

PlantIT: Containerized phenotyping in the cloud

Wesley Bonelli¹, Suxing Liu¹, Chris Cotter¹, Megan Flory¹, Maria Luck¹, and Alexander Bucksch¹

¹University of Georgia

November 21, 2022

Abstract

Continuous collection and analysis of high-resolution phenotype data is critical to develop crops resilient to the consequences of climate change. Though web-accessible tools for parallel, reproducible scientific workflows render big data increasingly tractable, software for plant science remains inadequate for large-scale precision agriculture. Cyberinfrastructure must present minimal barriers to entry, accommodate rapidly changing dependencies, support a wide variety of use cases, and weave together sensors at the edge, laptops, clusters, and cloud storage into a coherent virtual workspace. PlantIT is a web portal intended as such an environment. Platforms like PlantIT and its precursor DIRT [1] permit efficient phenotyping and equip geographically distributed researchers with a code-optional interface. Workflows are published in Docker images, deployed as Singularity containers to public or private computing resources, and monitored in real time. Data are stored automatically in the CyVerse Data Store and can be annotated according to the MIAPPE [2] standard. GitHub integration provides versioning and repositories can be activated with a single configuration file, like Travis or GitHub Actions. Containers allow for a range of use cases, including image-based trait measurements, 3D reconstructions, morphological growth simulations, and crop modeling. Pseudo-batch/stream processing is also necessary; as data scales, manual batch jobs rapidly become infeasible, and (re-)analysis must occur upon arrival in near-real-time. We suggest web-accessible phenotyping automation software may address bottlenecks and help reveal undiscovered relationships between genes, traits, and the environment.

PlantIT: Containerized phenotyping in the cloud

Wesley Bonelli¹, Suxing Liu¹, Chris Cotter¹, Megan Flory^{1,2}, Maria Luck^{1,2}, Alexander Bucksch^{1,3,4}
¹Department of Plant Biology, ²Lamar Dodd School of Art, ³Institute of Bioinformatics, ⁴Warnell
School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602

ABSTRACT

Continuous collection and analysis of high-resolution phenotype data is critical to develop crops resilient to the consequences of climate change. Though web-accessible tools for parallel, reproducible scientific workflows render big data increasingly tractable, software for plant science remains inadequate for large-scale precision agriculture. Cyberinfrastructure must present minimal barriers to entry, accommodate rapidly changing dependencies, support a wide variety of use cases, and weave together sensors at the edge, laptops, clusters, and cloud storage into a coherent virtual workspace. PlantIT is a web portal intended as such an environment. Platforms like PlantIT and its precursor DIRT [1] permit efficient phenotyping and equip geographically distributed researchers with a code-optional interface. Workflows are published in Docker images, deployed as Singularity containers to public or private computing resources, and monitored in real time. Data are stored automatically in the CyVerse Data Store and can be annotated according to the MIAPPE [2] standard. GitHub integration provides versioning and repositories can be activated with a single configuration file, like Travis or GitHub Actions. Containers allow for a range of use cases, including image-based trait measurements, 3D reconstructions, morphological growth simulations, and crop modeling. Pseudo-batch/stream processing is also necessary; as data scales, manual batch jobs rapidly become infeasible, and (re-)analysis must occur upon arrival in near-real-time. We suggest web-accessible phenotyping automation software may address bottlenecks and help reveal undiscovered relationships between genes, traits, and the environment.

Keywords: phenotyping, data science, big data, HPC/HTC, containers, image processing, simulation

REFERENCES

- (1) Das, A., Schneider, H., BurrIDGE, J., Ascanio, A.K.M., Wojciechowski, T., Topp, C.N., Lynch, J.P., Weitz, J.S., Bucksch, A., Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant Methods*. 11, 51 (2015).
- (2) Krajewski, P., Chen, D., Ćwiek, H., van Dijk, A.D.J., Fiorani, F., Kersey, P., Klukas, C., Lange, M., Markiewicz, A., Nap, J.P., van Oeveren, J., Pommier, C., Scholz, U., van Schriek, M., Usadel, B., Weise, S., "Towards recommendations for metadata and data handling in plant phenotyping," *J. Exp. Bot.* 66, 5417–5427 (2015).