# Catalyzing Genomic Discoveries: Leveraging rsIDs for Enhanced Trait-Driven Interoperability and Genomic Function Transfer

Sharon Wei1, Marcela K Tello-Ruiz1, Vivek Kumar1, Andrew Olson1, Kapeel Chougule1, Doreen Ware1,2

1 Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

2 USDA ARS NEA, Plant Soil & Nutrition Laboratory Research Unit, Ithaca, NY, 14853, USA

The Reference SNP cluster ID (rsID) is a unique identifier for genetic variations (GVs) at specific genome positions, enabling standardized referencing across databases and studies. While widely used in human research, its application in plant research has been limited due to insufficient support. However, the European Variation Archive (EVA) has now assigned millions of rsIDs to plant genomes, including key agricultural species like those in Gramene and SorghumBase. This development allows GVs to be identified by rsIDs rather than being tied to specific genome assemblies, simplifying data integration and enhancing marker-based breeding. Gramene has adopted rsIDs to consolidate genetic variation knowledge, improve phenotype prediction, and enhance trait-based marker discovery. Currently, rsIDs have been integrated into four crop genomes: Sorghum (41M), Rice (27M), Maize (78M), and Grape (0.3M). As more pan-genomes are sequenced, mapping rsIDs from reference genomes to these pan-genomes is more efficient than calling GVs for each accession. Using EVA's variant remapping pipeline, a 98% mapping accuracy was achieved across different assembly versions and 87% across Sorghum pan-genomes. This initiative not only advances plant genetic research but also promotes wider rsID adoption in agriculture, improving database interoperability and the effectiveness of genetic markers in breeding programs. Gramene's work is funded by USDA ARS (8062-21000-051-00D).