

## W115: GOBii, a Scalable Genomics Data Management System with Rapid Data Extract Times and Integration with Downstream Genomic Selection Analysis Pipelines

The Genomic Open-Source Breeding informatics initiative (GOBii) has built a genomics data management system that is highly scalable and has focused on data extract performance for large genomics data files. We have benchmarked several SQL and noSQL open-source data management systems with a view to managing large scale genomics data, and have determined that the HDF5 file system outperformed other data management systems both in loading and extract times. In order to also accommodate metadata management, we have designed and developed a hybrid system based on Postgres for sample and marker metadata management and HDF5 for the large genomics files. The HDF5 genomics files are stored in two different orientations to enable rapid extract in either sample or marker-fast formats. The system is flexible enough to be used across different crops and with diverse marker and sequence-based platforms. We are now working to integrate the genomics data extracts with downstream genomic selection applications in Galaxy.

## **Presenting Author**

Yaw A. Nti-Addae Genomic Open-Source Breeding Informatics Initiative

## Authors

Victor Jun M. Ulat International Maize and Wheat Improvement Center (CIMMYT)

Dave Matthews Cornell University

Guilhem Sempéré CIRAD, UMR Intertryp

Valentin Guignon Bioversity International

<u>Pierre Larmande</u> University of Montpellier,



Find Similar

Jon Renner University of Minnesota

Adrien Petel CIRAD

Elizabeth Jones Genomic Open-Source Breeding Informatics Initiative Project

Kelly Robbins Cornell University

## **View Related Events**

Day: Monday, January 14, 2019