Docker containers for Bioconductor

Docker allows software to be packaged into containers: self-contained environments that contain everything needed to run the software. Containers can be run anywhere (containers run in modern Linux kernels, but can be run on Windows and Mac as well using a virtual machine called Docker Toolbox. Containers can also be deployed in the cloud using Amazon EC2 Container Service or other cloud providers.

With Bioconductor containers, we hope to enhance

- *reproducibility*: If you run some code in a container today, you can run it again in the same container (with the same tag) years later and know that nothing in the container has changed. You should always take note of the tag you used if you think you might want to reproduce some work later.
- *ease of use*: With one command, you can be running the latest release or devel Bioconductor. No need to worry about whether packages and system dependencies are installed.
- *convenience*: Sometimes you just want a fresh R with no packages installed, in order to test something; or you typically don't have microarray packages installed but suddenly you need to do a microarray analysis. Containers make this easy.

Our aim is to provide up-to-date containers for the current release and devel versions of Bioconductor, and some older versions. Bioconductor's Docker images are stored in Docker Hub; the source Dockerfiles are in Github.

Our release images are based on rocker/rstudio and built when a Biocondcutor Release occurs. Our devel images are based on rocker/rstudio-daily and built weekly with the latest versions of R and Bioconductor packages.

For each supported version of Bioconductor, we provide several images:

- *base2*: Contains R, RStudio, and a single Bioconductor package (`BiocInstaller`, providing the `biocLite()` function for installing additional packages). Also contains many system dependencies for Bioconductor packages. Useful when you want a relatively blank slate for testing purposes. R is accessible via the command line or via RStudio Server.
- *core2*: Built on *base2*, so it contains everything in *base2*, plus a selection of core Bioconductor packages. See the full list.
- *protmetcore2*: Built on *core2*, so it contains everything in *core2*, plus a selection of core Bioconductor packages recommended for proteomic and metabolomics analysis.
- *metabolomics2*: everything in *protmetcore2*, plus select packages from the Metabolomics biocView.

Current Containers

Maintained by the Bioconductor Core Team: bioc-issue-bot@bioconductor.org

- bioconductor/devel_base2
- bioconductor/devel_core2
- bioconductor/release_base2
- bioconductor/release_core2

Maintained by Steffen Neumann: sneumann@ipb-halle.de

Maintained as part of the “PhenoMeNal, funded by Horizon2020 grant 654241”

- bioconductor/devel_protmetcore2
- bioconductor/devel_metabolomics2
- bioconductor/release_protmetcore2
- bioconductor/release_metabolomics2

Maintained by Laurent Gatto: lg390@cam.ac.uk

- bioconductor/devel_proteomics2
- bioconductor/release_proteomics2

Legacy Containers

The following containers are legacy and no longer updated. They have been kept to retain previous versions available via tags:

- bioconductor/devel_base
- bioconductor/devel_core
- bioconductor/devel_flow
- bioconductor/devel_microarray
- bioconductor/devel_proteomics
- bioconductor/devel_sequencing
- bioconductor/devel_metabolomics
- bioconductor/release_base
- bioconductor/release_core
- bioconductor/release_flow
- bioconductor/release_microarray
- bioconductor/release_proteomics
- bioconductor/release_sequencing
- bioconductor/release_metabolomics

Using the containers

The following examples use the `bioconductor/devel_base2` container. Note that you may need to prepend `sudo` to all `docker` commands.

Prerequisites: On Linux, you need Docker installed and on Mac or Windows you need Docker Toolbox installed and running.

To run RStudio Server:

```
docker run -p 8787:8787 bioconductor/devel_base2
```

You can then open a web browser pointing to your docker host on port 8787. If you’re on Linux and using default settings, the docker host is `127.0.0.1` (or `localhost`, so the full URL to RStudio would be `http://localhost:8787`). If you are on Mac or Windows and running `Docker Toolbox`, you can determine the docker host with the `docker-machine ip default` command.

Log in to RStudio with the username `rstudio` and password `rstudio`.

If you want to run RStudio as a user on your host machine, in order to read/write files in a host directory, please read this.

To run R from the command line:

```
docker run -ti bioconductor/devel_base2 R
```

To open a Bash shell on the container:

```
docker run -ti bioconductor/devel_base2 bash
```

Note: The `docker run` command is very powerful and versatile. For full documentation, type `docker run --help` or visit the help page.

Modifying the images

There are two ways to modify these images:

1. Making changes in a running container and then committing them using the `docker commit` command.
2. Using a Dockerfile to declare the changes you want to make.

The second way is the recommended way. Both ways are documented here.

List of packages installed on the *core2* container

These packages, plus their dependencies, are installed:

- BiocInstaller
- OrganismDbi
- ExperimentHub
- Biobase
- BiocParallel
- biomaRt
- Biostrings
- BSgenome
- ShortRead
- IRanges
- GenomicRanges
- GenomicAlignment
- GenomicFeatures
- SummarizedExperiment
- VariantAnnotation
- DelayedArray
- GSEABase
- Gviz
- graph
- RBGL
- Rgraphviz
- rmarkdown
- httr
- knitr

- BiocStyle

Acknowledgements

Thanks to the rocker project for providing the R/RStudio Server containers upon which ours are based.

Contact us: support.bioconductor.org

- About