Using *R* to Empower a New Plant Biology

Naim Matasci[1,*], Matthew Vaughn[2], Nirav Merchant[1,3]

1. The iPlant Collaborative, BIO5 Institute, University of Arizona
2. The iPlant Collaborative, Texas Advanced Computing Center, University of Texas at Austin.
3. Biotechnology Computing Facility, BIO5 Institute, University of Arizona
*Contact author: nmatasci@iplantcollaborative.org

The iPlant Collaborative is a U.S. National Science Foundation program to build a cyberinfrastructure for the plant sciences that will enable new conceptual advances through integrative, computational thinking. Teams composed of plant biologists and computer scientists work on grand challenges that address fundamental questions in plant biology, in particular focusing on elucidating the link between genotype and phenotype and on building a tree of life to represent the evolutionary history of all green plants.

Countless computer programs address individual aspects of these grand challenges but the complexity and sheer magnitude of the problem makes it almost impossible for individual researchers or research groups to tackle such problems alone.

iPlant provides a cyberinfrastructure that allows researchers in the plant science community to collaborate and share and integrate data and algorithms. Furthermore, by providing access to High Performance Computing resources it permits the generation and analysis of very large datasets as well as the usage of computationally intensive algorithms.

Given the growing importance of *R* in virtually every field of biology and bioinformatics, iPlant's cyberinfrastructure has been enriched by an interface to facilitate the development, execution and distribution of *R* scripts.

Developers can work on constructing new tools in Atmosphere, iPlant's virtual environment, through an IDE and access data and workspaces located in their iPlant home directory on in shared locations. All I/O operations are handled through iRODS, the Integrated Rule-Oriented Data System which permits the sharing of data and provides direct read access to plant genomes and other large files. The use of iRODS also ensures consistency of data across the different access point to iPlant's cyberinfrastructure (Atmosphere, API, Discovery Environment). Through iPlant's API users and developers can execute computationally intensive scripts on Texas Advanced Computing Center's HPC resources. Finally, by providing a simple metadata description, developers can create GUIs for their applications and make them available to colleagues and collaborators only or to the community at large through iPlant's Discovery Environment.

References

Atmosphere: http://atmo.iplantcollaborative.org/
iRODS: https://www.irods.org/
iPlant's Discovery Environment: http://preview.iplantcollaborative.org/de/