

Bioc Technical Advisory Board Minutes

6 June 2024

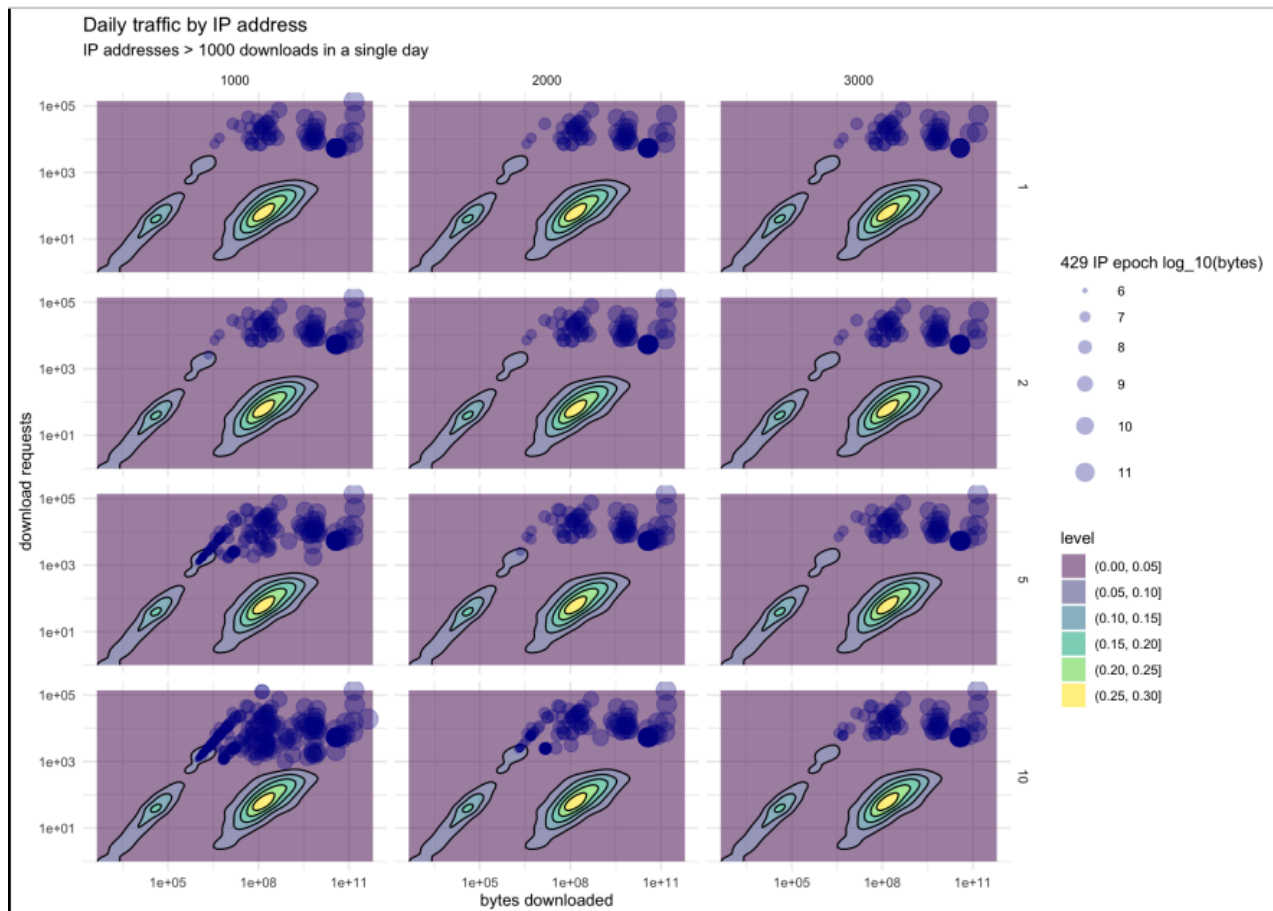
Present: Vince Carey, Laurent Gatto, Lori Kern, Charlotte Soneson, Rafael Irizarry, Kasper Hansen, Wolfgang Huber, Erdal Cosgun, Helena Crowell, Levi Waldron, Michael Love, Hervé Pagès, Marcel Ramos, Alexandru Mahmoud, Robert Shear, Henrik Bengtsson, Ludwig Geistlinger, Davide Risso, Stephanie Hicks, Jen Wokaty

Apologies: Sean Davis

:03 - :04 [Previous minutes](#) approved

:04 - :17 Cloud working group update

- Discussion of Cloudfront data outflow costs. 52TB shipped in April, 44TB in May. Monthly cost to us for this service, at least \$2000, not originally budgeted.



- Mitigation concepts
 - Move very large packages for BSgenomes and SNPlocs to egress-free storage, access via AnnotationHub.

- Introduce throttling for very high-volume consumers.
 - Hadley has made an offer to provide a distribution channel.
- Is this a question of package dependencies (i.e., is there a large package that is responsible for most of the volume)?
 - We could provide e.g. individual chromosomes in addition to the whole genome (e.g. for testing/example purposes).
 - Most of it appears to be retries, 'honest mistakes', but not easy/timeconsuming to pinpoint exactly what is the cause.
 - A large fraction of the volume is from GitHub Actions.
 - BiocManager could detect if the installation is done via a CI platform and redirect to a different download source.
- Can we block IPs (or introduce throttling)? Blocking specific packages would be harder and more disruptive.
- Making sure to use caching on GitHub Actions would help (but requires users setting this up correctly).
- It could unfortunately also be a matter of malevolent adversaries.

:17 - :22 Refactoring rtracklayer.

- Rtwobitlib is on CRAN.
- rtracklayer has lots of reverse dependencies, and does a lot of things.
- Import/export different types of files (gff, 2bit, bigWig, ...).
- Interact with UCSC genome browser (https requests, parse html responses). Susceptible to changes on the UCSC side.
- New UCSC.utils package with low-level functions interacting with the UCSC genome browser via the REST API. Will be used in GenomicFeatures, txdbmaker, ...
- Import/export functionality would be easier to maintain if it was broken up into several smaller packages, each dealing with a specific file type.

:22 - :29 Mambaforge issue for basilisk.

- Conda "defaults" channel should not be used as it includes packages with "no commercial use" licensing.
- Transitioning to alternate forge is complex as silent installation on windows seems to have fallen away in recent distributions.
- Also <https://github.com/rstudio/reticulate/issues/1583> is interesting regarding virtual environments and reticulate::install_python. Can we simplify python dependency management?
- Separate environment for each package/version vs possible shared environment/python stack. Package version conflicts will be difficult to manage with the latter setup.

:29 - :31 Zarr v3 in python finalizing mid July

- How should we interact with the python Zarr?
- Would wrapping the python package avoid the requirement for a delayed backend / some kind of R representation?

:31 - :33 Community/technical topics

- EMBL-EBI looking for a group leader:
<https://www.ebi.ac.uk/research/research-group-leaders-at-embl-ebi>. They recruit on potential, not past achievement.
- Migrating support site from Biostars to Discourse. Important because consultative support for our BioStars fork is no longer available.
 - Some hiccups with the current site currently (e.g. lack of pagination beyond the main page).
 - No objections to the switch from the board.
 - Get a supported migration, where current content will be moved over to Discourse - don't want to lose current questions/answers/reputation.

:33 - :50 Governance

- Discussion of proposed revisions to governance document.
- An important consideration is the approach to working with NumFOCUS to facilitate funding of small projects.
- TAB membership: attendance, nomination, voting considerations. How to increase/ensure diversity in the board?
 - Diversity is not unidimensional.
 - Taking diversity into account in voting, reserving spots to represent various groups.
 - NIH asks applicants to provide information that helps understand how their participation aligns with the purpose to promote diversity. We currently also ask the applicants to comment on how they will contribute to the diversity of the board.
 - Specify concrete goals in terms of diversity.

:50 - :60 Other

- Last day to vote for BiocAwards is tomorrow June 7th. If you need the voting form please contact Lori
- Build system
 - Currently 2 builds per week (but should increase to daily for devel soon).
 - Can we report when the next snapshot will be taken?