

Using docker for teaching

Peter Humburg

Published at *Genomic Campfire* on Mar 25 2015

Recently I gave an introductory workshop to eQTL analysis¹. I decided to use docker to provide a uniform working environment for everyone. This was my first serious use of docker and a somewhat mixed experience.

My main motivation for using docker was that the participants brought their own laptops for the practical part of the course and there was a risk that we would get bogged down with software issues. Providing a docker [docker image](#) with all the required software and most of the relevant data seemed like a good way to ensure everyone had the same software environment, essentially allowing me to deal with the configuration upfront.

The image is based on one of the excellent [Bioconductor images](#) which provides an RStudio server as well as a collection of Bioconductor packages. All I really needed to add were a few more R packages that were required for the course. I also used this image to run all the R code examples to ensure that they would work on the day.

That all sounds rather good in theory but unfortunately it wasn't quite so easy in reality. While preparing the workshop everything went pretty smoothly² and I felt confident that there wouldn't be any surprises with the examples I had prepared due to incompatible R package version or the like (this turned out to be the case). The first minor issue presented itself when I realised that downloading the image from DockerHub would take too long to do it on the day. Still, it was easy enough to instruct participants to install docker, or [boot2docker](#), and pull the image beforehand. While that presented a minor obstacle for some it certainly was easier than getting everyone to install the latest R version, RStudio and a list of packages. However, the fact that docker doesn't support Windows and OS X natively meant that the software environment wasn't quite as homogeneous as I had hoped.

The main issue arose because I had to provide additional data on the day. The final part of the course involved the use of a dataset for which I couldn't provide the genotypes publicly. So I asked the participants to download these data at the start of the course and provided [instructions](#) on how to make the data directory accessible from within the docker container. This didn't go well at all. Most people encountered some issue with this and some never managed to access the data on their machine. As a result we spent quite some time trying to sort out software configuration issues, the very thing I had hoped to avoid.

I still think that using docker images like this is a good idea and would do it again but I will certainly think twice about trying to share data with the host machine.

¹the material is available [here](#)

²apart from the fact that the docker builds occasionally failed because of issues with Debian package repositories but I don't think that is docker's fault