

## **BioBuilds: A Model for Long Term Sustainability of Open Source Bioinformatics**

*Chris Mueller, Varshal Davé, Thomas Burnet, and Cheng Lee*

From the original FASTA software suite to BLAST and the Human Genome Project, and now with the proliferation of tools for next generation sequencing data analysis, bioinformatics has a long history of using Open Source software. For many bioinformaticians, Open Source tools are a natural extension of their normal workflows and development philosophies. More importantly, the robust Open Source community enables the creation and sharing of common tools and methods that benefit everyone.

However, as bioinformatics extends beyond academic and research applications and gains traction in diagnostic and applied markets, there is a stronger focus on using, rather than developing, Open Source tools. This is a great evolutionary step for bioinformatics that highlights how the field is maturing. To maintain relevancy, it is important that the Open Source community grow with the new opportunities.

On one end, simple challenges such as the time and effort required to maintain installed Open Source tools must be addressed. Many Open Source tools require extensive hands-on time to install, with build and dependency errors often derailing the process entirely for all but the most committed, impacting adoption and productivity. Seemingly simple updates can require days of effort.

The problem is compounded in regulated environments, where tools must have clear provenance and additional support for verification and validation protocols, leading to a high level of uncertainty around using Open Source tools at all. It is unrealistic to expect every hospital pathology lab to employ a software engineer simply to maintain a collection of tools.

At a higher level, many of the tools developed in the last decade to support sequencing and other 'omics applications are nearing the end of their value as research targets. For example, short-read alignment methods, up until recently a go-to topic for Masters and PhD students, have very little research value left in them. As the funding agencies wind down support for this research to focus their resources on new topics, support for many important tools will fade.

To begin to address these issues, we have launched the BioBuilds project. BioBuilds is distribution of Open Source bioinformatics tools suitable for deployment in research, commercial, and regulated environments. BioBuilds' primary mission is to ensure long-term support for Open Source bioinformatics tools across a broad range of research, diagnostic, and industrial applications.

As a distribution of tools, BioBuilds provides pre-built binaries for many Open Source bioinformatics tools that "just work" out of the box on a number of platforms. All tools are built and tested and the package includes all dependencies. With vendor support, BioBuilds also includes binaries optimized for specific platforms. Like BioBrew, CloudBioLinux, and other projects that provide packaged tools, BioBuilds believes that the first step in long term support is ensuring ready access to tools.

As an organization, BioBuilds goes beyond just providing tools. To support our mission of long term support, we are actively working on funding and sponsorship models to ensure the resources are available to provide continued support for Open Source bioinformatics tools. We are working closely with academic, industry and non-profit partners, including hardware and instrument vendors, diagnostic suppliers, and other scientific software foundations such as NumFocus, Boost, and OpenMPI, to develop a model that ensures continued resources and financial support will be available for these tools long after their research support has ended.

Bioinformatics has recently reached the point where its value is clear outside of academic and research pursuits. For Open Source bioinformatics to maintain relevancy as applied markets develop, it is essential to develop the mechanisms to ensure long term support is available for tools and develop healthy relationships with the industries that benefit from it. BioBuilds is a first step towards this goal.