Bioc Technical Advisory Board Agenda

7 November 2019

Attending: Martin Morgan, Vincent Carey, Aaron Lun, Sean Davis, Kasper Hansen, Laurent Gatto, Levi Waldron, Matt Ritchie, Michael Lawrence, Rafael Irizarry, Aedin Culhane, Wolfgang Huber, Charlotte Soneson
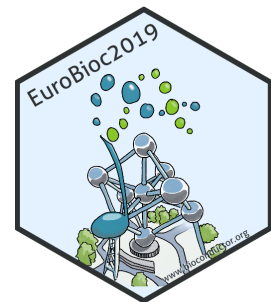Regrets: Stephanie Hicks

Schedule

:00 - :05 Minutes and action items from previous meetings
- Vote to approve minutes.

:05 - :25 Material for review
- Funding
    - U41: (NHGRI; core) Renewal due in January
    - U24: Monthly coordination meetings; private slack channel
    - AnVIL: https://anvilproject.org: bi-weekly bioc meetings, public slack; weekly tech calls; 'developer' days in mid-December
    - CZI seed network: monthly meetings, private slack
    - Opportunities?
        - CZI funding proposal (on slack)?
- Events
    - BiocAsia: Sydney Dec 5-6.
    - BiocEurope: Brussels Dec 9-10. Active planning slack and logo!
    - Bioc2020: Boston July 29-3. Public slack. Monthly planning meetings, active planning slacks, lots of participation!
        - Bioc2021: interest from Marc Carlson / Seattle
    - CSAMA 21-26 June 2020
    - Other
        - Meetups, etc
        - BIRS hackathon, Banff, 2020 (Aedin Culhane, Elana Fertig, Kim-Anh Lê Cao); slack channel
        -
- Committees
    - Code of Conduct Committee (Levi, with Stephanie)
        - Researching possible improvements to CoC and procedures, no updates
    - Conference (Levi):
        - BioC2020 in Boston, July 29-31; website available.

- in progress: website, workshops build system, invited speakers. Venue secured. Greatly expanded organizing committee since last open call.
            - "Mark your calendar" announcements coming up
        - Emerging Topics (Vince):
- Working groups
        - Developer forum: third session (Stephanie discussing *Hub and serialization formats; open discussion). [YouTube](); 24 people listened for 1361 minutes.
        - Build system / continuous integration (Vince): [slack]() activity
- Other:
        - Bioconductor Foundation Seed Funds (Stephanie / Vince)
            - [Stephanie] a rough draft (here) of the idea. Vince and Martin -- I believe I responded to your comments and my understanding of where we are is that the executive board would consider the idea. At the moment we have applications tentatively set to open Nov 15 (not sure if that's still the plan?) to be able to have a round of funding for this year before Bioc2020.
        - Europoean Bioconductor Foundation (Laurent / Wolfgang)
            - Consultation with EMBL legal services [response on-going], Torsten Hothorn (R Foundation)
            - Seems feasible; requiring commitment from a 'core' group
            - Intention: develop at BiocEurope
        - October release [announcement]()
            - 94 new software packages
            - Some emphasis on documenting builder configuration, so Windows / Mac devel delayed (should be available 'now')
            - Attention paid to production of annotation resources, with some initial changes scheduled
            - Relatively many changes in R-4.0 breaking more than average number of packages
        - Docker / container [discussion]()
            - Rationalize Bioconductor offerings around 'biocondcutor_full' (providing system dependencies) approach
            - Things that containers cannot do that a native installation could?

Topics

:25 - :35 BioC community survey (Matt)

- 107 respondents (as at 25th October 2019), mostly male (77.57%), post-docs (32.71%), from Europe (52.33%), dry lab researchers (80.37%), based in academia/university (87.85%), using R for 5-10 years and Bioconductor for 2-5 years (37.38%), use base R (84.11%), tidyverse (39.25%), Rstudio (84 responses) and Python (63 responses).

- Top 3 uses of Bioconductor reported were for bulk RNA-seq (82 responses), single cell genomics (52 responses) and DNA-seq data (41 responses). Help found via vignette/manual (46 responses) or support site (33 responses)
- Satisfaction levels seemed high across most areas (BioC training material was lowest scoring, with a tie between 'OK' and 'Satisfied').
- Barriers to contributing: S4 is mentioned a few times. Best aspects of BioC: community occurred many more times under this section than in the 'Worst aspects,' which is good.
- New activities. Hackathons
- Most respondents had not heard of or used ExperimentHub.
- Things to potentially improve: ways to increase female participation, reach more wet lab researchers, S4 training material to help new developers.
- More complete response summary (may need to select an HTML viewer)
- TODO: Formulate next steps — Matt

:35 - :45 Advisory Boards (Martin)

- TAB membership term limits
    - Goal is ⅓ re-election per year; remember, individuals can stand for re-election
    - Soliciting interest from current members who would rather participate in other aspects of the project / feel their efforts would benefit other activities / etc. Contact Martin.
- Community Advisory Board
    - Vote: Approve draft governance document
    - Identify TAB / Executive nominees this meeting / month.
    - Executive takes next steps to formalize membership
- Scientific Advisory Board membership will be updated, with input from stakeholders

:45 - :59 Discussion: core team technical priorities

- Possible directions to seed discussion
    - Annotation / Ensembl / BiomaRt strategies
    - Goals for containerization & cloud computing
    - Strategies for making Bioconductor technical contributions more impactful
    - Interoperability, especially with Python (RStudio / reticulate; Aaron's basilisk discussion as reference)
    - Specific areas of technical need
- Discussion:
    - [Kasper] Developer documentation for HDF5+DelayedArray in a SummarizedExperiment. We also need user documentation but developer is more critical. Too few people understand the complexities and we have substantial hope resting on this development. [Kasper]
    - [Kasper] Web interface to browse AnnotationHub + ExperimentHub

- [Stephanie] -- Do we think Bioc should be thinking much more seriously about other on-disk data representation formats beside HDF5? For some reason I wasn't aware of two (potentially major) problems of using HDF5 files: (1) limited parallel read/write access and (2) not designed for distributed computing -- so you need local files (potentially many copies of them) for distributed computing. To be fair, I have only used HDF5 locally (on a cluster or downloaded via ExperimentHub), but this is relevant for the world of spatial transcriptomics where raw images can be petabytes and a lot of distributed computing is used from what I understand. I know Bioc has [LoomExperiment](#) and [ZarrExperiment](#), but it's unclear how much bioc community support there is for either of these or even stability (considering the latter is only only on Github and says still under construction). However, there seem to be problems with e.g. zarr too (it can result in large number of individual files for very large arrays, so can cause issues on legacy file systems that can't handle large file numbers). Maybe this deserves a larger discussion on #bigdata-rep?
- [Aedin] Spatial transcriptomics/proteomics object class. Recently starting working on MIBI-Tof, mass-tags data.
- [Kasper] That the build system makes config.log available
- [Levi, Kasper] ExperimentHub: making it easier to develop and update packages
- TODO: secondary 'ad hoc' discussion to identify and summarize priorities / next steps, reporting at next TAB meeting