Bioc Technical Advisory Board Minutes

6 April 2023

Attending: Vince Carey, Stephanie Hicks, Levi Waldron, Laurent Gatto, Lori Shepherd Kern, Charlotte Soneson, Mike Smith, Michael Love, Bob Shear, Jennifer Wokaty, Alexandru Mahmoud, Marcel Ramos, Aedin Culhane, Robert Gentleman
Regrets: Wolfgang Huber, Sean Davis, Davide Risso, Kasper Hansen, Shila Ghazanfar, Rafael Irizarry

:03 - :10 TAB/project administration

- Minutes from previous meeting approved
- Release of 3.17 is moving along. 3.18 must allow time for introspection and reduction in technical debt. Bob Shear has analyzed our AWS presence and found many under-maintained components. A bitwarden vault has been produced to manage privileged information and two-factor authentication protocols have been introduced. An approach to articulating and managing SOPs is in place; these are all still private.
- Two big objectives of the 3.18 development cycle are a) improved automation, b) renovation of general annotation resources.
    - AnnotationForge org.Xx packages (in AnnotationHub) - produces packages for 1000 organisms, software is very old, very slow and not fault-tolerant. Meet with Jim and Marc to think of strategies. Re-evaluate the pipeline that is run for a release, and what information is provided. E.g. newer APIs may be available.
- Of general interest, a supplement opportunity, software tools for open science.
- Proposal for the core to have a reading activity starting with the flamingo book (free online version) [Thanks to Bob Shear for the reference.]
- Consolidating slack communications: Nice table/slackr work from Maria Doyle that should help us reduce the number of channels and prospectively guide discussions to proper location.

:11 - :30 Book production

- Jacques Serizay (Pasteur Institute) contacted Wolfgang and Vince - he is developing 'Orchestrating Hi-C Analysis with Bioconductor' (https://jserizay.com/OHCA/)
- Suggestion to have editions of books with pinned versions of all Bioc and CRAN package in use, the author is responsible for managing transitions. Docker?
- Mike Smith is managing the MSMB book, Stephanie/Lukas the development of the OSTA book.
- OSCA - Aaron proposed handing maintenance to the core, a team was assembled to manage it.
- To date we don't have guidelines on book production or principles on what these are in the project. There was an offer for hardcopy of OSCA material for ISMB but several remarks on slack indicated that hardcopy would be undesirable.

- The solution put in place for OHCA involves Docker, GitHub Actions, the GitHub container registry, and GitHub pages. Hi-C is "methodologically less volatile" than single cell, which may make maintenance somewhat easier. However, maintenance effort with respect to changes in upstream dependencies is "long-tailed" - average stability estimates may not be sufficient to manage risk.
- Should we have a "book working group"? Technical definition, standards, interop/interdependency, ease of access need to be hammered out.
- Comments:
    - The idea of having the books is good, but more formal guidelines/instructions on what it means to maintain the book and develop new content is needed.
    - Independent chapters vs monographs (cross-references easier with a monograph, but becomes less modular - e.g. for deprecating a single chapter, or if one part is failing). Cross-references in OSCA implemented via the rebook package.
    - Packages developed for testing integrity of static outputs (e.g. images) - https://vdiffr.r-lib.org/
    - In MSMB, each chapter can be built independently (builds on GitHub Actions every 2 weeks). Uses R Markdown/quarto options for referencing across chapters. ~250 dependencies, at least one chapter fails perhaps every 6 weeks.
    - Resources listed under bioconductor.org/help - for some of these, we welcome contributions, for others, we may not be ready. Which resources are kept up-to-date, which are static. Incentivize community members to contribute. Proposal (under development) for modification of the text in this box: bioconductor.org slash help
    - Could we have a feedback system - users could tag whether something is working, not working, useful, outdated, etc. Not limited to books - would be useful to all packages, documentation, vignettes, etc (the same is true for making sure that results/figures are still consistent with the text in vignettes).

:31 - :35 Cloud working group update

- Met on March 21 to discuss new ideas and developing standards.
- Vince also shared the idea of telemetry to guide cloud infrastructure selection and personas involved with cloud technology.
- This led to a discussion on benchmarking packages to develop infrastructure recommendations to users and distributing the recommendations through a template.
- The next meeting will be at 12PM ET April 18. Erdal will share a test environment for benchmarking. He will also share cloud security issues as a starting point to develop a cloud security roadmap. We'll also see a demo of rworkflows.

:35 - :40 Volume of package review

- Volatile, but please see the issue tracker
- Estimate of time from submission to acceptance? Extensive analysis was done here: https://llrs.dev/post/2020/07/31/bioconductor-submissions-reviews/

:41 - :45 Discussion about whether a function should return an object or directly save results to disk.

- Should we add a new guideline element?
- Should BiocCheck identify write-to-disk events and warn about them?
- Tasks that can easily fill a disk are becoming more common.
- How many existing Bioc packages already write to disk?
- Issue would be if a function writes to disk without the user being aware (needs to be a user-controlled option, or provide a message).

:45 - :47 GPUs

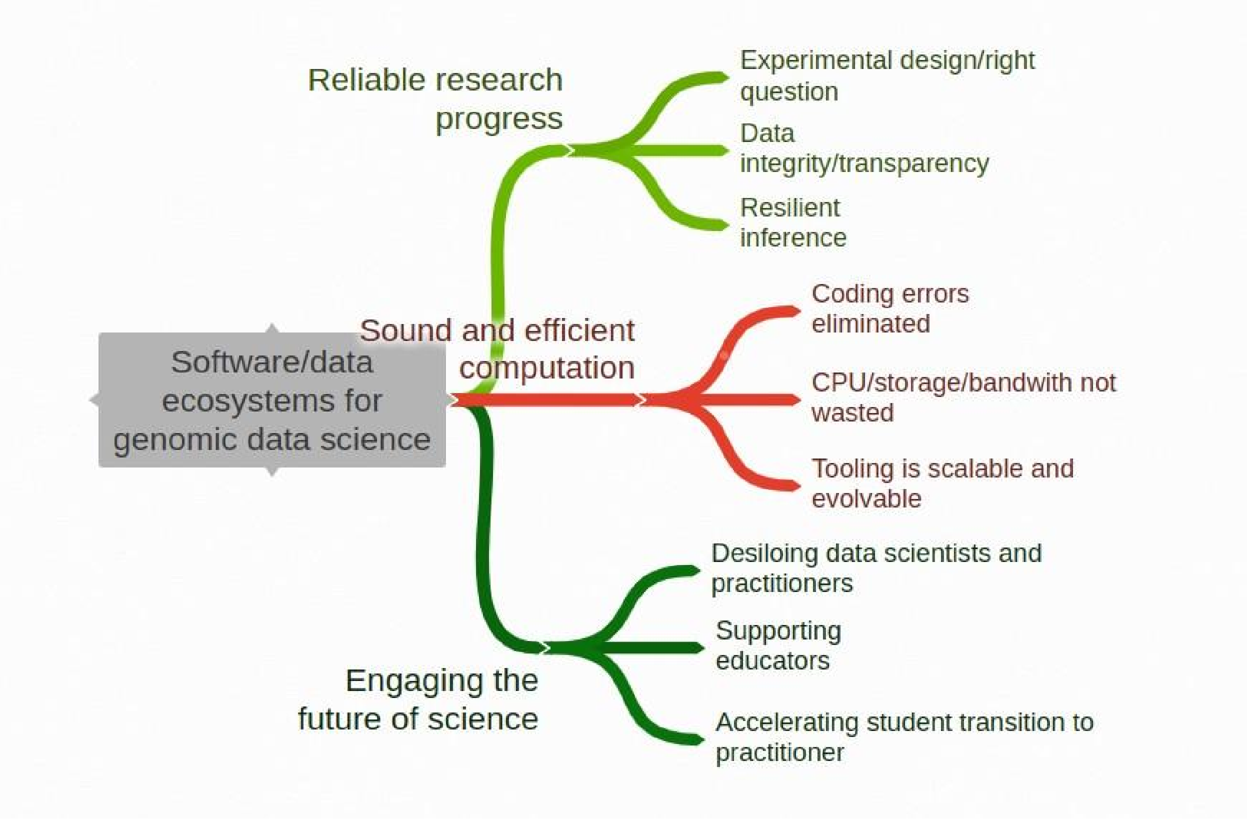
- Are TAB members regularly using functionality that would be helped by access to GPUs?

:48 - :50 Expiring member terms, new member elections

- Four terms end this year.
- TAB exec terms also expire.
- Should timing of nominations be moved closer to the conference? The CAB is planning this from next year.

:51 - :60 Open discussion

- alabaster for language-agnostic data representations.
- RBioFormats blowing up memory with qptiff inputs.
- Schematic for thinking about values of the ecosystem:



- The bicycle principles for short-format training in genomic data science
    - Evidence collection - test 'before/after workshop' knowledge. Something we should consider for the conference workshops?
    - Regular offering of an introductory workshop.
- Suggestion for the BioC2023 'Meet the TAB' session - bring up some of the 'difficult/thorny problems' for discussion with the community
    - books
    - technical debt
    - offloading the core team
    - annotation data
- Bioc is participating in Galaxy Training Network Smorgasbord 2023