Reproducible Bioinformatics Analysis module

BIOS-IN5410/9410 - Day 3

Lex Nederbragt

November 17, 2025

Program for today

- Recap day 2
- Collaborating through git and GitHub
- Wrap up: Reproducible Bioinformatics Analysis

Recap: a basic workflow

- Make changes to file(s)
- Inspect your changes with `git diff`
- Stage (some of) your changes with `git add`
- Commit staged changes with `git commit -m`

Hands on: `git` and Github continued

Follow the instructions on GitHub.

Wrap up: a basic workflow

- Update your local repo with `git pull origin main`
- Make your changes and stage them with `git add`
- Commit your changes with `git commit -m`, and
- Upload the changes to GitHub with `git push origin main`

Collaboration tips

- Be sure that you have an updated version of the repository you are collaborating on
    - do a `git pull` before making changes
- Better to make many commits with smaller changes, rather than of one commit with massive changes
    - small commits are easier to read and review

Advanced collaboration

- **Propose** changes
  - Instead of **making** changes
- This is called “submitting a Pull request”

What not to put under version control

- Raw data (these should not change)
- Intermediate files (usually)
- Result files that can easily be regenerated
  - Exception: if data and results are small
- Large files
  - Megabyte-sized and larger
  - Github limit: 100 MB

Collaboration and sharing

- github.com
  - Free or paid
  - Private repositories
    - With or without collaborators
  - Public repositories

Collaboration and sharing

- github.uio.no
    - Free for users with a UiO account
    - Private repositories
        - With or without collaborators
    - Public repositories
        - uio login required (!)

The three “Why’s” of version control

1. Reproducibility

- For your future self
- For your lab-mates and collaborators
- For the person who accuses you of making up your results…

The three “Why’s” of version control

2. Fixability

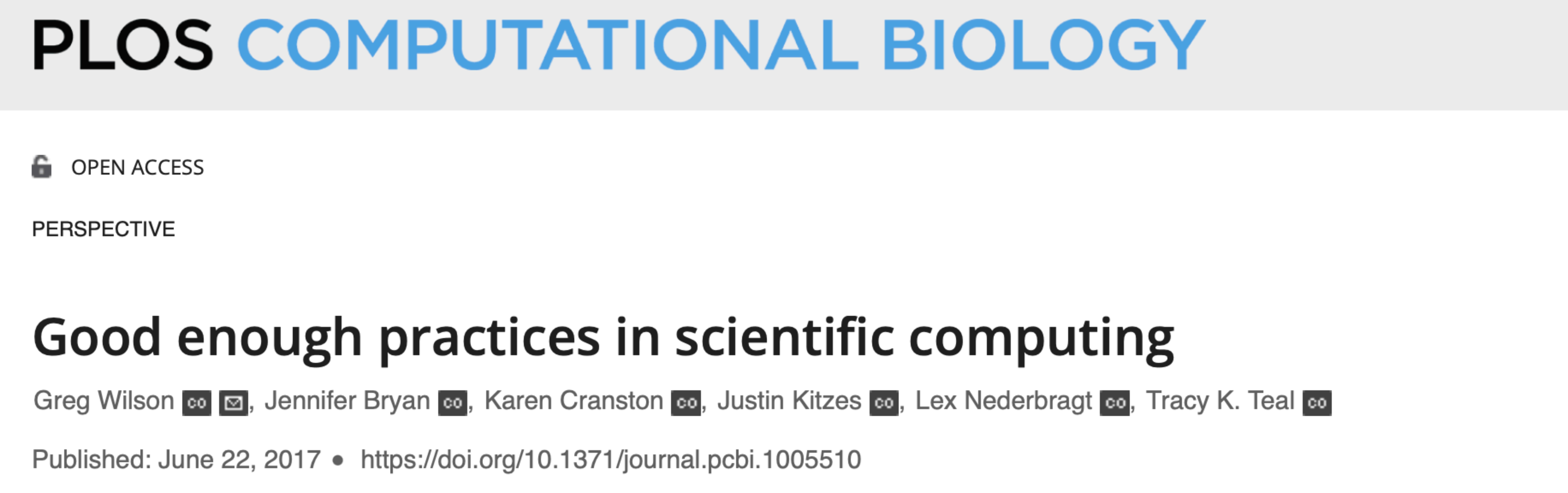
- Why does figure 4 look different now from last week?
- “Version control makes you fearless”

The three “Why’s” of version control

3. Sharing and Collaboration

Makes it easy to share projects and update them, sometimes simultaneously, with other collaborators

Example



PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

PERSPECTIVE

Good enough practices in scientific computing

Greg Wilson, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, Tracy K. Teal

Published: June 22, 2017 • https://doi.org/10.1371/journal.pcbi.1005510

Written in markdown, using git and github

Good enough practices in scientific computing

1. Data Management
2. Software
3. Collaboration
4. Project Organization
5. Keeping Track of Changes
6. Manuscripts

Reproducible Bioinformatics Analysis

- Example: blast

Reproducible Bioinformatics Analysis

Resources

- Wilson et al., “Good enough practices in scientific computing”: https://doi.org/10.1371/journal.pcbi.1005510
- Wilson et al. “Best Practices for Scientific Computing”: https://doi.org/10.1371/journal.pbio.1001745
- Noble, “A Quick Guide to Organizing Computational Biology Projects”: https://doi.org/10.1371/journal.pcbi.1000424
- Sandve et al. “Ten Simple Rules for Reproducible Computational Research”: https://doi.org/10.1371/journal.pcbi.1003285

For the coming days

- Writing something in a text file: commit it to version control