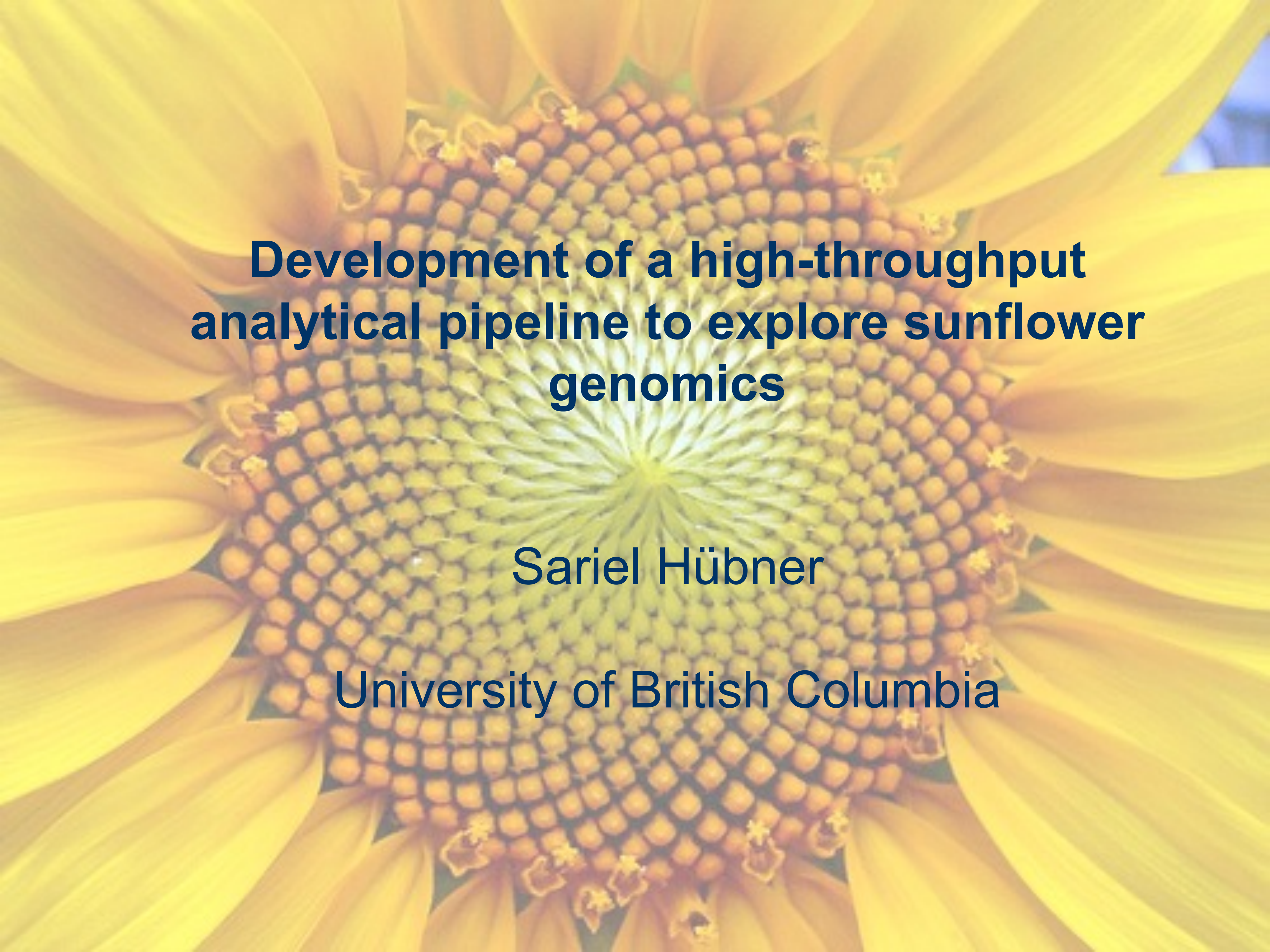


Development of a high-throughput analytical pipeline to explore sunflower genomics

Sariel Hübner

University of British Columbia

Exploring sunflower genomics

Diversity

Association

Variation

LD

Ancestry

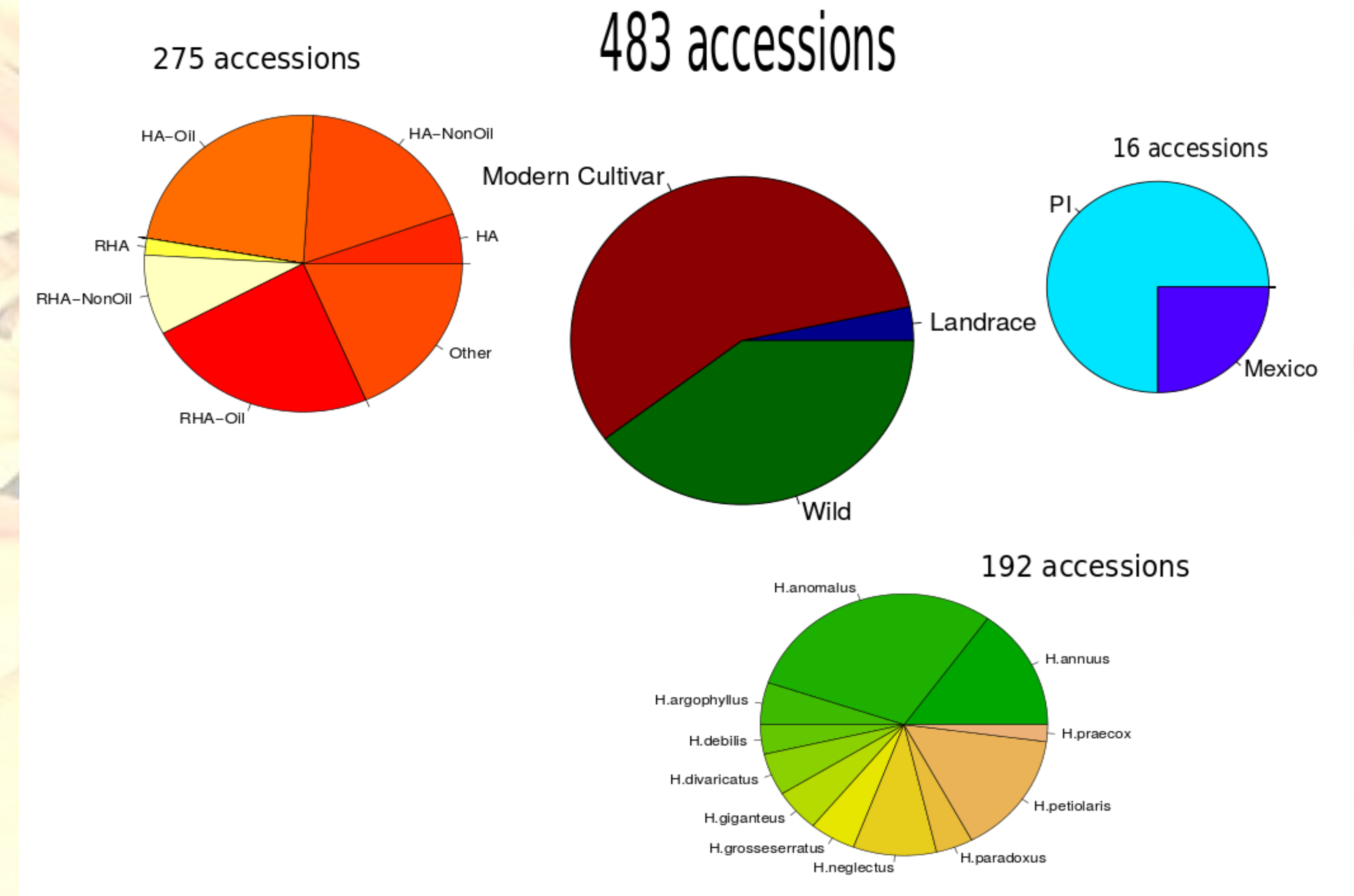
Selection

Introgression

Differentiation

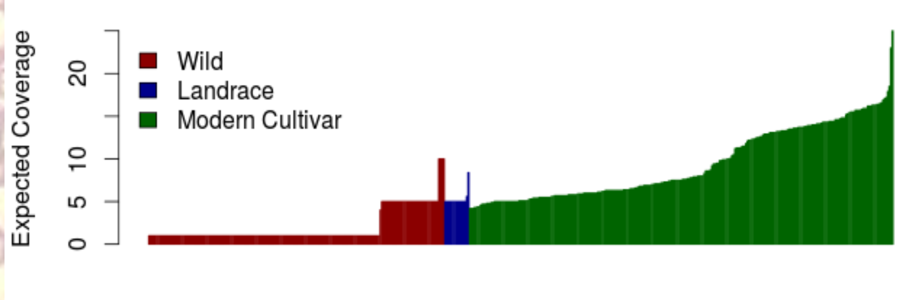
Reference Genome

Resequencing sunflower accessions

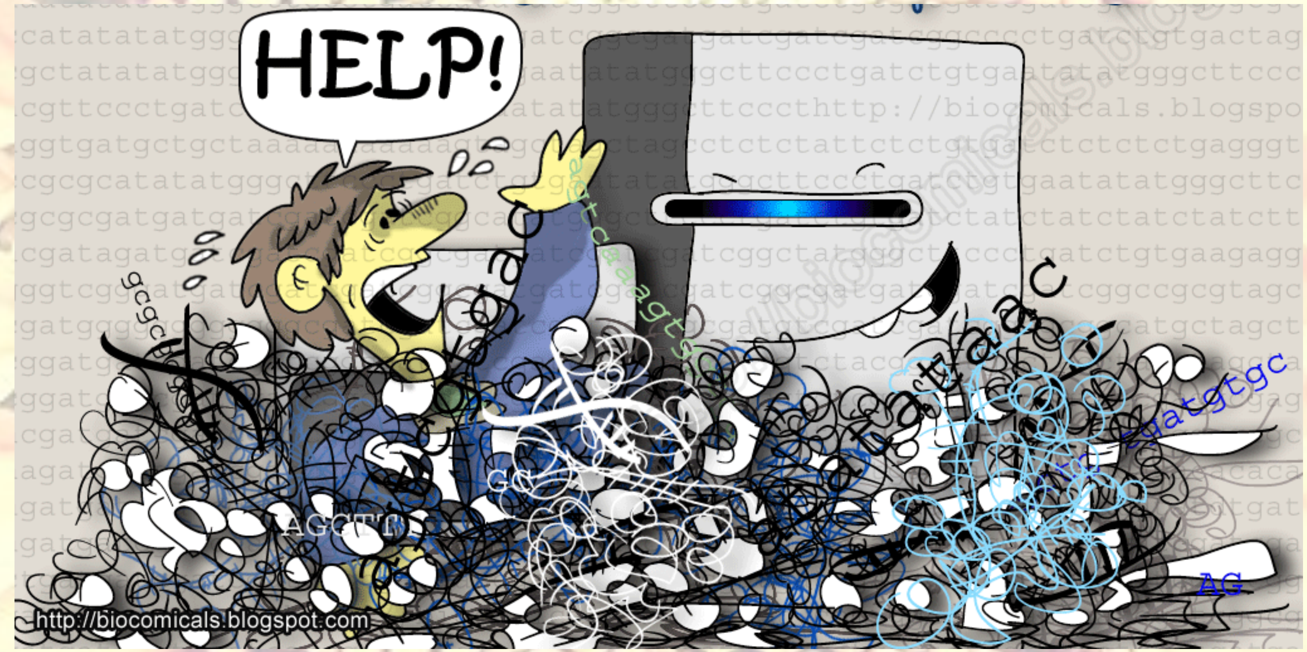


Sequence coverage across accessions

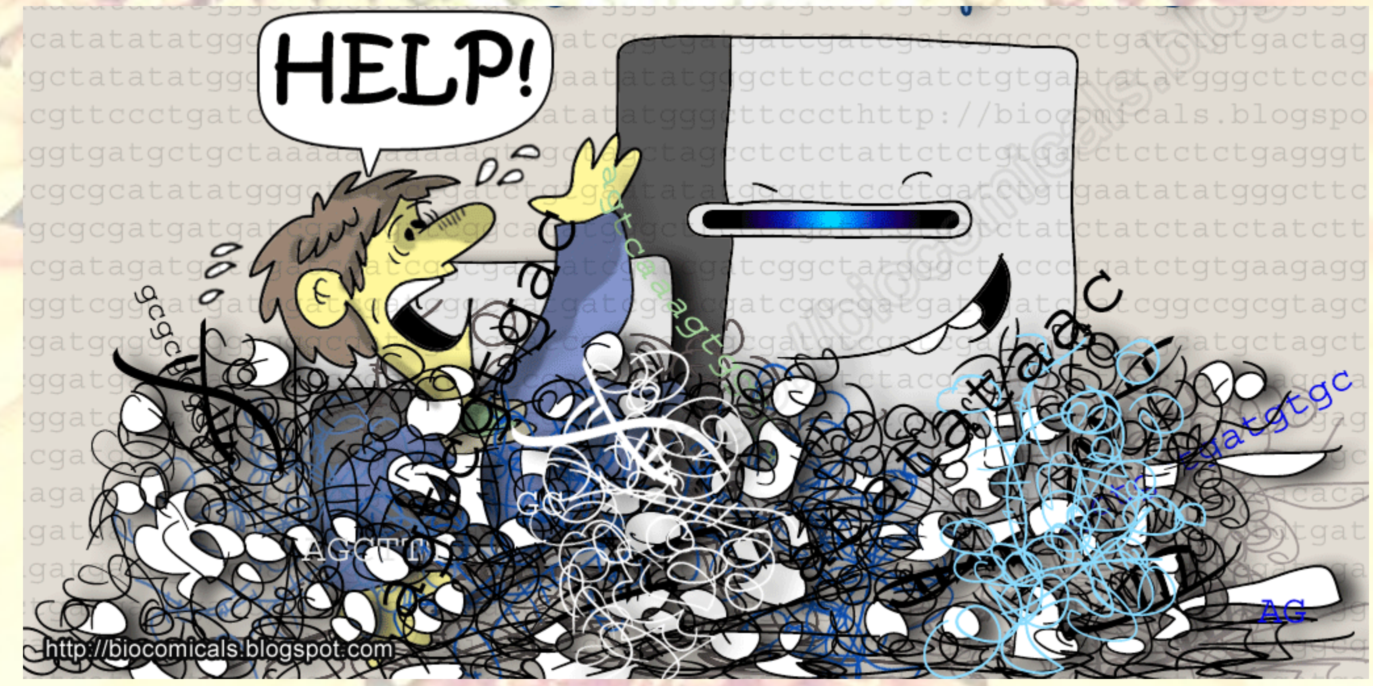
Σ



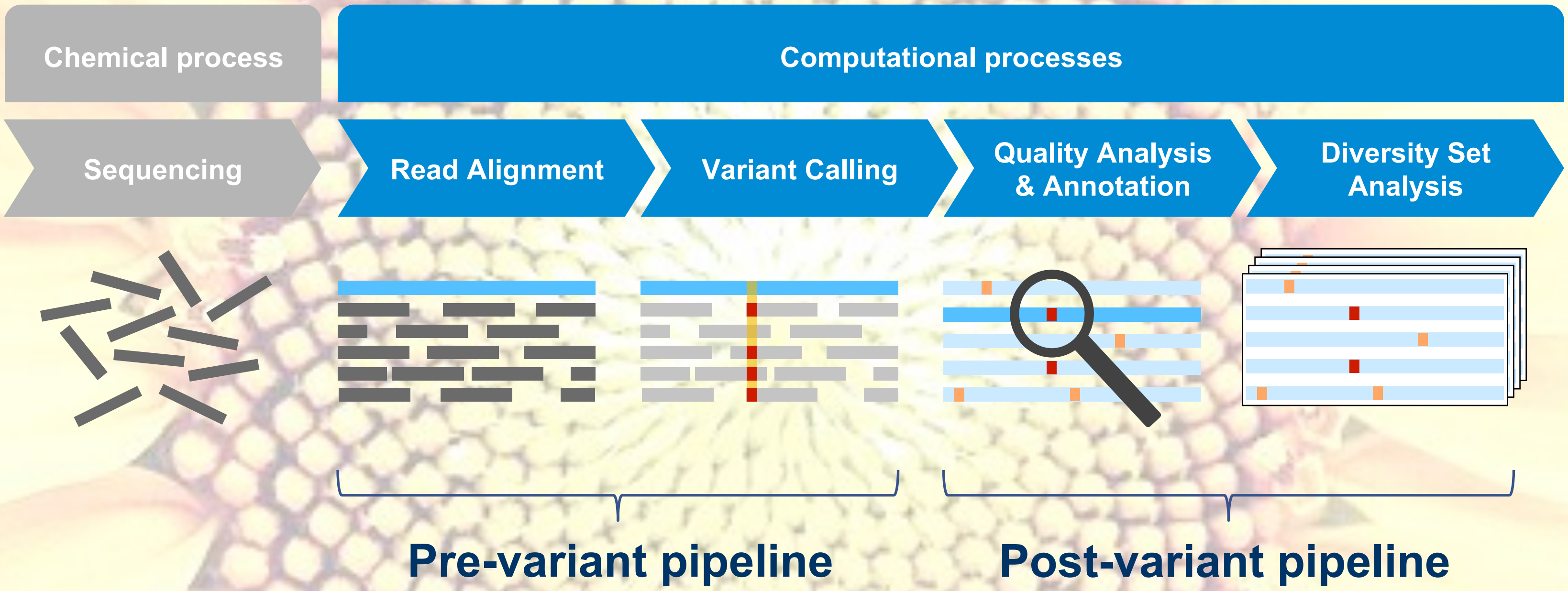
≈ 10Tb


HELP!
http://biocomicals.blogspot.com


SAP HANA® Solution
Redefines In-memory Computing


HELP!
http://biocomicals.blogspot.com

Overview genomics analysis process



Chemical process

Computational processes

Sequencing

Read Alignment

Variant Calling

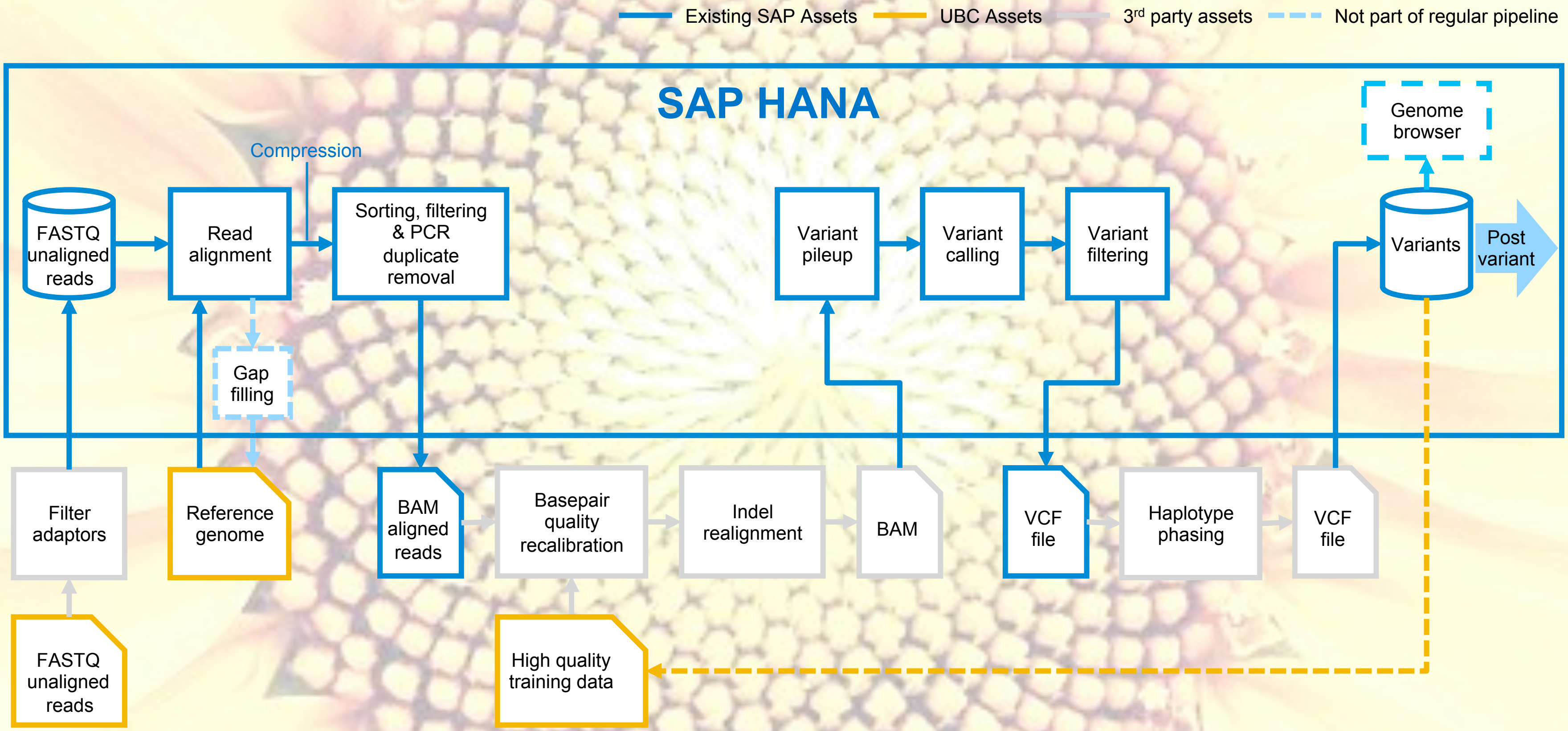
Quality Analysis & Annotation

Diversity Set Analysis

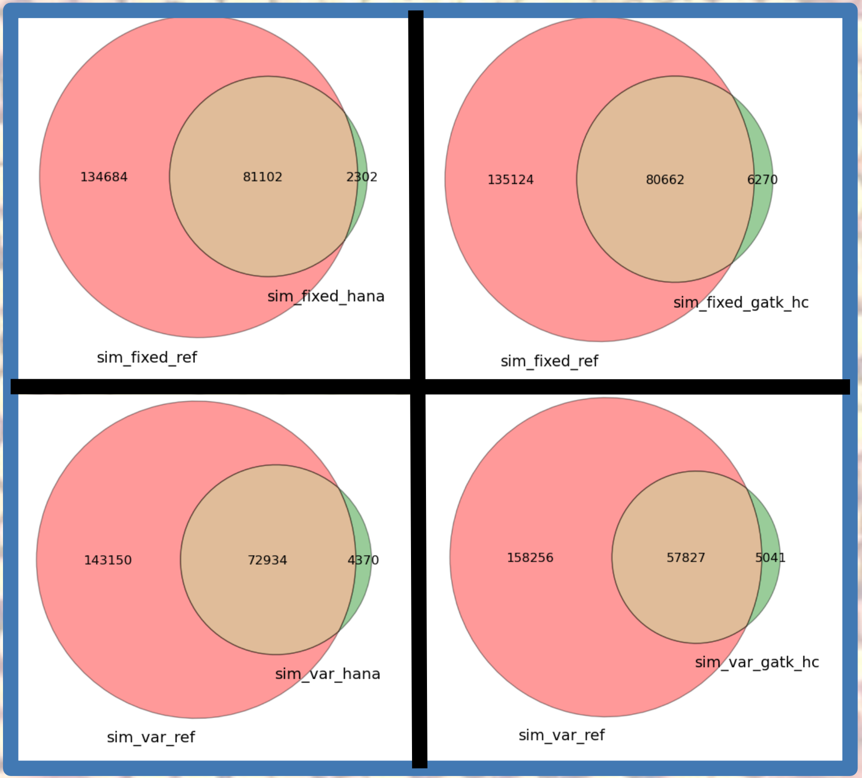
Pre-variant pipeline

Post-variant pipeline

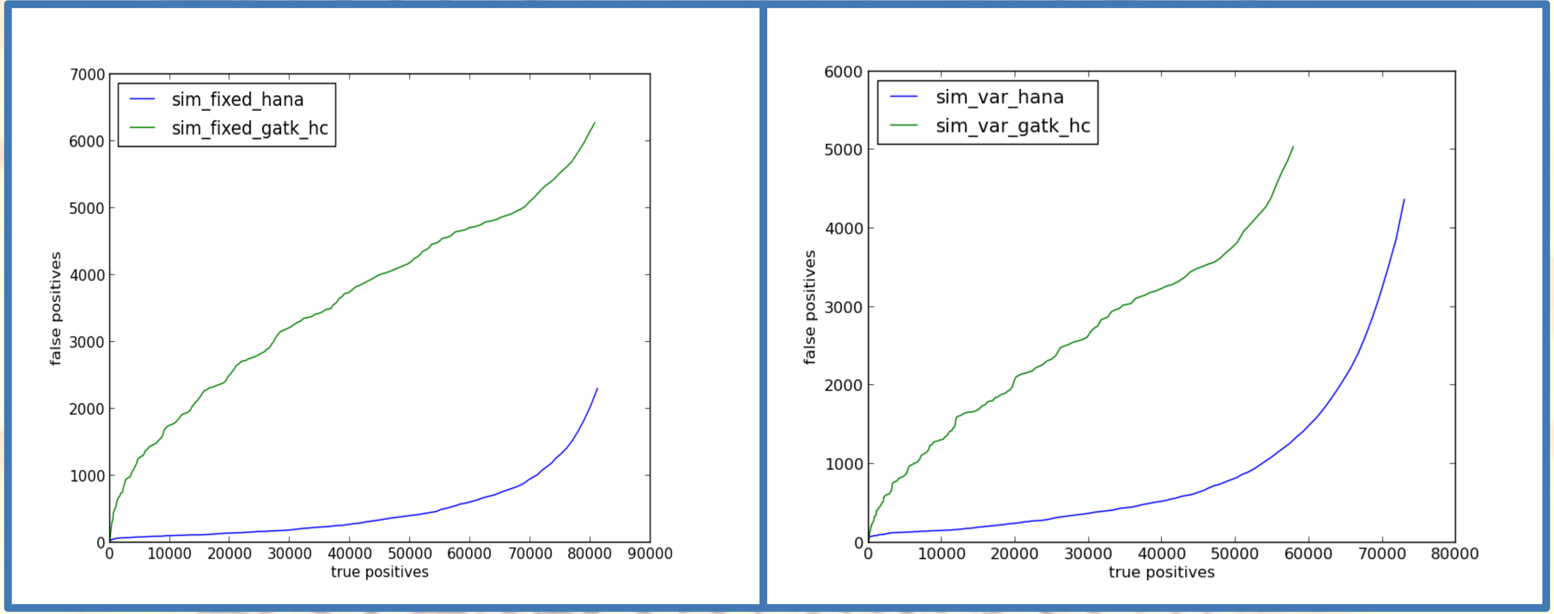
Pre-variant pipeline



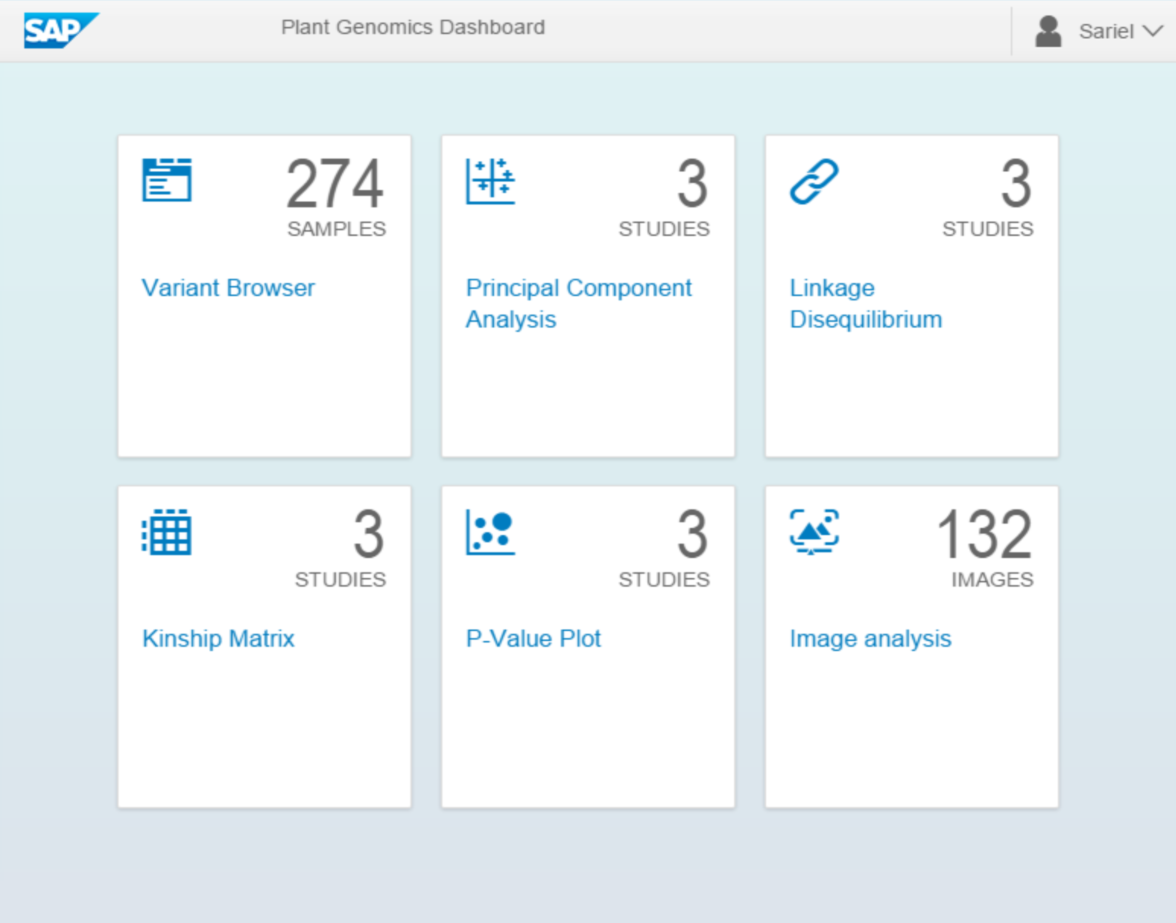
Benchmarking: SAP vs. GATK



Benchmarking: SAP vs. GATK



Plant genomics application



Plant Genomics Dashboard

Sariel

274 SAMPLES
Variant Browser

3 STUDIES
Principal Component Analysis

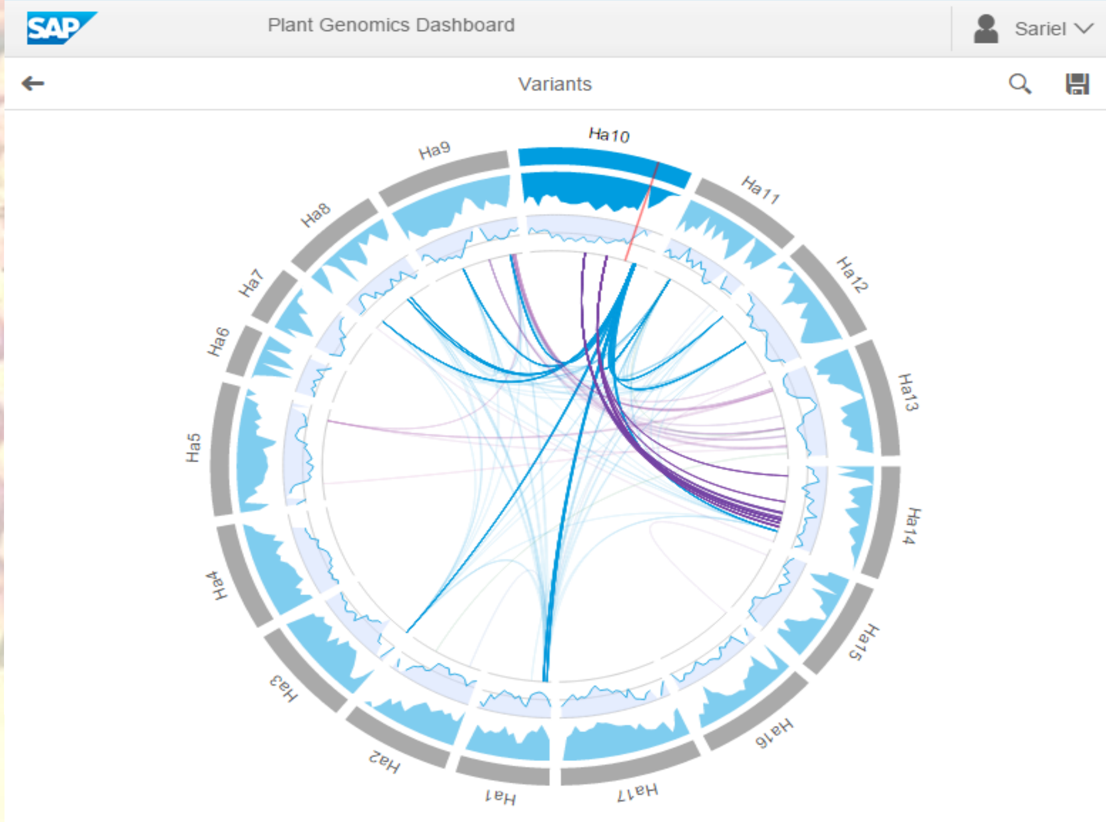
3 STUDIES
Linkage Disequilibrium

3 STUDIES
Kinship Matrix

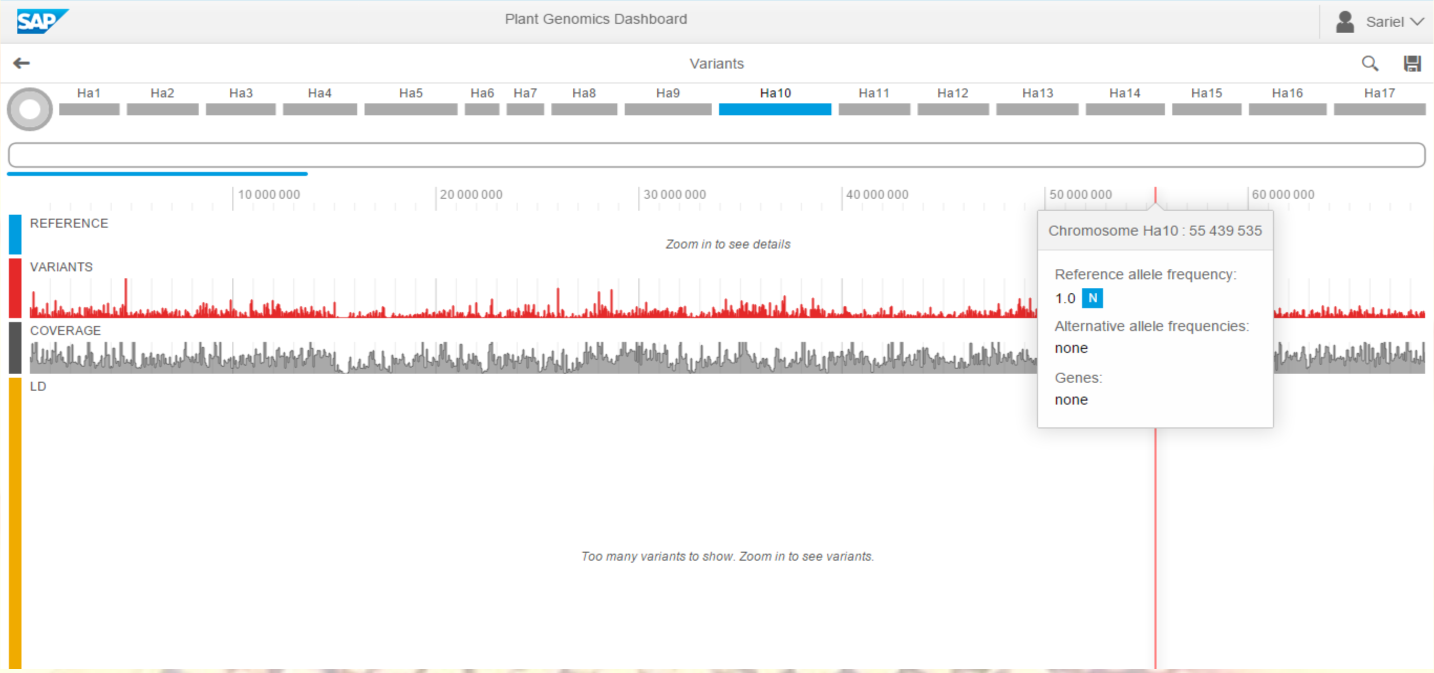
3 STUDIES
P-Value Plot

132 IMAGES
Image analysis

High quality SNPs: 72,305,390
After filtering (MAF = 10%; NA<30%): 4,129,575



Chromosome Ha10



Zoom in to see details

Chromosome Ha10 : 55 439 535

Reference allele frequency:
1.0 N

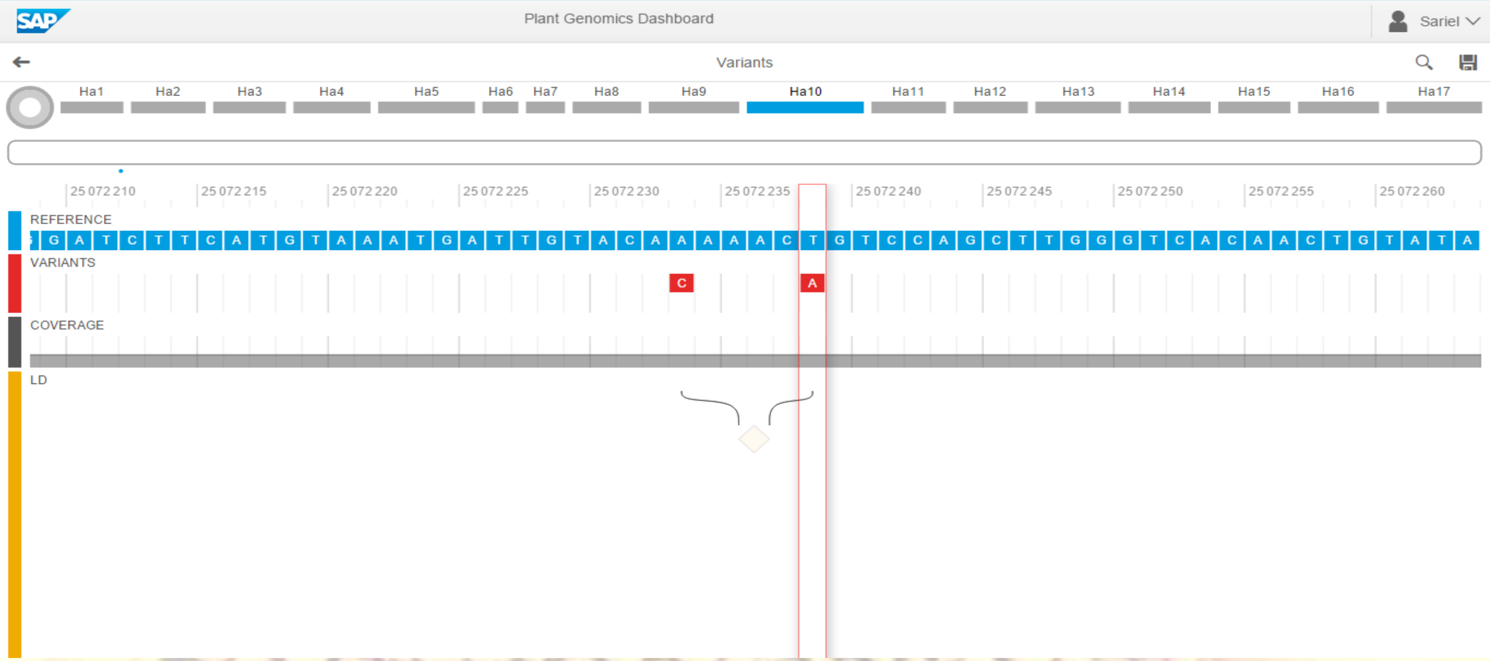
Alternative allele frequencies:
none

Genes:
none

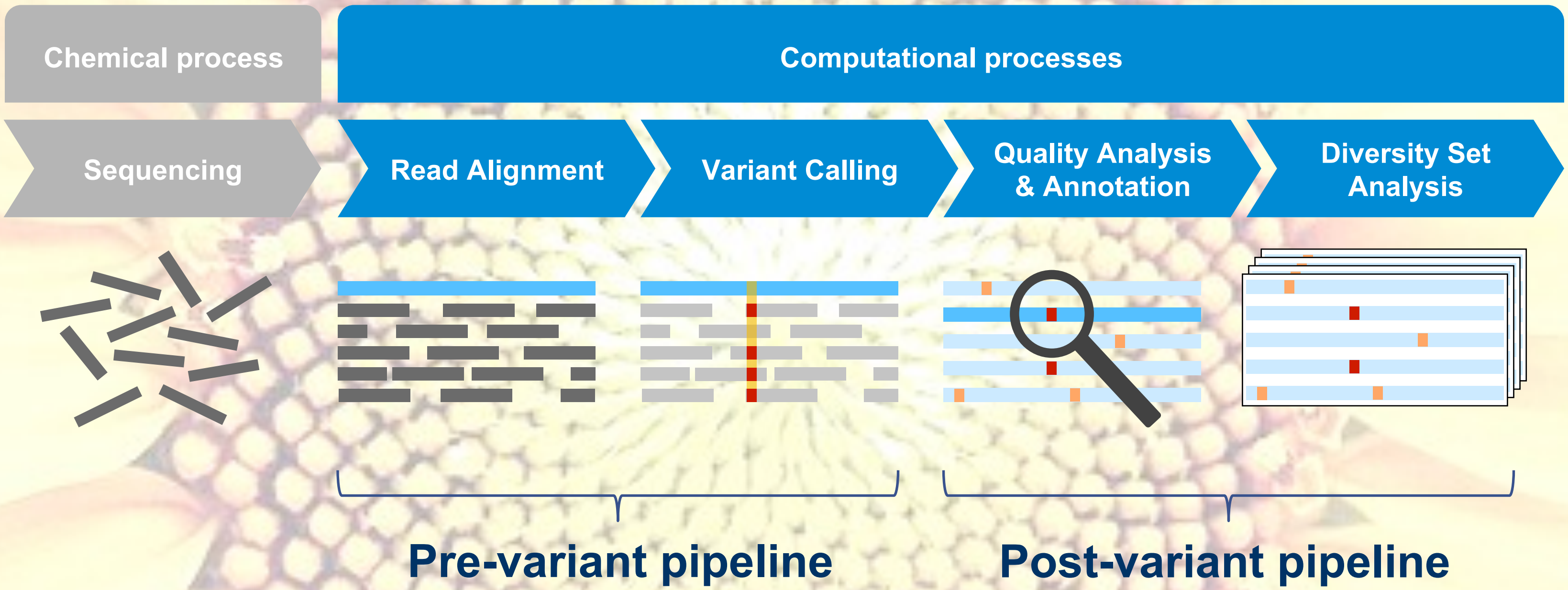
Too many variants to show. Zoom in to see variants.

Chromosome Ha10

Chromosome Ha10



Overview genomics analysis process



Chemical process

Computational processes

Sequencing

Read Alignment

Variant Calling

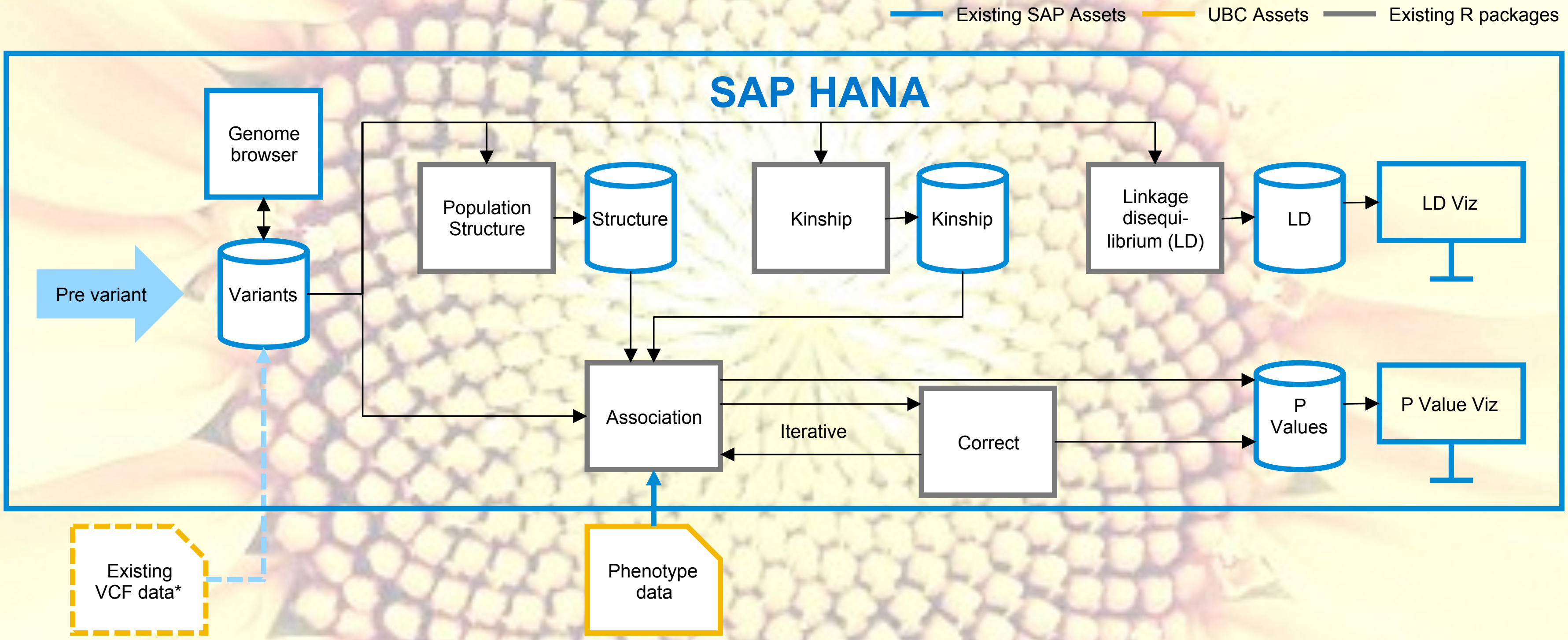
Quality Analysis & Annotation

Diversity Set Analysis

Pre-variant pipeline

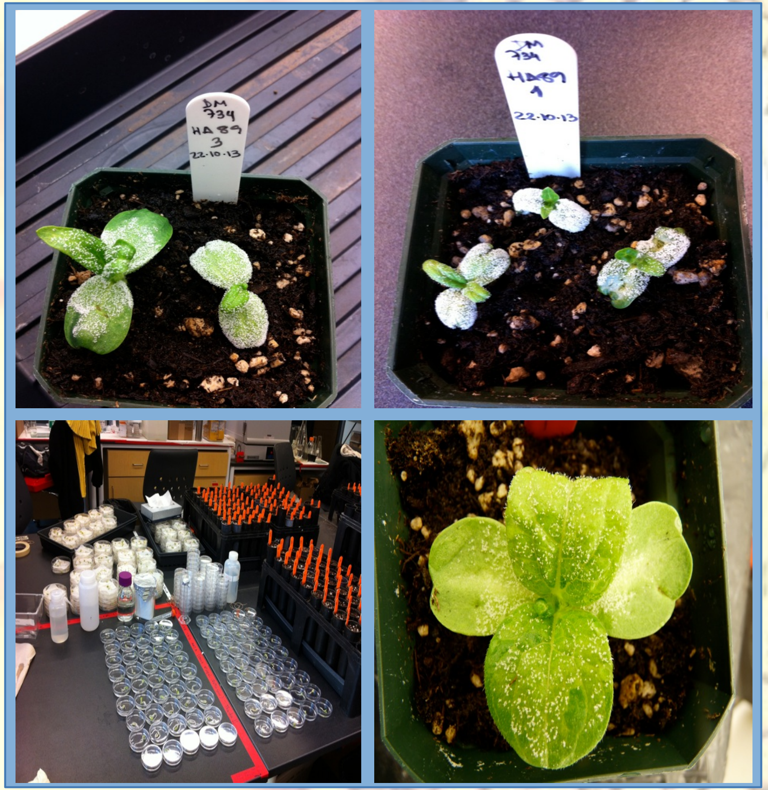
Post-variant pipeline

Post-variant pipeline



Downey mildew resistance in sunflower

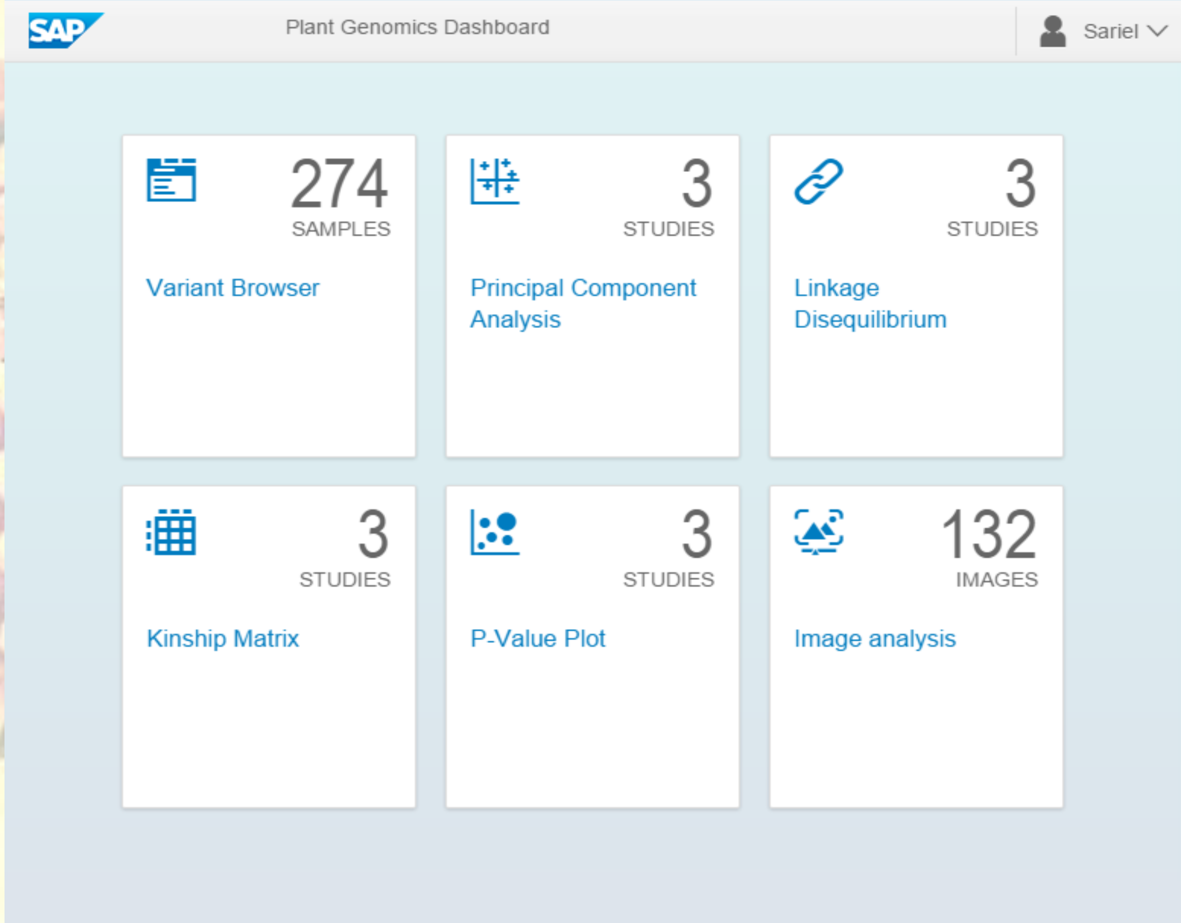
Natalia Bercovich



- Presence of spores in cotyledons
- Presence of spores in leaves
- Spores density in cotyledons

$$(\prod_{i=1}^{n} t_i)^{\frac{1}{n}} = \sqrt[3]{pC \times dC \times pL}$$

Plant genomics application



274 SAMPLES
Variant Browser

3 STUDIES
Principal Component Analysis

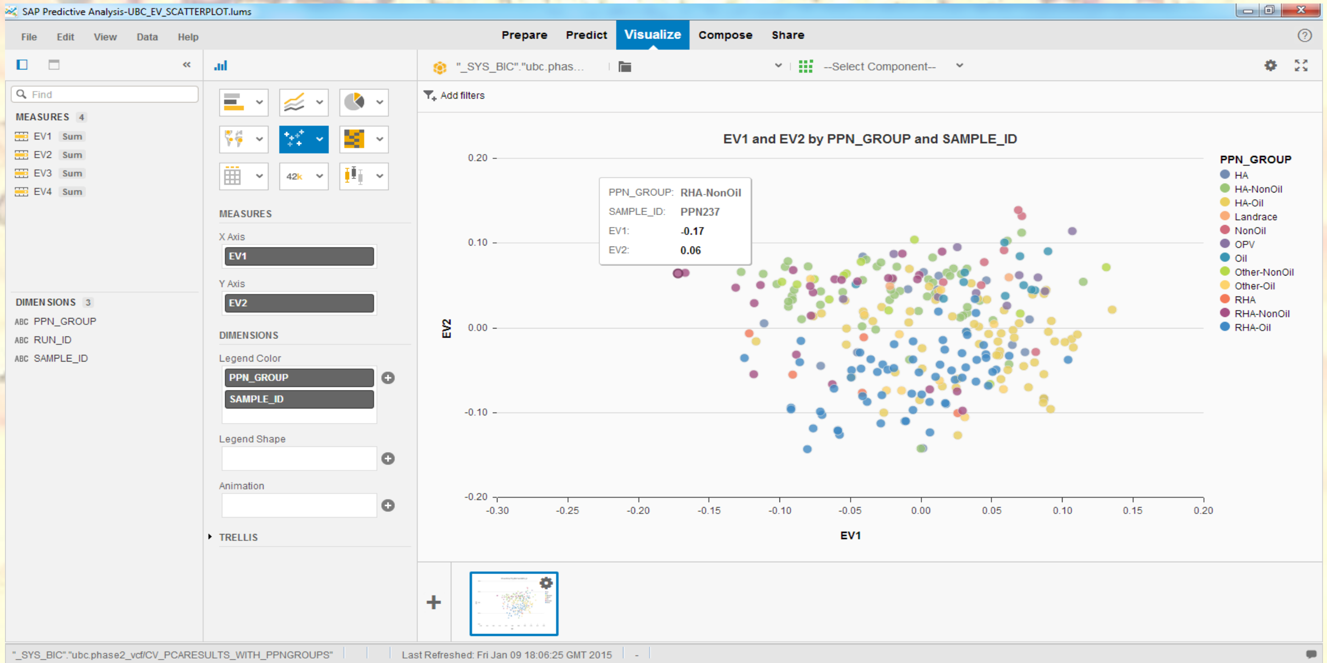
3 STUDIES
Linkage Disequilibrium

3 STUDIES
Kinship Matrix

3 STUDIES
P-Value Plot

132 IMAGES
Image analysis

Population structure based on 1.5M SNPs



X Axis: EV1

Y Axis: EV2

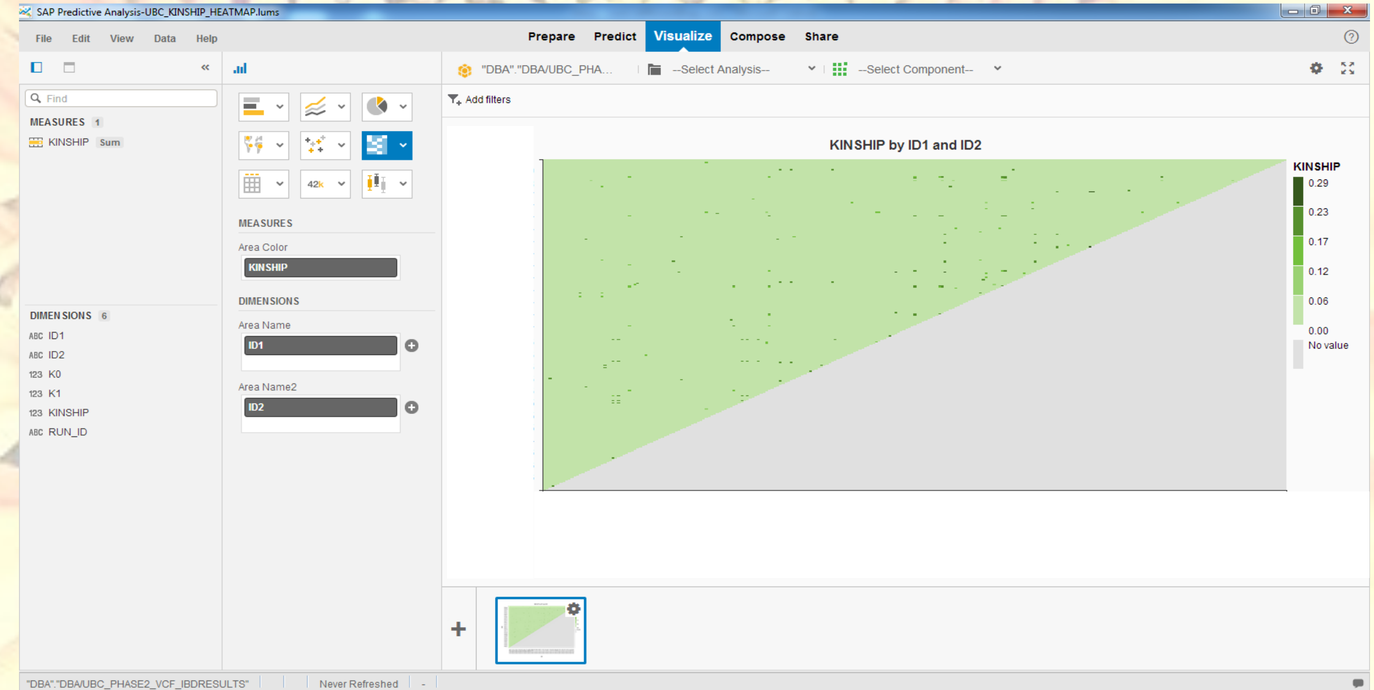
Legend Color: PPN_GROUP, SAMPLE_ID

PPN_GROUP: RHA-NonOil
SAMPLE_ID: PPN237
EV1: -0.17
EV2: 0.06

"_SYS_BIC"."ubc.phase2_vcf/CV_PCARESULTS_WITH_PPNGROUPS"

Last Refreshed: Fri Jan 09 18:06:25 GMT 2015

Kinship matrix based on 1.5M SNPs



MEASURES

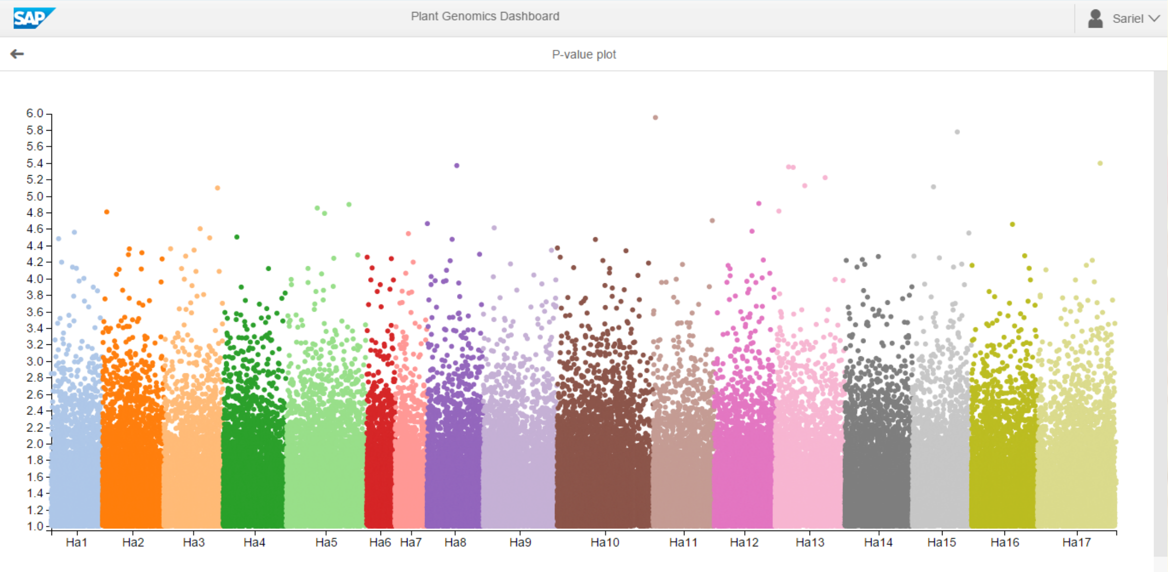
Area Color: KINSHIP

DIMENSIONS

Area Name: ID1

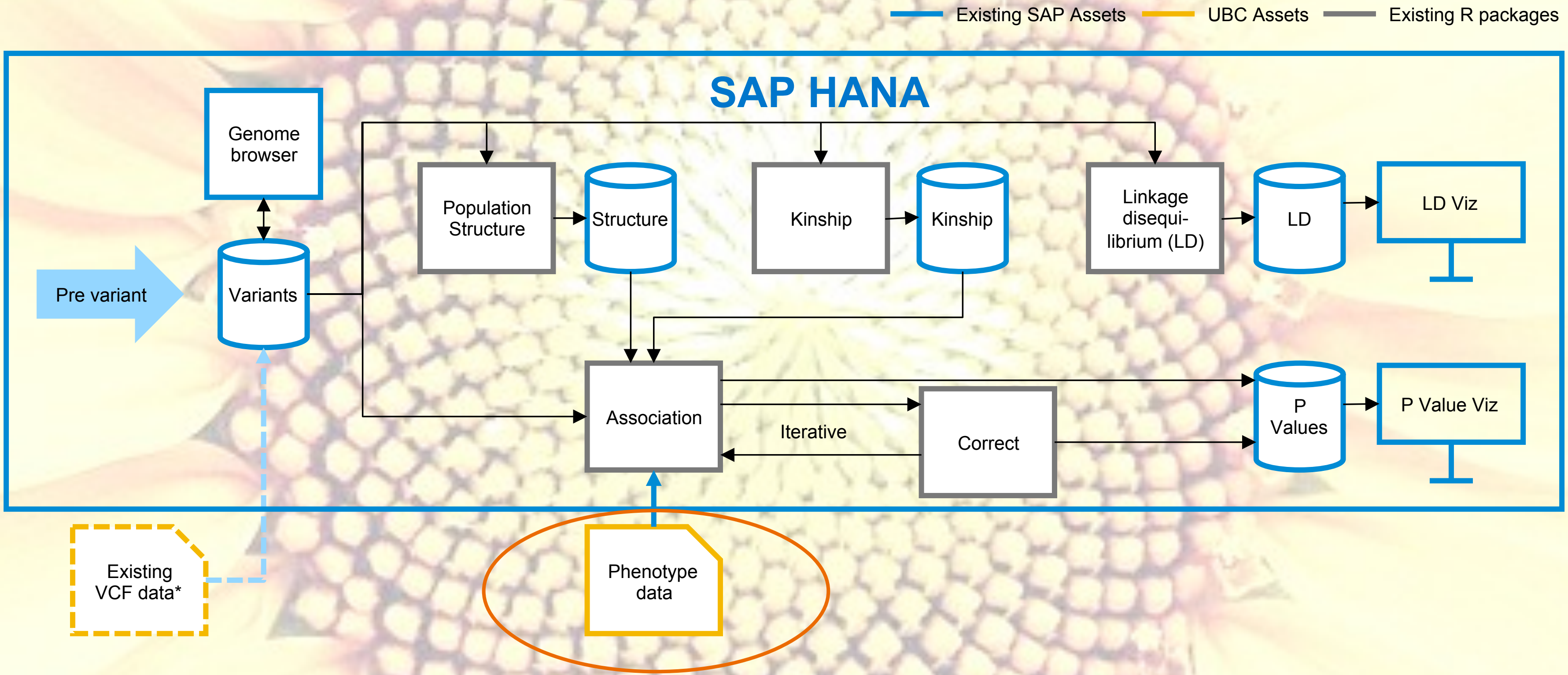
Area Name2: ID2

Association mapping of Downey mildew resistance in sunflower



Filtering at any stringency using Bonferroni, BH-FDR, SimpleM

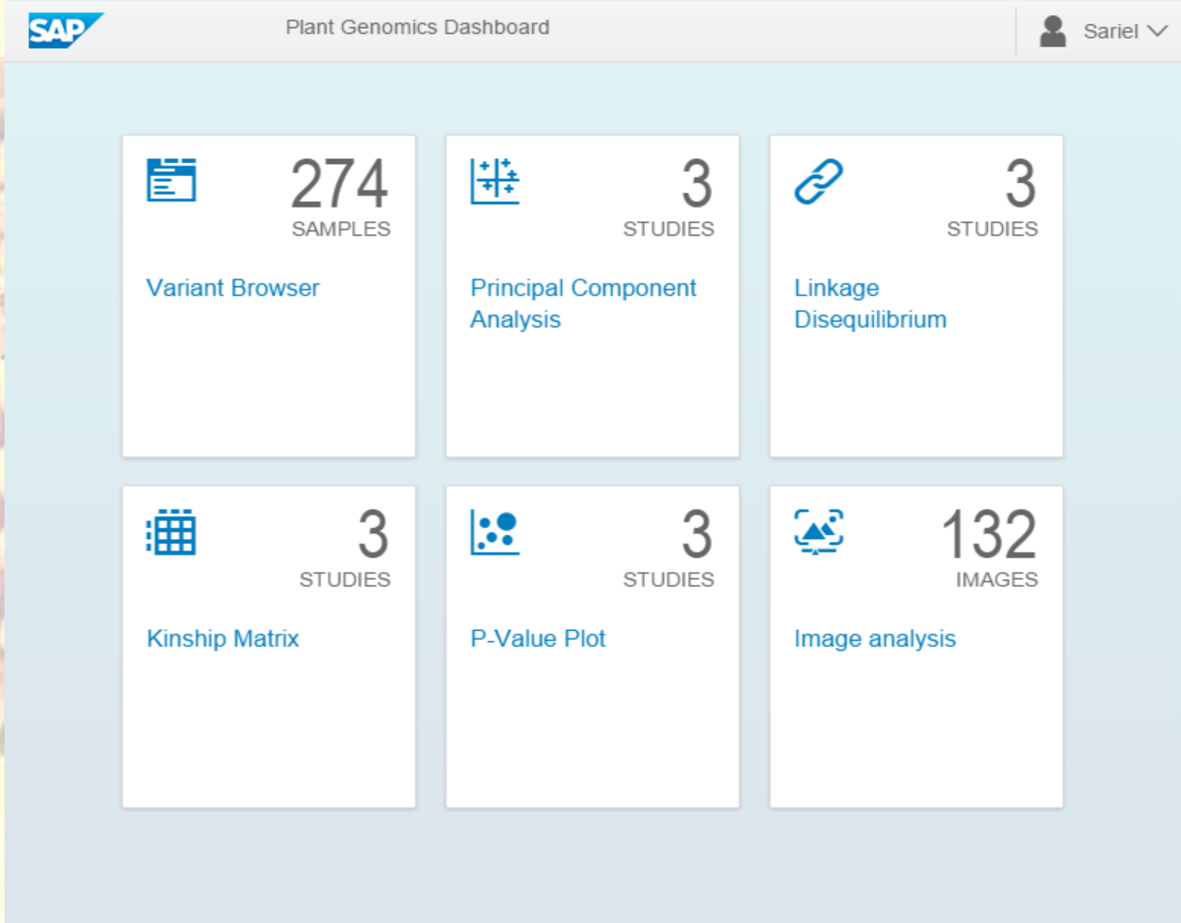
Post-variant pipeline



Jerusalem artichoke variation

Dan Bock and Michael Kantar

Plant genomics application



Plant Genomics Dashboard

274 SAMPLES Variant Browser	3 STUDIES Principal Component Analysis	3 STUDIES Linkage Disequilibrium
3 STUDIES Kinship Matrix	3 STUDIES P-Value Plot	132 IMAGES Image analysis

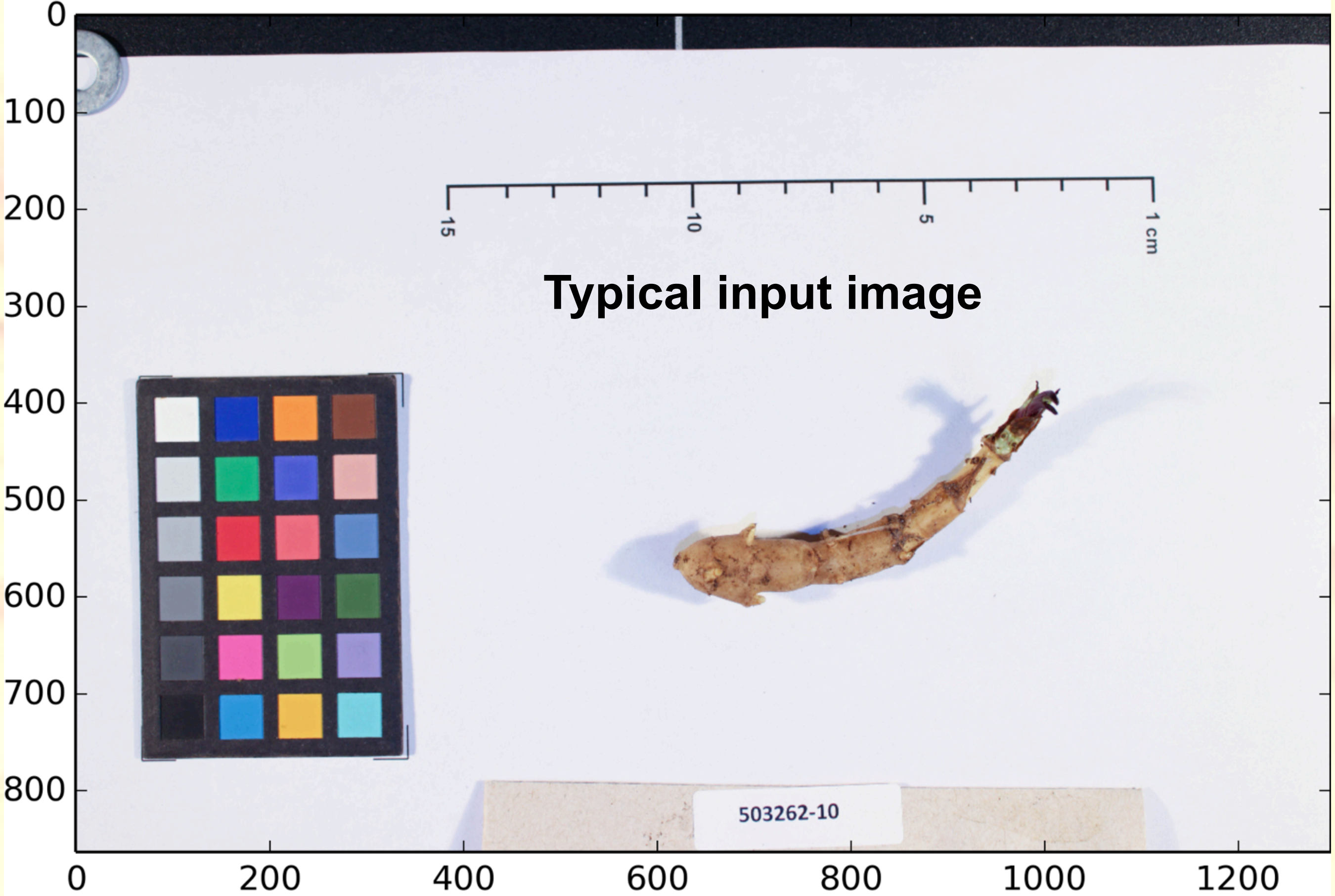
Disclaimer

The information in this document is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This document is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information on this document is not a commitment, promise or legal obligation to deliver any material, code or functionality. This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This document is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP intentionally or grossly negligent.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Lab Preview

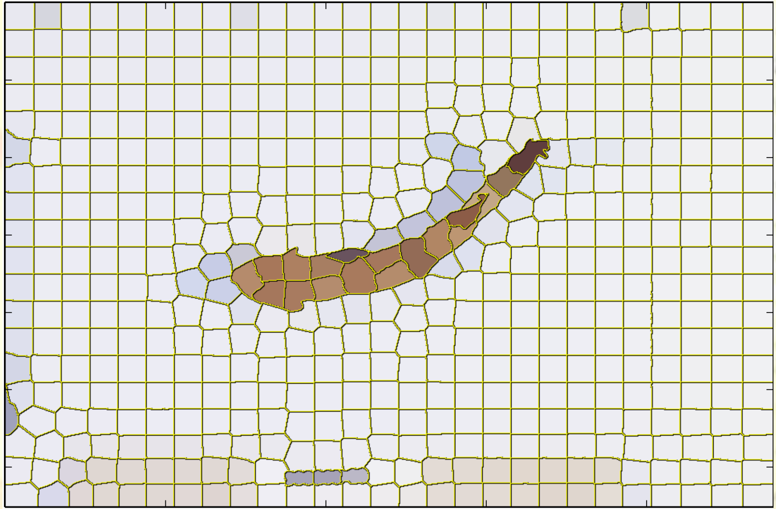
Phenotype analysis for plant genomics



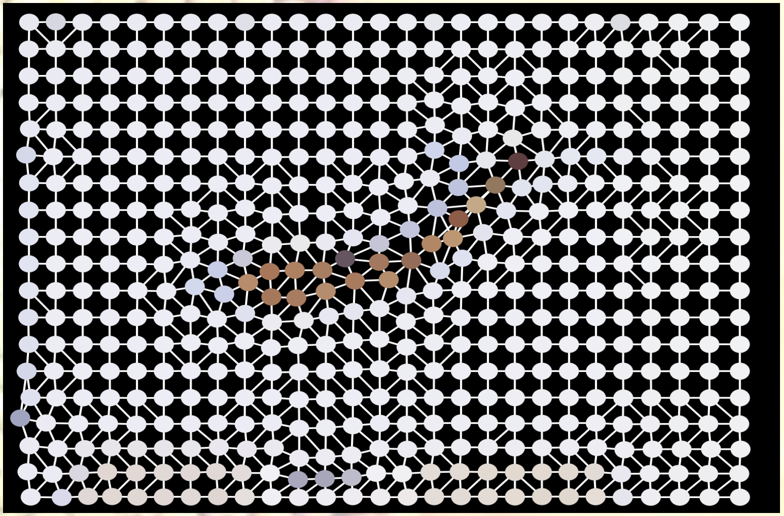
Typical input image

Lab Preview

Image processing algorithms



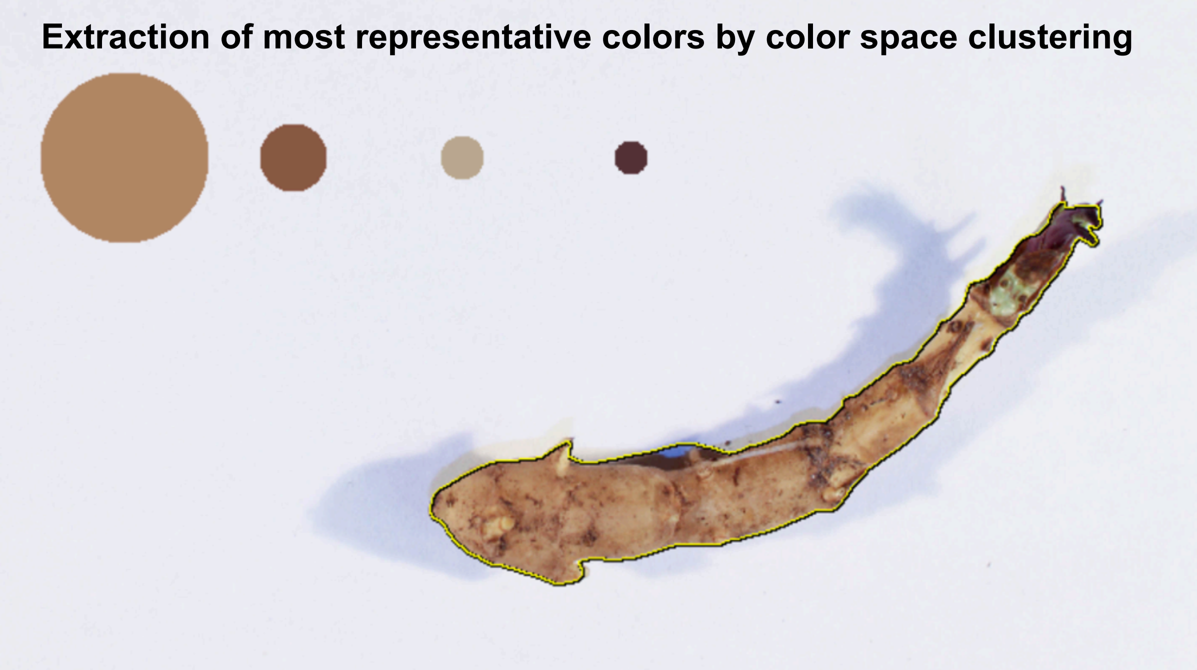
Superpixel pre-segmentation for data reduction



Graph-based representation of superpixels for segmentation by graph clustering

Lab Preview

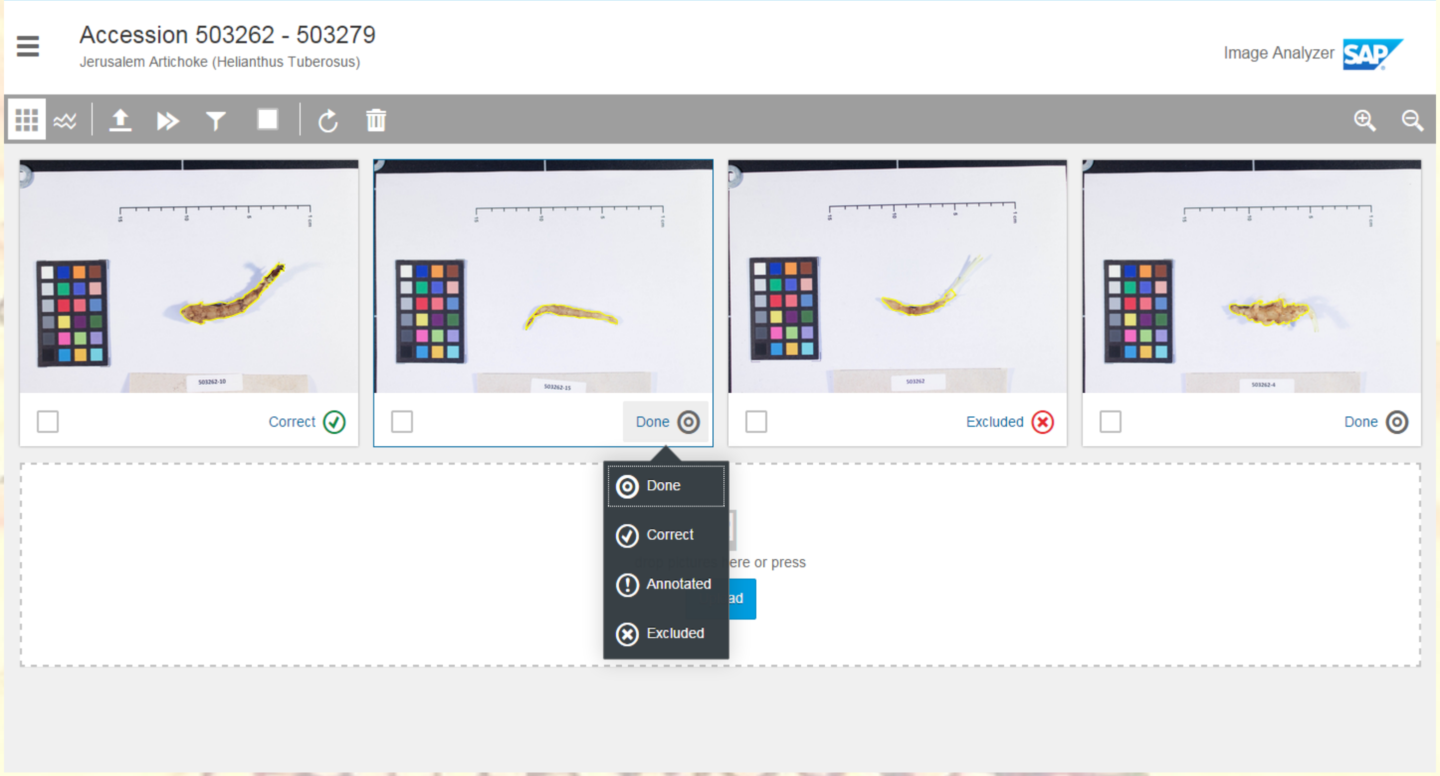
Image processing algorithms



Extraction of most representative colors by color space clustering

Lab Preview

Image processing algorithms

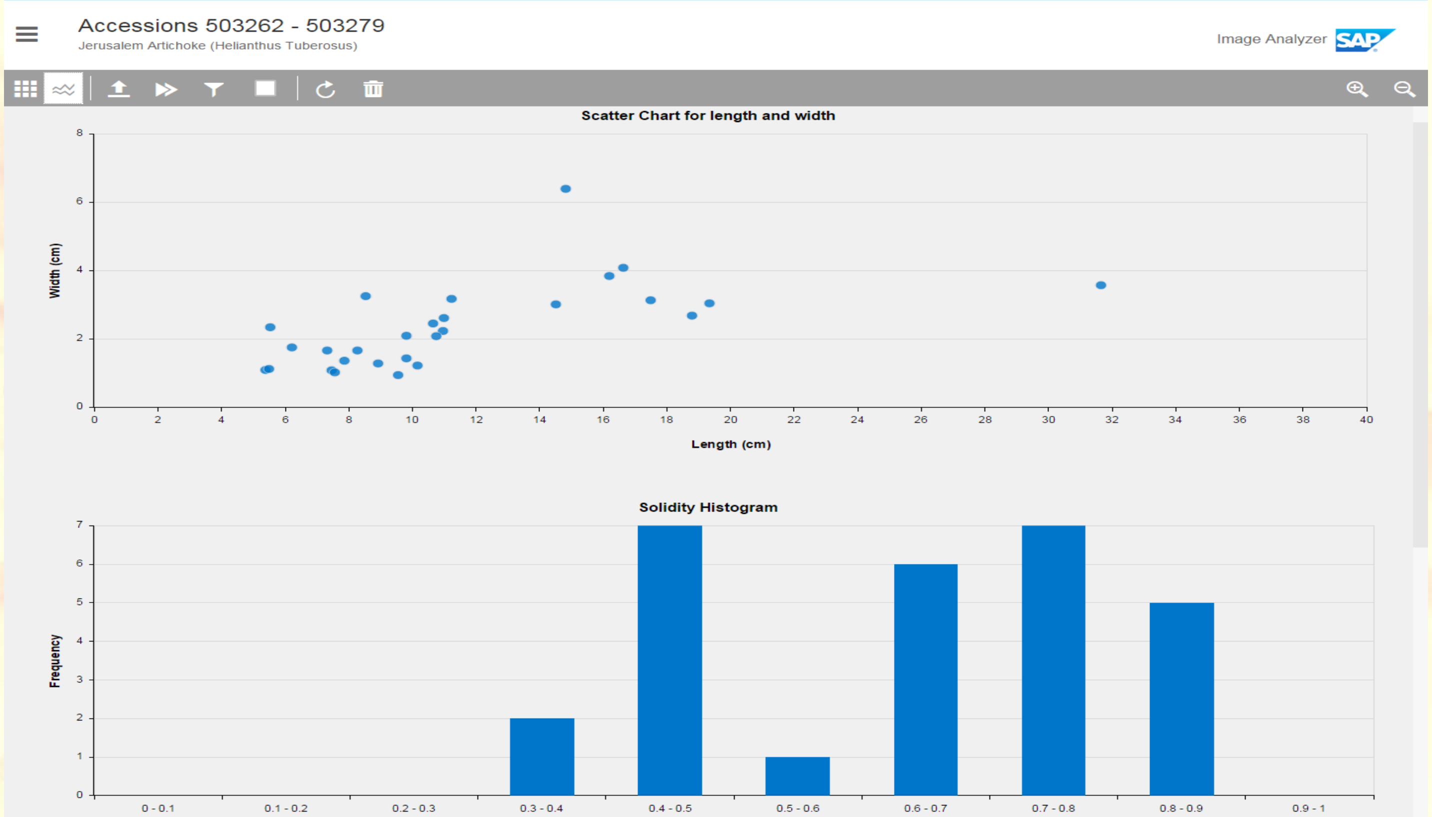


Accession 503262 - 503279

Jerusalem Artichoke (Helianthus Tuberosus)

Lab Preview

Image processing algorithms



Accessions 503262 - 503279

Jerusalem Artichoke (Helianthus Tuberosus)

Lab Preview

What's next?

- Finish alignments (2-3 weeks)
- Process alignments (2-3 weeks)
- Variant calling (10-14 days)
- Phasing, imputation and correction
- Association mapping (meta-analysis)
- Structural and copy number variation
- Diversity, differentiation, selection

Thanks!

Emanuel Ziegler

Mark Heffernan

Christian Bock

Timo Kubach

Markus Noga

Thomas Holst

David Blumenthal-Barby

Loren Rieseberg

Greg Baute

Natalia Bercovich

Dan Bock

Michael Kantar

Marco Todesco

Jenifer Mandel