Bioc Technical Advisory Board Minutes

1 April 2021

Attending: Vince Carey, Levi Waldron, Charlotte Soneson, Nitesh Turaga, Lori Shepherd, Maya Reed McDaniel, Kayla Interdonato, Michael Love, Erdal Cosgun (guest), Kasper Hansen, Shila Ghazanfar, Laurent Gatto, Sean Davis, Hervé Pagès, Stephanie Hicks, Rafael Irizarry, Aedin Culhane, Robert Gentleman, Aaron Lun, Hector Corrada Bravo, Wolfgang Huber, Marcel Ramos
Apologies: Martin Morgan

Meeting agenda:

:01 - :03) 2021-03-04 minutes approved

:03 - :10) Introductions for Erdal Cosgun Ph.D., Microsoft Genomics

:10 - :30) Erdal Cosgun presentation on Microsoft Genomics Data Lake and Azure support for Bioconductor infrastructure activities

- Microsoft genomics solutions: Research & Discovery (ML/AI, Bioc on Azure, Genomics data science VM, Genomics Data Lake), Automation & Scale (Cromwell on Azure), Clinical Genomics (Microsoft Genomics service on Azure)
- Working with Broad and Verily on Terra
- Mirror Bioc Docker image on MCR (4.9K pulls)
- http://bioconductor.org/help/docker/#msft
- Azure R/Bioc data science VM - customized with genomics content (R/Bioc, GATK, …). Two options: Windows Server 2016/Ubuntu 18.04. Can be customized with other content. Discussion at https://social.microsoft.com/Forums/en-US/home?forum=dsvm
    - https://github.com/microsoft/genomicsnotebook
    - Azure account needed for deployment of VM
- Genomics Data Lake (Illumina Platinum Genomes, ClinVar, gnomAD, 1000 Genomes, ENCODE, …). Could also share Bioc datasets here.
    - For each resource - landing page + SAS Token
    - Tool for exploring the data sets (don't need Azure account) "Azure Storage Explorer"

:30 - :60) Q&A for Erdal/discussion

- Linux VM built in an iterative process (adding Bioc packages successively), Windows binaries all built on the Bioc build system - how to build the Windows binaries similarly to the Linux approach. Should be the end goal.
- Compiling OSCA book on Azure would also be good.
- Workshop material/Orchestra
    - 'cloud doesn't change anything on how we do builds', just changes the infrastructure, not necessarily any new capabilities.

    - most users are not cloud-savvy - even getting an Azure account can be a barrier in itself. Can we provide a simpler 'gateway' so users won't necessarily recognize that they are using a cloud service?
- Difficult to budget for cloud usage currently (in grants etc)
- Another common complication/barrier: how to get files 'in' and 'out' of the cloud storage in an intuitive manner? Impediment to more 'ad-hoc' usage of cloud service.
    - Storage Explorer -> Storage accounts in Azure more accessible compared to other cloud services
- What could be hosted on Azure from the Bioc side?
    - Packages, Hubs
    - No strict need to copy - can be mirrored (but a copy is helpful for the VMs)
    - No specific migration plan/priorities yet
    - If Hubs would be moved to the data lake, what would be the impact on users?
- GitHub Actions can generate macOS binaries - can we generate macOS binaries in Azure?
- How can we take advantage of using VMs on the Azure cloud to run daily builds? In particular, on Windows - produce Windows binaries.
    - Can run builds in VMs in the cloud
    - Need to provide our requirements.
    - Can the builds be parallelized or do they have to be run on one machine?

Items of note:

- Core developer team has expressed an interest in attending TAB meetings and have been invited to this meeting.
- Bioconductor NHGRI U24 was awarded 12 March 2021. See the NIH reporter for details.
- Revisions to the organizational structure and detailed plans for release 3.14 are in development.
- Aedin contacted the NIH NHGRI to request if we could submit an R13 application to fund Bioconductor annual meetings. We have a call with the program officer Monday, April 5th at 3:30 – 4:00pm EST.
- Hardware difficulties with a deployment of the build system at BWH have led to delays in build system reporting. Additional redundancy is being introduced. Huge thanks to Hervé for adapting to these challenges.
- Three letter-of-intent submissions to CZI EOSS Cycle 4: Aedin and CAB, Vince "diversity followup for Cycle 1 project", Vince "genetics".

Greatest Hits March:

- Reports for checking status of reverse dependencies were added to the build reports (e.g. https://bioconductor.org/checkResults/3.12/bioc-LATEST/rhdf5/) by Hervé, and Mike Smith added a dependsOn argument to BiocPkgTools::problemPage(). More details on slack.
- 126 submissions received for BioC2021. Registration will open shortly.

- March Developer's Forum with the working group on object-oriented programming in R: https://www.youtube.com/watch?v=_QsFRiOBjt8. A new slack channel #r7 was created for further discussion.
- BiocManager now knows about the book repo, and books can thus be installed using e.g. BiocManager::install("OSCA.advanced").
- New chapters are being added to the OSCA books (see #osca-book slack channel).
- The #humble-brag slack channel was initialized - keep an eye out for more Bioc-related greatest hits.

CAB update:

- Levi presented slides on governance at last CAB meeting
- Elected 5 new members:
    - Photos on webpage.
    - Profiles of New CAB Members
      Daniela Cassol, University of California, Riverside, USA. website,
      Katerina Kechris, University of Colorado, Denver, USA. website,
      Estefania Mancini, Centre for Genomic Regulation, Spain. Google Scholar
      Kevin Rue-Albrecht, University of Oxford, UK. website.
      Mike Smith, EMBL, Germany. website
- CAB Bioc2021 awards.
    - Established subcommittee, creating page on Bioconductor website,
    - Will approve nominations form at April meeting. Do the TAB want input on form or to vote on nominees?
- Events
    - H3 Africa: Bioconductor will present 2 workshops (They requested two of Levi's BioC2020 workshops, Sean will provide Orchestra) at H3 Africa, April 26-30 2021.
- Funding.
    - Submitted (EOSS4-000000020) a Letter of Intent to the CZI EOSS Cycle 4. It would employ 2 people, a web developer and training & support officer, fund the virtual meeting platform (eg Airmeet), Carpentries membership and Sean's Orchestra. Vince Carey also submitted an aligned proposal (EOSS-0000000218).
    - Goals are "content" focused
        - redevelop the website; mobile-friendly, integrate Bioconductor web presence YouTube/web etc and include 'Kaggle style' continuous engagement
        - develop training course infrastructure; fund Carpentries membership and global instructor training, develop educational modules, and support closed captioning/translation services. The idea is to train and equip instructors in different countries so courses can be provided in local languages.