

Dear editor,

First of all, we would like to thank reviewer for his careful reading of our manuscript, for his positive comments and for his relevant remarks. We also would like to apologize for the delay we took to answer and submit a revised version of our article. We have teaching responsibilities at the University and during the last months, they have been very time consuming. We thank you, and the reviewer for your understanding.

You can find below point-by-point answers to all the remarks made by the reviewer. We hope our explanations and the modifications we made in the main text and Gitbook, will clarify the interest of the FAIR\_bioinfo courses.

Thomas Denecker & Claire Toffano-nioche

## Reviewer : Lex Nederbragt

*The GitBook and course clearly address a pressing need for training in the Life Sciences. The book is comprehensive, and treats many, if not most of the tools that are relevant. I applaud the authors for doing such a thorough job of documenting their workflow for FAIR bioinformatics.*

*But herein also lies a challenge: the GitBook feels almost like a 'brain dump' of the authors. There are long descriptions of tools and only brief explanations on how to use them. Individual commands, choice of parameters, or alternative choices of these parameters are not explained, even though the target audience description says 'No prior computational skills are required to complete the entire course'.*

All these remarks are correct. We have indeed selected the information given according to our audience. The aim of this training is to enable Biologists to understand the problems of reproducibility in order to ask Bioinformaticians. We have added the commands so that they can reproduce them but without expecting them to be totally independent at the end of the training.

*There is no lesson plan that helps others to design course activities, hardly any exercises. Is each lesson taught in one day? Over several days (which is what I think would be needed)? The exercises that are there are basically asking the learner to find a lot out on their own, there is very little guidance.*

This information was indeed missing. We added it in the Jose paper and in the Gitbook. It should be counted as an hour and a half of class per session. We left 1 month between sessions to give the learners time to mature the concepts.

*There are no worked examples, although one could say the entire book is one huge worked example.*

Indeed, we made only one example voluntarily. We wanted to show a complete pipeline of classical bioinformatics analysis.

*I found no description of how the lessons were taught, the GitBook is a book, more than a lesson plan.*

We completed the Gitbook.

*There is a lot of ground being covered and tools are often taught very superficially. As an example, the git lesson is around 3/4 page. In comparison (maybe a bit unfair, but still), the Software Carpentry git lesson takes 3 hours at least, and than it does not even cover branching!*

Our goal was not for learners to know how to use the tools to perform the analysis, but rather for them to know that they exist. We cited many examples orally and in the Gitbook so that they could go further if they wanted to redo everything themselves from their own data. This is an initiation to reproducibility and we give them some guidelines so that they can train themselves.

*So in all, I think the GitBook is an OK reference for self study by people familiar with the bioinformatic tools used. But FAIR\_Bioinfo as it stands is not a stand-alone lesson that other teachers can easily adopt to their needs.*

With the help of all the remarks, we have made corrections to the Gitbook and we hope that it is now easier to use it again.

*I would like to finish my review with a list of suggestions what can be done to improve the module:*

- *check the order of teaching tools, e.g. the shell command should be known when git is taught*

The order has been changed. The course starts with the shell.

- *explain much more of what each command means, alternatively, assuming preexisting knowledge of the shell and/or the bioinformatic tools used*

As explained above, this is not the purpose of this course.

- *provide the slides in English*

The Gitbook contains all the slides translated into figures.

- *take it much slower, use more time per tool/part of the workflow*

It would always take more and more time to do things perfectly, and we are fully aware of this. Nevertheless, the learners did not have more time to devote to the training. We wanted to maximize the number of participants.

- *add timings, and instructor notes: how is someone going to teach this section?*

This information has been added to the Gitbook.

- *add much more exercises, some scaffolded, and sparingly totally open exercises ('now run this on your own')*

The entire course is a replayable example. We have given additional resources (e.g. Software carpentry) if they wish to practice .

- *prioritize, so with suggestions what instructors could skip if they have not enough time to teach the entire module*

This information has been added to the Gitbook (when is possible).

- *make the learning outcomes for each chapter explicit*

This information has been added to the Gitbook.

- *help learners to install the necessary software to run the bioinformatics steps*

We explained how to install conda and docker which are the two necessary dependances. If they want to install a tool that doesn't go through these two tools, we verbally advised them to go to Bioinformaticians.

- *add contributor guidelines so that instructors using the module can give you input*

We've added paragraphs about this in the Gitbook to each chapter.

---

*I will now comment on why I did not mark some of boxes from the checklists.*

## General checks

- **License:** Does the repository contain a plain-text LICENSE file with the contents of a standard license? (OSI-approved for code, Creative Commons for content)

*The license is describes as CeCILL FREE SOFTWARE LICENSE AGREEMENT. I am unsure whether this license is compliant.*

The version 2.1 of CeCILL is officially recognized by OSI as an *Open Source* license (see <https://opensource.org/licenses/category> "International Licenses").

## Documentation

- **A statement of need:** Do the authors clearly state the need for this module and who the target audience is?

*Only in the paper, not in the GitBook.*

We add it in the introductory chapter of the Gitbook (need: <https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/#objectives> ; target audience: <https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/#schedule> )

- **Installation instructions:** Is there a clearly stated list of dependencies?

*Not in the GitBook, nor in the paper, and I did not find a requirements.txt file. However, there is a docker container.*

One of the objectives of the course is to raise awareness of the value of reproducible installation of tools. For this purpose, we explain in the course the installation and use of the two necessary tools which are: docker and conda. And as indicated, all the tools are available in the docker image associated with the course. We have therefore not edited any requirement.txt file.

- **Usage:** Does the documentation explain how someone would adopt the module, and include examples of how to use it?

*There are instructions on how to run the entire analysis pipeline in the paper. I had one issue with that, see [thomasdenecker/FAIR\\_Bioinfo#3](#). There is little to no explanation on how to adopt the module.*

In order to be able to use the file selector locally, you need a volume share in Docker and the computer. To do this, add to the docker command -v your\_local\_address:/srv/shiny-server. For more details, the official documentation is here <https://docs.docker.com/storage/volumes/> and in our Gitbook is here: <https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/a-trip-to-the-sea#image-to-container-how-to-launch-an-image-greater-than-docker-run> .

- **Community guidelines:** Are there clear guidelines for third parties wishing to 1) Contribute to the module 2) Report issues or problems with the module 3) Seek support

*No.*

We placed our code in github and the lesson in gitbook. These two resources have their own functionalities for third parties to contribute, report issues, and seek support.

## **Pedagogy / Instructional design (Work-in-progress: reviewers, please comment!)**

- **Learning objectives:** Does the module make the learning objectives plainly clear? (We don't require explicitly written learning objectives; only that they be evident from content and design.)

*One can distill the learning objectives, but I hesitate to say that they are made 'plainly clear'.*

In the conclusion of each chapter, we have added what has been learned

- **Pedagogy:** Does the module seem easy to follow? Does it observe guidance on cognitive load? (working memory limits of 7 +/- 2 chunks of information)

*No. There is a lot of information that is presented with fairly limited explanations. Cognitive load will be high. Examples:*

*- <https://fair-bioinfo.gitbook.io/fair-bioinfo/fair-principles#how-to-use-the-fair-data-principles> does not introduce the shell and bash, jumps straight into git (ls in [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory) is not explained, etc)*

We swapped chapters as now the bash shell language (chapter 2: <https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#the-bash-language>) is presented before git (chapter 3: [https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/code\\_memory#the-git-tool](https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/code_memory#the-git-tool)).

*- <https://fair-bioinfo.gitbook.io/fair-bioinfo/some-first-definitions#what-is-a-terminal> mentions terminal commands, but does not teach them, teaching happens after git in <https://fair-bioinfo.gitbook.io/fair-bioinfo/its-not-magic>*

We have deleted the chapter “some first definition” and we have repositioned the presentation and launch of the terminal before any command line action (chapter 2: <https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#the-terminal-the-swiss-army-knife>)

*- [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory) shows the use of cd which is not taught until next session/chapter*

With the exchange of the shell and git chapters, the cd command is now seen in the shell chapter (chapter 2: [https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#initiate-the-fair\\_project](https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#initiate-the-fair_project)) before being used in git chapter (chapter 3: [https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/code\\_memory#initialise-git](https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/code_memory#initialise-git)).

*- <https://fair-bioinfo.gitbook.io/fair-bioinfo/some-first-definitions#a-couple-of-ssh-keys> is going to lead to confusion, high cognitive load, could wait until it is needed rather than here*

We have deleted the chapter “some first definition” and we have moved the presentation of an ssh key just before its use to the “a trip to the sea” chapter (<https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/a-trip-to-the-sea#a-couple-of-ssh-keys>)

*- <https://fair-bioinfo.gitbook.io/fair-bioinfo/its-not-magic>: wget, tar and sed commands are not explained; also '2. wget (EBI): no installation required because wget is usually present by default.' wget is not installed by default on the Mac.*

We have added explanations for the wget, tar (<https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#how-to-download-web-content-with-command-line>), and sed commands (used to replace windows carriage returns with unix ones,

<https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#preparation-of-a-design-file>)  
. We have added URL links to help Mac user for installing wget (english or french pages).

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/its-not-magic#using-a-loop> *If you have never used a for loop in bash, this for j in \$(tail -n +2 .././conditions.txt) is quite a big leap!*

We have re-written the paragraph about the loop concept and we have decomposed the command (which we did orally) to explain more progressively this complex command (<https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/its-not-magic#the-loop-concept>).

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/its-not-magic> *uses sudo apt-get update sudo apt-get install tree to install tree, but this will not work on all systems (e.g., Mac); lti would also help to explain sudo*

We have removed the visualization comfort provided by the tree command, so we no longer need the apt-get commands before chapter 5, in a docker container in an unix system. We have also added an explanation of the sudo command (<https://app.gitbook.com/@fair-bioinfo/s/fair-bioinfo/a-trip-to-the-sea#containerization-with-docker>).

-

<https://fair-bioinfo.gitbook.io/fair-bioinfo/ive-got-the-power#creating-the-snakefile-of-our-project> *has the commands for the pipeline, but these are not run or explained outside of snakemake. As an example, what actual command is run, and what it does, could be explained: bowtie2 -x {params.index} -U {input.file} -S {output.sam} > {output.out}*

We have chosen not to detail the launch orders for bioinformatics analysis tools in order not to distract from the focus of the course, which is to increase reproducibility. Moreover, any other analysis could replace the example we have chosen.

- **Instructional design:** Is the instructional design deliberate and apparent? For example, exploit worked-example effects; effective multi-media use; low extraneous cognitive load.

*There are no worked example; there are useful illustrative images, but not multi-media; at times, extraneous cognitive load is high (see also elsewhere). Can the slides be translated to English?*

The scope of the concepts covered in the course is large, leaving little space/time with the time constraints we had (8x1h30). We have chosen an application example to make the concepts concrete but we have not added other examples.

The most important slides have been translated and included as figures in the gitbook. We do not plan to translate the slides because the specificity of our course is that it is in French and many resources already exist in English.

## JOSE paper

- Does it describe how it has been used in the classroom or other settings, and how someone might adopt it?

*Maybe yes to the first, no to the second*

- Could someone else teach with this module, given the right expertise?

*Yes, but see comments elsewhere on 'book versus lesson plan'*

- Does the paper tell the "story" of how the authors came to develop it, or what their expertise is?

*Yes to the first (briefly), no to the second.*

- **References:** Do all archival references that should have a DOI list one (e.g., papers, datasets, software)?

*I could find a link to a paper in the GitBook, but that did not have a doi; no doi's for datasets or software.*

To respond to these requests, we have completed the paper as follow:

*how it has been used in the classroom:*

Each session lasted an hour and a half ; they were carried out at the rate of one session per month.

*How someone might adopt it:*

The slides of each session are available on the FAIR\_Bioinfo github in French. They have been translated into English and inserted as figures in the Gitbook. English speaking instructors can very easily create slides from the Gitbook and these figures.

*what their expertise is:*

This course was born from the combination of our personal experiences in reproducibility acquired during our research work in Bioinformatics.

*no doi's for datasets or software:*

For the concrete example, we chose a set of sequence data that are stored in international databases and gave their identification numbers. For the tools in the analysis workflow, we made the installations with conda by specifying the version numbers of the tools in the dockerfile. The identification numbers and version numbers of the software are considered equivalent to the doi by the scientific communities to which they refer.

---

Some minor issues:

## PDF

- *Reproducible --> Reproducible*

The typographical error has been corrected.

- *the figure is misplaced, it appears between the heading of table 1 and table 1 itself*

It seems that the problem comes when generating the PDF. We'll have to check this in the final generation.

## [https://github.com/thomasdenecker/FAIR\\_Bioinfo](https://github.com/thomasdenecker/FAIR_Bioinfo)

- *I ran the workflow successfully without having to use sudo. Is running as root really necessary?*

No, only in a linux environment. We left the sudo because the majority of the participants had a linux system. But we added this remark (in the paper and in the gitbook).

## GitBook

- *There is quite some remaining French language, some of it should be translated*

We thank you and we tried to track them to correct them.

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/fair-principles#how-to-use-the-fair-data-principles> *"Thus, we do not follow the following principles" --> it looks like you are following these principles?*

We've corrected that part.

- *same page: "Tools used = references in their field" I am not sure what that means?*

We used the most common and widely used tools to perform this analysis.

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/data-presentation> *--> could study system be introduced more? It is not even mentioned this is a green alga, and what the goal of the experiment was*

It is true that in the Gitbook, we did not put the experimental design contrary to the oral presentation (slide). However, we have added the reference of the article whose title indicates that the study model is the green alga *ostreococcus tauri*.

- *same page: fastq q-scores are not explained*

We have chosen not to go into detail when the commands are specific to the analysis. This is a perfect example of a specific commands. Indeed, if we didn't have fastq files, we



wouldn't do this command but others adapted to our files. On the other hand, all the commands for reproducibility are described in detail

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/some-first-definitions> it does not become clear that the expectation is the learner installs a text editor

Indeed, in the gitbook we only put the installation links for text editors. During the oral presentation, we showed live how to install it using the documentation.

- [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory#what-when-why-git-and-github](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory#what-when-why-git-and-github) calls git a 'version manager' which I would not use (elsewhere it is called 'version control tool', which is OK)

We made the correction in the gitbook.

- [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory#using-git](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory#using-git) repeatedly has git statut

Rehearsal is voluntary and is part of the pedagogy of teaching. We wanted to show the changes at all stages to show participants that Git sees everything.

- same page: 'Let's take the example of the family photo at the Sunday meal to illustrate the rest' I found that confusing and not an analogy that helps with understanding

This example worked very well orally... We will think about illustrations to make it clearer in writing.

- same page, the first \$ git diff has a sentence that is not from its output: The file will have its original line endings in your working directory.

We have correct this by copying the terminal outputs.

- [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory#how-to-create-a-new-branch](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory#how-to-create-a-new-branch) h-git-branch-less-than-newbranchname-greater-than git branch is not a git command I know, and as far as I know should be git checkout -b <newBranchName>

Indeed, this command is a combination of both. We have shown the two ways of doing it (one written and the other oral).

- same page, just below: branche --> branch

The typo error has been corrected.

- [https://fair-bioinfo.gitbook.io/fair-bioinfo/code\\_memory#how-to-switch-changes-from-one-branch-to-the-main-branch-git-merge-less-than-branchname-greater-than](https://fair-bioinfo.gitbook.io/fair-bioinfo/code_memory#how-to-switch-changes-from-one-branch-to-the-main-branch-git-merge-less-than-branchname-greater-than) 'How to switch changes from one branch to the main branch?' --> I would not use 'switch' here

We show how to switch branches using git checkout. We have replaced “switch” by “move toward”.

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/lover#table-exploration-with-dplyr> is a weird stand-alone demo that learners probably think they will have to execute, but I think they did no load the data

Concerning dplyr, the goal is just to show the simplicity of the dplyr syntax compared to the "classical" syntax. We didn't go further than this presentation (no live demonstration). The idea is to push users towards new R practices and to inform them.

- <https://fair-bioinfo.gitbook.io/fair-bioinfo/notebooks#anaconda> 'As regards reproducibility, doubts may persist.' --> I do not understand the meaning of this remark.

If Anaconda is updated, we do not know if the installed tools are not updated too (depends on the setting). If yes, we lose the reproducibility. We have deleted that information.