What you can expect of me

The instructor will be prepared for class, on time and not leave early. He will also be respectful of you and your opinions. Overall, the instructor wants to foster a kind and respectful class environment where all students can express themselves and share their opinions. This means that meaningful and constructive dialogue is encouraged in this class and it requires a degree of mutual respect, willingness to listen, and tolerance of opposing points of view. Respect

for individual differences and alternative viewpoints will be maintained at all times in this class. One's words and use of language should be temperate and within acceptable bounds of civility and decency. Finally, the instructor will reply to emails and grade tests as soon as possible (and provide positive criticism) to allow students mastering the material presented in class.

16 This course was designed with you in mind

The instructor has developed this course to provide a welcoming environment and effective, equitable learning experience for all students. If you encounter barriers in this course, please bring them to my attention so that I may work to address them.

16.1 This class's community is inclusive

Students in this class represent a rich variety of backgrounds and perspectives. The [Biological Sciences department](#) is committed to providing an atmosphere for learning that respects diversity and creates inclusive environments in our courses. While working together to build this community, we ask all members to:

- Share their unique experiences, values, and beliefs, if comfortable doing so.
- Listen deeply to one another.
- Honor the uniqueness of their peers.
- Appreciate the opportunity we have to learn from each other in this community.
- Use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the campus community.
- Recognize opportunities to invite a community member to exhibit more inclusive, equitable speech or behavior—and then also invite them into further conversation. We also expect community members to respond with gratitude and to take a moment of reflection when they receive such an invitation, rather than react immediately from defensiveness.
- Keep confidential any discussions that the community has of a personal (or professional) nature, unless the speaker has given explicit permission to share what they have said.
- Respect the right of students to be addressed and referred to by the names and pronouns that correspond to their gender identities, including the use of non-binary pronouns.

16.2 We use each other's preferred names and pronouns

Please let the instructor know of your preferred or adopted name and gender pronoun(s), and he will make those changes to his own records and address you that way in all cases.

To change to a preferred name so that it displays on all BSU sites, including Canvas and the course roster, contact the Registrar's Office at (208) 426-4249. Note that only a legal name change can alter your name on BSU official and legal documents (e.g., your transcript).

16.3 This course is accessible to students with disabilities

The instructor recognizes that navigating your education and life can often be more difficult if you have disabilities. I want you to achieve at your highest capacity in this class. If you have a disability, the instructor needs to know if you encounter inequitable opportunities in this course related to: - Accessing and understanding course materials. - Engaging with course materials and other students in the course. - Demonstrating your skills and knowledge on assignments and exams.

If you have a documented disability, you may be eligible for accommodations in all of your courses. To learn more, make an appointment with the university's [Educational Access Center](#).

16.4 For students responsible for children

The instructor recognizes the unique challenges that can arise for students who are also parents or guardians of children. If you have any specific needs related to this topic, please contact the instructor asap.

17 Academic integrity

To create a welcoming, engaging, and effective learning environment, the instructor expects all of us to exhibit behavior that reflects Boise State's [Statement of Shared Values](#). The Shared Values emphasize academic excellence, caring, citizenship, fairness, respect, responsibility, and trustworthiness. In keeping with these values, the instructor expects students in this course to uphold the standards outlined in the Boise State University [Student Code of Conduct](#).

18 Student well-being

If you are struggling for any reason (e.g., COVID, relationship, family, or life's stresses) and believe these may impact your performance in the course, the instructor is encouraging you to contact the Dean of Students at (208) 426-1527 or email deanofstudents@boisestate.edu for support. Additionally, if you are comfortable doing so, please reach out to me and I will provide any resources or accommodations that I can. If you notice a significant change in your mood, sleep, feelings of hopelessness or a lack of self worth, consider connecting immediately with Counseling Services (1529 Belmont Street, Norco Building) at (208) 426-1459 or email healthservices@boisestate.edu.

18.1 What do you need?

The university has many resources designed to support you as a learner and human being. Among these are:

- [Albertsons Library](#) provides a treasure trove of physical and electronic resources.

- As you enter the library, straight ahead you'll find the Reference Desk, where librarians can help you find the information and resources you need.
- The Circulation Desk lets students borrow various technologies.
- The MakerLab on the second floor offers tools for student use, and there are friendly staff in the MakerLab to help you learn how.
- The [Writing Center](#) offers individual consultations tailored to your needs, including making sense of writing assignment instructions, brainstorming, crafting a thesis, organizing an essay, revisions, citations, and more.
- [Counseling Services](#) helps you tap into your strengths and find resources to deal more effectively with concerns that impact your pursuit of personal and academic goals. It emphasizes prevention and early detection and provides a broad spectrum of short-term counseling, consultative, evaluative, teaching, and training functions. Counseling staff consists of licensed counselors, psychologists, and closely supervised trainees/post-graduate interns.
- The [Gender Equity Center](#) provides specialized, brief, no-cost, confidential support for campus members affected by:
 - Identity-related harassment.
 - Sexual harassment, sexual assault, unhealthy relationships, and stalking.
 - Life experiences related to gender, sexual orientation, and allyship.
 - Life transitions and crisis.
 - Personal, financial, or academic roadblocks.
- [Multicultural Student Services](#) works with all students interested in equity and social justice, students from different cultures, countries and ethnicities. The staff understands cultural differences and perspectives, and knows what it's like living in a dominant culture. The training and the programs Multicultural Student Services provides can help you navigate through political correctness and into respectful dialogue. This center also provides:
 - Support and guidance for planning cultural festivals and events.
 - In-center tutoring and studying lounge.
 - Volunteer opportunities.
 - Mentoring and Training.
 - Programs and events.
- Food assistance: If you are hungry and cannot afford to purchase food, the campus has some resources to help you. You can visit the [campus food pantry](#) or [get free meals in the campus dining hall](#).

19 References

- BAKER, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452–454. Available at: <https://doi.org/10.1038/533452a>.
- BONE, R.E., J.A.C. SMITH, N. ARRIGO, and S. BUERKI. 2015. A macro-ecological perspective on crassulacean acid metabolism (CAM) photosynthesis evolution in afro-madagascan drylands: Eulophiinae orchids as a case study. *New Phytologist* 208: 469–481. Available at: <http://dx.doi.org/10.1111/nph.13572>.
- BOULTER, J., M.L. OROZCO MORALES, N. PRINCIPE, and C.M. TILSED. 2023. What is

- kindness in science and why does it matter? *Immunology & Cell Biology* 101: 97–103. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/imcb.12580>.
- BRITISH ECOLOGICAL SOCIETY ed. 2014a. A guide to data management in ecology and evolution. British Ecological Society.
- BRITISH ECOLOGICAL SOCIETY ed. 2014b. A guide to getting published in ecology and evolution. British Ecological Society.
- BRITISH ECOLOGICAL SOCIETY ed. 2014c. A guide to peer review in ecology and evolution. British Ecological Society.
- BRITISH ECOLOGICAL SOCIETY ed. 2014d. A guide to reproducible code in ecology and evolution. British Ecological Society.
- CARROLL, S.R., E. HERCZOG, M. HUDSON, K. RUSSELL, and S. STALL. 2021. Operationalizing the CARE and FAIR principles for indigenous data futures. *Scientific Data* 8: 108. Available at: <https://doi.org/10.1038/s41597-021-00892-0>.
- COLE, H.J., D.G. GOMES, and J.R. BARBER. 2022. EcoCountHelper: An r package and analytical pipeline for the analysis of ecological count data using GLMMs, and a case study of bats in grant teton national park. *PeerJ* 10: e14509. Available at: <https://doi.org/10.7717/peerj.14509>.
- DONOHO, D.L. 2010. An invitation to reproducible computational research. *Biostatistics* 11: 385–388. Available at: <http://dx.doi.org/10.1093/biostatistics/kxq028>.
- ELLESTAD, P., M.A. PÉREZ-FARRERA, and S. BUERKI. 2022. Genomic insights into cultivated mexican vanilla planifolia reveal high levels of heterozygosity stemming from hybridization. *Plants* 11: 2090. Available at: <https://doi.org/10.3390/plants11162090>.
- FREEDMAN, L.P., I.M. COCKBURN, and T.S. SIMCOE. 2015. The economics of reproducibility in preclinical research. *PLOS Biology* 13: e1002165. Available at: <https://doi.org/10.1371/journal.pbio.1002165>.
- GANDRUD, C. 2015. Reproducible Research with R and RStudio. C. Gandrud [ed.], CRC Press.
- GUANGCHUANG, Y., S.D. K., Z. HUACHEN, G. YI, and L.T.T. YUK. 2017. Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8: 28–36. Available at: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12628>.
- MARKOWETZ, F. 2015. Five selfish reasons to work reproducibly. *Genome Biology* 16: 274. Available at: <https://doi.org/10.1186/s13059-015-0850-7>.
- MUNAFO, M.R., B.A. NOSEK, D.V.M. BISHOP, K.S. BUTTON, C.D. CHAMBERS, N.P. DU SERT, U. SIMONSOHN, et al. 2017. A manifesto for reproducible science. *Nature human behaviour* 1: 0021.
- PENG, R.D., and S.C. HICKS. 2021. Reproducible research: A retrospective. *Annual Review of Public Health* 42: 79–93. Available at: <https://doi.org/10.1146/annurev-publhealth-012420-105110>.
- SAREWITZ, D. 2016. The pressure to publish pushes down quality. *Nature* 533: 147–147. Available at: <https://doi.org/10.1038/533147a>.
- SMITH, J.F., T.H. PARKER, S. NAKAGAWA, J. GUREVITCH, ECOLOGY, and T.(TOOLS FOR T. IN EVOLUTION) WORKING GROUP. 2016. Promoting transparency in evolutionary biology and ecology. *Systematic Botany* 41: 495–497. Available at: <http://www.bioone.org/doi/abs/10.1600/036364416X692262>.

- SOLTIS, D.E., V.B. SMOCOVITIS, K.K. PHAM, M.B.S. CORTEZ, A.L. SMITH, and P.S. SOLTIS. 2023. Rethinking the ph.d. Dissertation in botany: Widening the circle. *American Journal of Botany* 110: e16136. Available at: <https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.1002/ajb2.16136>.
- TRISOVIC, A., M.K. LAU, T. PASQUIER, and M. CROSAS. 2022. A large-scale study on research code quality and execution. *Scientific Data* 9: 60. Available at: <https://doi.org/10.1038/s41597-022-01143-6>.
- WAGNER, A.S., L.K. WAITE, M. WIERZBA, F. HOFFSTAEDTER, A.Q. WAITE, B. POLDRACK, S.B. EICKHOFF, and M. HANKE. 2022. FAIRly big: A framework for computationally reproducible processing of large-scale data. *Scientific Data* 9: 80. Available at: <https://doi.org/10.1038/s41597-022-01163-2>.
- WICKHAM, H., and G. GROLEMUND. 2017. R for data science: Import, tidy, transform, visualize, and model data. 1st ed. O'Reilly Media, Inc. Available at: <http://r4ds.had.co.nz>.
- WILLIAMS, J.W., A. TAYLOR, K.A. TOLLEY, D.B. PROVETE, R. CORREIA, T.B. GUEDES, H. FAROOQ, et al. 2023. Shifts to open access with high article processing charges hinder research equity and careers. *Journal of Biogeography* 50: 1485–1489. Available at: <https://doi.org/10.1111/jbi.14697>.
- WOJAHN, J.M.A., S.J. GALLA, A.E. MELTON, and S. BUERKI. 2021. G2PMineR: A genome to phenome literature review approach. *Genes* 12: 293. Available at: <https://doi.org/10.3390/genes12020293>.

20 Appendix 1

Citations of all R packages used to generate this report.

- [1] J. Allaire, Y. Xie, C. Dervieux, et al. *rmarkdown: Dynamic Documents for R*. R package version 2.21. 2023. <https://CRAN.R-project.org/package=rmarkdown>.
- [2] C. Boettiger. *knitcitations: Citations for Knitr Markdown Files*. R package version 1.0.12. 2021. <https://github.com/cboettig/knitcitations>.
- [3] M. C. Koohafkan. *kfigr: Integrated Code Chunk Anchoring and Referencing for R Markdown Documents*. R package version 1.2.1. 2021. <https://github.com/mkoohafkan/kfigr>.
- [4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2022. <https://www.R-project.org/>.
- [5] H. Wickham, J. Bryan, and M. Barrett. *usethis: Automate Package and Project Setup*. R package version 2.1.6. 2022. <https://CRAN.R-project.org/package=usethis>.
- [6] H. Wickham, R. François, L. Henry, et al. *dplyr: A Grammar of Data Manipulation*. R package version 1.0.9. 2022. <https://CRAN.R-project.org/package=dplyr>.
- [7] H. Wickham, J. Hester, W. Chang, et al. *devtools: Tools to Make Developing R Packages Easier*. R package version 2.4.4. 2022. <https://CRAN.R-project.org/package=devtools>.
- [8] Y. Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*. ISBN 978-1138700109. Boca Raton, Florida: Chapman and Hall/CRC, 2016. <https://bookdown.o>

[rg/yihui/bookdown](https://yihui.org/bookdown).

- [9] Y. Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.33. 2023. <https://CRAN.R-project.org/package=bookdown>.
- [10] Y. Xie. *Dynamic Documents with R and knitr*. 2nd. ISBN 978-1498716963. Boca Raton, Florida: Chapman and Hall/CRC, 2015. <https://yihui.org/knitr/>.
- [11] Y. Xie. *formatR: Format R Code Automatically*. R package version 1.12. 2022. <https://github.com/yihui/formatR>.
- [12] Y. Xie. “knitr: A Comprehensive Tool for Reproducible Research in R”. In: *Implementing Reproducible Computational Research*. Ed. by V. Stodden, F. Leisch and R. D. Peng. ISBN 978-1466561595. Chapman and Hall/CRC, 2014.
- [13] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.42. 2023. <https://yihui.org/knitr/>.
- [14] Y. Xie and J. Allaire. *tufte: Tufte’s Styles for R Markdown Documents*. R package version 0.12. 2022. <https://github.com/rstudio/tufte>.
- [15] Y. Xie, J. Allaire, and G. Golemund. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman and Hall/CRC, 2018. ISBN: 9781138359338. <https://bookdown.org/yihui/rmarkdown>.
- [16] Y. Xie, C. Dervieux, and E. Riederer. *R Markdown Cookbook*. Boca Raton, Florida: Chapman and Hall/CRC, 2020. ISBN: 9780367563837. <https://bookdown.org/yihui/rmarkdown-cookbook>.
- [17] H. Zhu. *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.3.4. 2021. <https://CRAN.R-project.org/package=kableExtra>.