Debian adopts and disseminates Bioinformatics Open Source Software

Steffen Möller, Charles Plessy, David Paleino, Andreas Tille and the Debian Community*

Looking back

In Bioinformatics we are used to associate technological progress with the advancements in wet-lab techniques that bring us a steadily increased influx of more and more novel data to manage and interpret. Over that we often forget that this is only possible since the IT sciences have evolved even more quickly. This allows us to keeping pace with the data stream while applying even more analyses.

When the first Bioinformatics Open Source Conferences were held in the late 90s, the Internet was still a recent event. To find data on the net was considered special. And that data came at no extra charge. GNU/Linux had emerged as the ubiquitous operating system, as free as the data that was analysed with it. And free were most tools for sequence analysis, with development often funded by the same institutions that funded the wet-lab production of the data. Free also became the Bio{Perl,Java,*} libraries that help analysing the sequence data.

These libraries and many accompanying tools are now being used in many different suites for the handling of biological data. Or they are being used by smallish scripts to help with analyses in smaller or larger research projects. They became a commodity. One has gained sufficient confidence in the community to always want the latest versions of these helpers. Many analysts take the existence of these tools for granted or use them as part of a larger tool while not being aware of them. And those distributing software that is depending on common libraries or tools need ways to ensure a trustworthy installation of the basic research infrastructure. They can give instructions to their users to install everything themselves, can ship precompiled binaries or – suggest a GNU/Linux distribution's packages.

Today

GNU/Linux distributions live from their users. Commercial distributions have opened up for packages organised by the community, i.e. via OpenSuSE and Fedora. Debian GNU/Linux has been a community-driven distribution ever since and with around 50 programs or libraries (plus dependencies) it is the distribution which ships the largest number of bioinformatics packages for the largest number of platforms.

In 2001 Debian introduced a concept now called Debian Blends, a platform for the presentation of software packages for communities with a distinct interest. Bioinformatics is well kept under the hood of the Debian-Med blend with some packages also being found under Debian-SciComp or Debian-Science. Individuals interested to see a bioinformatics software packaged will send an email to the mailing list or fill out a Request for Packaging. Quite often it is a Debian packager amongst the developers of the software or an enthusiastic user, who seeks his self-prepared package to be shipped with the distribution.

Looking forward

A major problem in bioinformatics is the local maintenance of remotely accessibly data. Every group performing research in this field is solving this to their local needs in some way, crafting a series of scripts, but this effort should somehow be shared.

Where Debian, like all other Linux distributions, need help, and this also is particularly obvious in the complex interplay of software in Bioinformatics, is the guidance of users in the interplay of multiple tools. There is yet no package for the education of users or for pre-assembled workflows that address frequently observed problems. This issues was raised in past discussions in the context of the possibility to prepare a BOSC liveCD, which may also be imaginable as an interplay of a regular Linux distribution with a set of Wiki pages to guide the users.

Availability

The Debian home page is `http://www.debian.org`, development on bioinformatics packages is best monitored on `http://debian-med.alioth.debian.org`.

---

* to whom correspondence should be addressed: debian-med@lists.debian.org