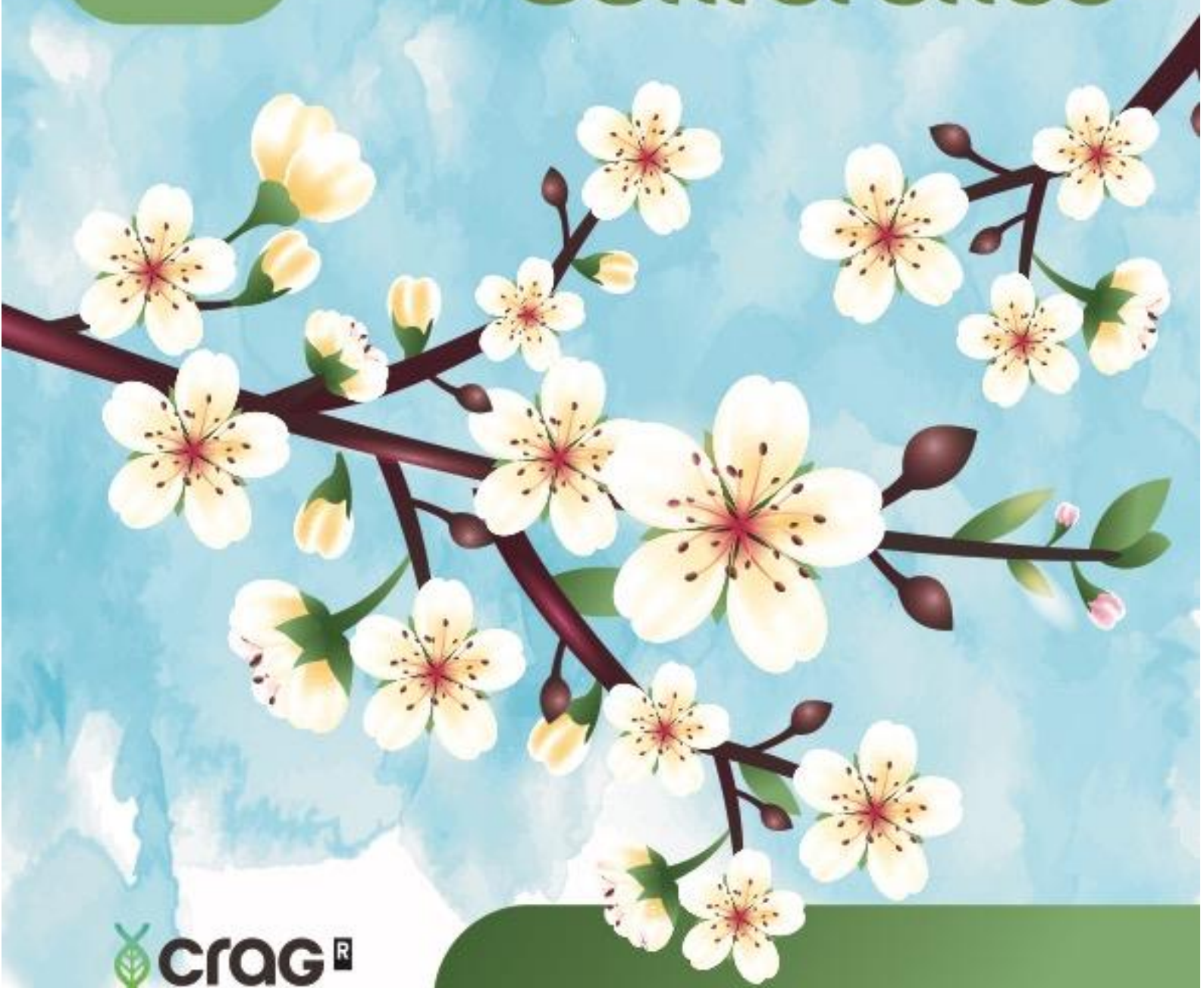


RGC12

# 12<sup>th</sup> Rosaceae Genomics Conference




 **crag**<sup>®</sup>  
CENTRE FOR RESEARCH  
IN AGRICULTURAL GENOMICS

**IRTA**<sup>®</sup>

Institut  
de Recerca i Tecnologia  
Agroalimentàries

6 - 9 May, 2025  
Costa Brava-Barcelona

 **GEYSECO**.es  
Asociación de Genetistas y Evolucionistas

SECRETARÍA TÉCNICA  
TEL +34 932 212 242

[www.rosaceaegenomics.com](http://www.rosaceaegenomics.com)

T4-PS2.26

## AXIOSAFE (AXIOM™ SNP ASSESSMENT AND FILTERING ENGINE): A NOVEL PYTHON PIPELINE FOR AXIOM™ GENOTYPING DATASETS

**Lorenzo Spina**<sup>1</sup>, **Diego Micheletti**<sup>1</sup>, **Giorgio Tumino**<sup>2</sup>, **Stijn Vanderzande**<sup>2</sup>, **Nicole Pradas**<sup>3</sup>, **Maria Jose Aranzana**<sup>3,4</sup>, **Nicholas P. Howard**<sup>5</sup>, **Michela Troggo**<sup>1</sup>, **Eric van de Weg**<sup>2</sup>, **Luca Bianco**<sup>1</sup>

<sup>1</sup>Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige TN, Italy  
University School for Advanced Studies IUSS Pavia, Pavia PV, Italy

<sup>2</sup>Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

<sup>3</sup>Centre for Research in Agricultural Genomics (CRAG), Barcelona, Spain

<sup>4</sup>Institut of Agrifood Research and Technology (IRTA), Genomics and Biotechnology, Barcelona, Spain

<sup>5</sup>Fresh Forward Breeding and Marketing B.V., Huissen, The Netherlands

### Abstract

The information encoded in genomes plays a crucial role in modern plant breeding. Among the technologies used to characterize plants' genetic makeup, SNP genotyping remains one of the most widely used methods. In particular, array-based genotyping has seen the Axiom™ (ThermoFisher) platform become increasingly popular, primarily due to the development of Axiom™ array designs for crops and fruit tree species such as apple, pear, almond, grapevine, walnut, hazelnut, and chestnut. Given the large amount of data that these arrays can produce, a standard analysis pipeline is needed to automatically and reproducibly manage and curate the resulting genotyping data before it can be used for breeding purposes.

This work presents AxioSAFE (Axiom™ SNP Assessment and Filtering Engine), a new data curation pipeline designed explicitly for Axiom genotyping datasets. AxioSAFE is implemented as a modular Python program which performs curation over separate and optionally independent analysis steps. The pipeline accepts the results from the Axiom Analysis Suite base calling software as input and is compatible with huge datasets. The key functionalities available in AxioSAFE are: (1) identification of non-diploid samples, (2) analysis of array signal data clustering patterns associated with problematic SNPs, (3) detection of duplicate samples, (4) pedigree inference for identifying Parent-Offspring and Parent-Parent-Offspring relationships, and (5) phasing of samples utilizing known pedigree information.

AxioSAFE performs data curation by marking for removal all samples and SNPs that fail one or more curation filters, resulting in safe-to-use and highly curated genotype datasets. All removed samples and markers are reported. Additionally, users can provide an external genotype reference dataset which will be integrated with the tested dataset before executing pedigree inference. AxioSAFE provides a fast, automated alternative to the manual curation of genotype data, delivering reliable results for downstream analyses in Breeding-By-Design (BSD) projects.