Bioc Technical Advisory Board Minutes

1 December 2022

Attending: Vincent Carey, Sean Davis, Hervé Pagès, Lori Shepherd, Robert Shear, Michael Love, Laurent Gatto, Charlotte Soneson, Marcel Ramos, Stephanie Hicks, Alexandru Mahmoud, Kasper Hansen, Davide Risso, Rafael Irizarry
Regrets: Wolfgang Huber, Levi Waldron, Robert Gentleman, Aedin Culhane, Shila Ghazanfar

:03 - :10 Approval of prior minutes; welcome to new Core Dev Robert Shear (DFCI).
- Quiz question: who maintains the named packages? Graphic source.
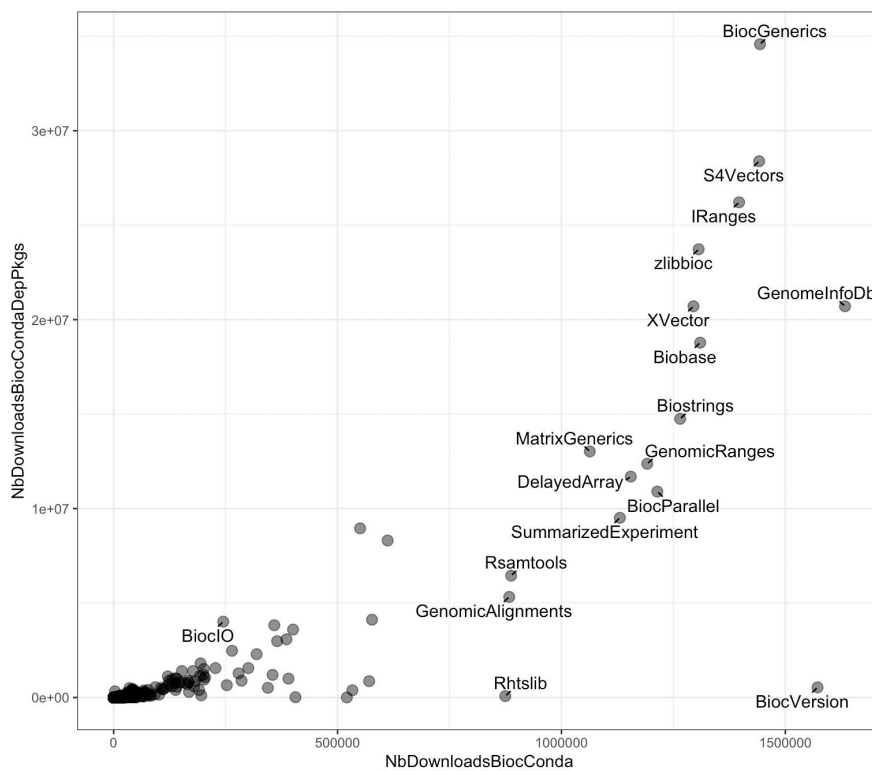
:10 - :13 RJournal submission
- As each R Journal issue approaches its deadline, we are invited to provide news from the project.
- The formatting/submission process has been simplified through the rjtools package.
- Sources for the article to be submitted this week are in R markdown. I would be happy to include more information on grants that TAB members or other Bioconductor contributors have received, other educational activities, etc. Please file PRs to the link given here if you would like more information to be added to the R Journal submission.
- To produce the PDF of the submission using rjtools, the command is
  `render("bioc.Rmd", rjournal_pdf_article())`. I encounter pandoc failure

when attempting to render to HTML in this way, but within a recent RStudio instance, the knit to rjournal web button succeeds.

:14 - :15 Governance and possible replacement of Robert Gentleman seat with an invited working group member.

:15 - :20 Website working group formed.
- Has anyone tried [cranlike](#)?
- Initial meeting: Decide on priorities and work towards smaller goals rather than complete initial redesign.
  - Priority is landing page redesign - too much information currently.
  - Functionality/search based on biocViews.
  - Better searchable site (eg. most recent first, sub searches and filters). Suggestion to exchange notes with Mike Smith, author of https://code.bioconductor.org/search.
  - Focus on release (potentially have a separate domain for devel) - especially for new users, how to get started.
  - The website should serve the wide range of users.

:20 - :23 Hacktoberfest (summary by Charlotte).
- [Hacktoberfest](#) concept: Have 4 pull/merge requests accepted between October 1 and October 31 to complete Hacktoberfest. The first 40,000 participants (maintainers and contributors) who complete Hacktoberfest can elect to receive one of two prizes: a tree planted in their name, or the Hacktoberfest 2022 t-shirt.
- [Slack announcement](#).
- Two meetups were held (Kevin & Kozo).
- [PRs for the bioc-project lesson in the Carpentries incubator](#) (please feel free to contribute as well!). Having well-defined issues in the GitHub repository is very helpful in order to solicit contributions (also unit tests).
- [Blog post](#) about awesome lists.

:23 - :40 Common classes and endorsed classes (thanks to Laurent Gatto for these details):
- Working group: http://workinggroups.bioconductor.org/currently-active-working-groups-committees.html#recommended-classes-and-methods
  - Current members: Lori Shepherd (lead),Vince Carey, Laurent Gatto, Johannes Rainer.
  - Invite: Lluis Revilla (has asked about this on slack) and Hervé Pagès.
  - Advertise on the bioc-devel mailing list and slack.
  - New WG repo: https://github.com/Bioconductor/BiocClassesWorkingGroup
- Discussion:
  - What are the 'official' classes (https://contributions.bioconductor.org/important-bioconductor-package-development-features.html#commonclass)?

- - ○ To what extent should these be enforced during package review? "In general, a package will not be accepted if it does not show interoperability with the current Bioconductor ecosystem."
    - ○ If not strictly enforced, at least require a wrapper function to convert to these?
    - ○ Procedure to establish a new 'official' class.
  - ● Comments:
    - ○ Strict enforcement may not be a good idea (new classes/innovation is needed at times). However, in many cases existing classes are enough.
    - ○ Encourage knowledge of/enhancement of existing classes.
    - ○ Articulate principles. Why is e.g. SummarizedExperiment important/valuable? Understand benefits obtained by using existing classes.
    - ○ Many successful innovations still inherit from existing classes.
    - ○ Possible use-case: BiocSet vs GSEAbase vs … for gene sets and enrichment analyses.
    - ○ Many submissions using data frames - could often use GRanges or other core classes instead.
    - ○ If a package is not using any of the Bioconductor classes, why submit to Bioconductor (and not CRAN)?
    - ○ Hard to know the long-term value of a package at submission time.
    - ○ Not just about classes, also about using the infrastructure/project in general.
    - ○ Think about the class system - why don't people use them? Need to make sure that the value is clear - what are the benefits?
    - ○ "glossy" - why be part of Bioconductor in general? What do you get out of coming into the system.

:40 - :45 Strictness of R CMD check/BiocCheck "--as-cran" – various remarks from Henrik Bengtsson.
- ● Relevant slack posts/GitHub issues:
  - ○ https://community-bioc.slack.com/archives/CLUJWDQF4/p1601057413025000
  - ○ https://community-bioc.slack.com/archives/CEQ04GKEC/p1622069560041900
  - ○ https://community-bioc.slack.com/archives/CEQ04GKEC/p1656372029195499
  - ○ https://community-bioc.slack.com/archives/CEQ04GKEC/p1666571545883419
  - ○ https://community-bioc.slack.com/archives/CEQ04GKEC/p1666617812557399
  - ○ https://community-bioc.slack.com/archives/CLUJWDQF4/p1584493059086500
  - ○ https://github.com/Bioconductor/BBS/issues/242
  - ○ https://github.com/HenrikBengtsson/Wishlist-for-R/issues/16
- ● Documentation of R CMD check --as-cran:
  - ○ https://cran.r-project.org/doc/manuals/r-devel/R-ints.html#Tools
- ● Most checks are already in place on the build machines.
- ● Shared R installation - packages can pass even if they don't declare all dependencies used in vignettes, tests etc, if some other package declared the dependency. Want to avoid that - ask R CMD check that tests, vignettes etc can be run with the declared dependencies. Enable for the devel branch only for now. May not work properly on Windows (need to test).

- The SPB manually sets many environment variables on submission (https://github.com/Bioconductor/packagebuilder/blob/master/check.Renviron).

:45 - :47 Should we submit a tutorial to ISMB/ECCB 23-27 July 2023 in Lyon, France (BioC2023 is 2-4 Aug): https://www.iscb.org/ismbeccb2023-submissions/tutorials
- Part of effort to reach out to e.g. the scverse team (maybe independently of tutorial - something like a retreat more effective?)
- Virtual tutorial is a possibility.
- Joint tutorial or satellite meeting to get groups together across languages.

:47 - :60 Large data resources: contribution and distribution best practices - OSN - Imaging/Zarr/TileDb.
- Do we want to go into the image analysis of very large-scale imaging data? Technology dependent, lots of engineering involved.
- Keep working on resources for describing best practices of analysis of large-scale data - show that Bioconductor is applicable in the area.
- Beth Cimini (Broad) will be a keynote at BioC2023 and is interested to hear how Bioc wants to integrate with image analysis pipelines. Learn from people working in the field.
- Some efforts are already being done to make large imaging data available (Ludwig Geistlinger). Also CZI is working on making data open.