Quantification and identification of allele specific proteins for polyploid non-model crops:
Proof of principle for 3 banana genotypes/phenotypes.
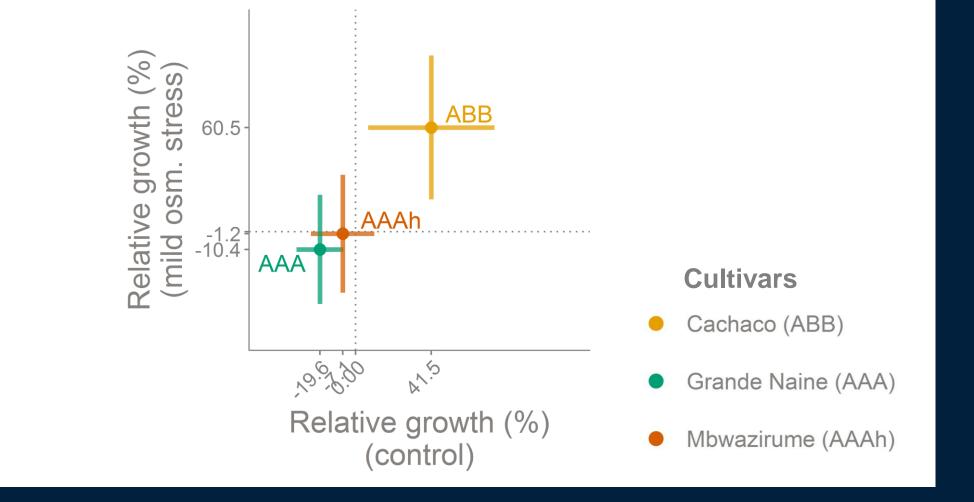
Jelle van Wesemael[1], Yann Hueber[2], Ewaut Kissel[1], Nadia Campos[1], Rony Swennen[1,3,4], Sebastien Carpentier[1,5]

KU LEUVEN

Bioversity International

Banana (Musa spp.) is the most important fruit crop in the world.
Cultivars are triploid, hybrid crosses of *M. acuminata* (A) and *M. balbisiana* (B)
Vulnerable to drought

Polyploidy results in a plethora of genomic, transcriptomic, and proteomic products controlling the phenotype.
Allelic variants are of interest for climate smart agriculture : flexibility towards the environment

Our goal: Mining the Musa biodiversity for drought tolerance:
Select allele specific proteins linked to the observed phenotypic differences among three contrasting genotypes by integration of transcriptomics and proteomics.

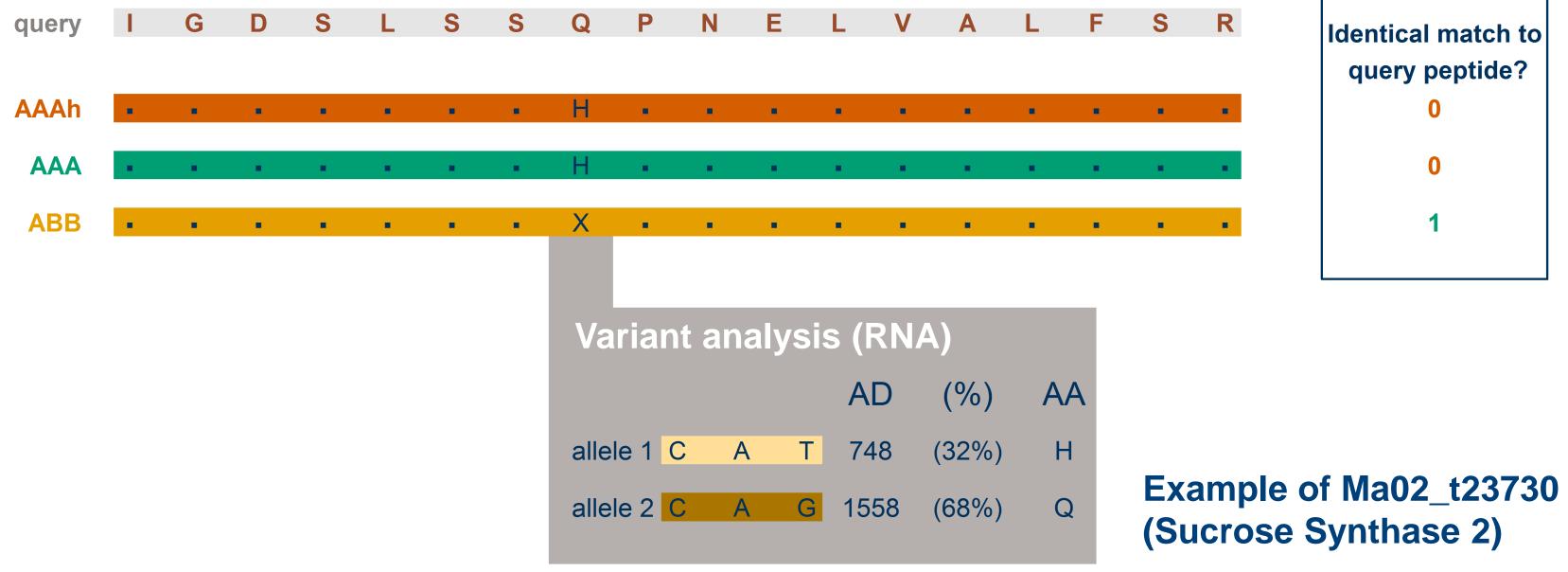## The ABB genotype shows more root growth under control and mild osmotic stress conditions.

Cultivars
- Cachaco (ABB)
- Grande Naine (AAA)
- Mbwazirume (AAAh)

---

260 amino acid polymorphisms (SAAP) are identified by alignment of the peptide sequences to every cultivar specific database

Alignment of peptide query to amino acid (RNA based) database (BLAST)

query	I	G	D	S	L	S	S	Q	P	N	E	L	V	A	L	F	S	R

AAAh ... H ...
AAA ... H ...
ABB ... X ...

Identical match to query peptide?
AAAh: 0
AAA: 0
ABB: 1

Variant analysis (RNA)

				AD	(%)	AA
allele 1	C	A	T	748	(32%)	H
allele 2	C	A	G	1558	(68%)	Q

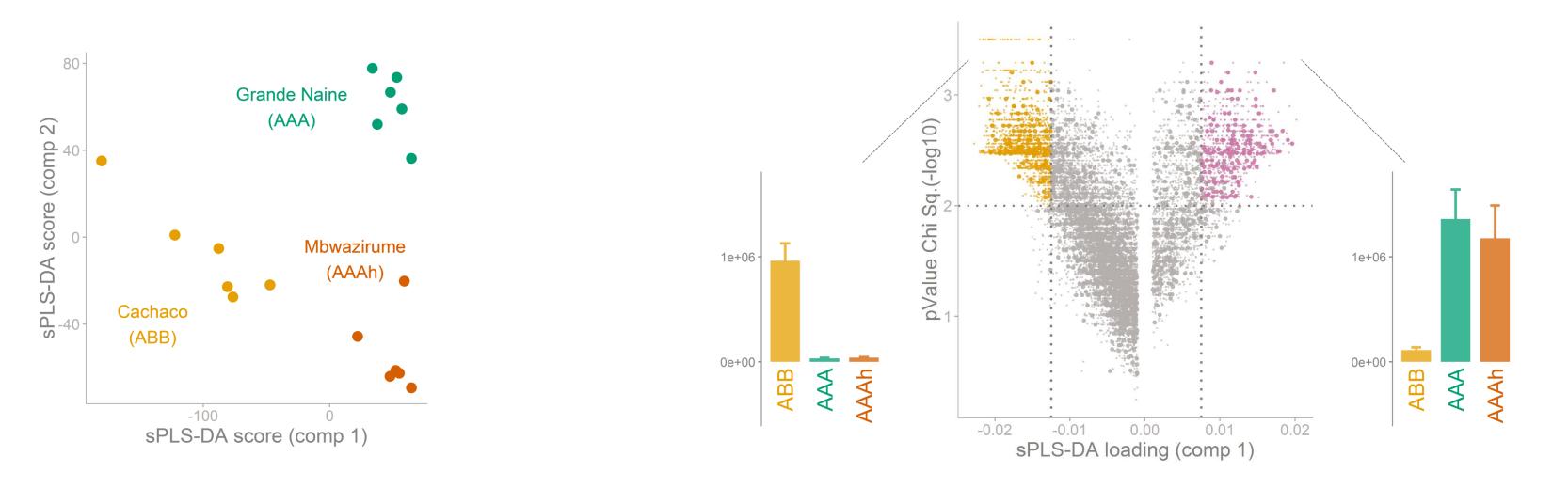**Example of Ma02_t23730 (Sucrose Synthase 2)**

Some polymorphisms in the drought tolerant phenotype (ABB) display homeolog expression bias

The variant specific read count (mRNA) was calculated for the ABB cultivar, a triploid hybrid cross:
1 *Musa acuminata* (A)
2 *Musa balbisiana* (B) genomes.

Thus we expect:
33% of reads A-isoform
66% of reads B-isoform

Read ratio B / (A+B) — Number of B specific alleles
Read ratio A / (A+B) — Number of A specific alleles

27 genes have 100 % B reads (expression level dominance)

---

130 additional allele specific genes are selected through combination of univariate and multivariate statistics based on the peptide abundance.

Multivariate: Sparse Partial Least Squares (sPLS-DA, mixomics, R)
Univariate: Kruskal-Wallis (p<0.05, BH corrected)

Cultivar specific genes are involved in general osmotic stress responses.
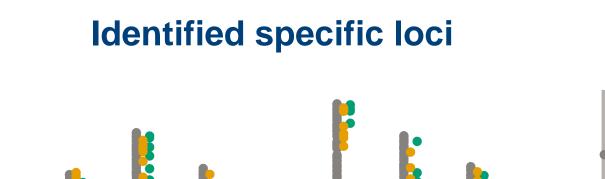
Phenotype of preference (ABB) invests in root growth with minimal setback under osmotic stress

The functions of identified alleles are enriched in general osmotic stress responses

GO-term	Description	p (Fisher)
GO:0006096	Glycolytic process	0.003
GO:0006457	Protein folding	0.015
GO:0006950	Response to stress	0.018

Mining the Musa genome by integration of transcriptomics with proteomics identifies 390 genes with allele specificity linked to the differential phenotype

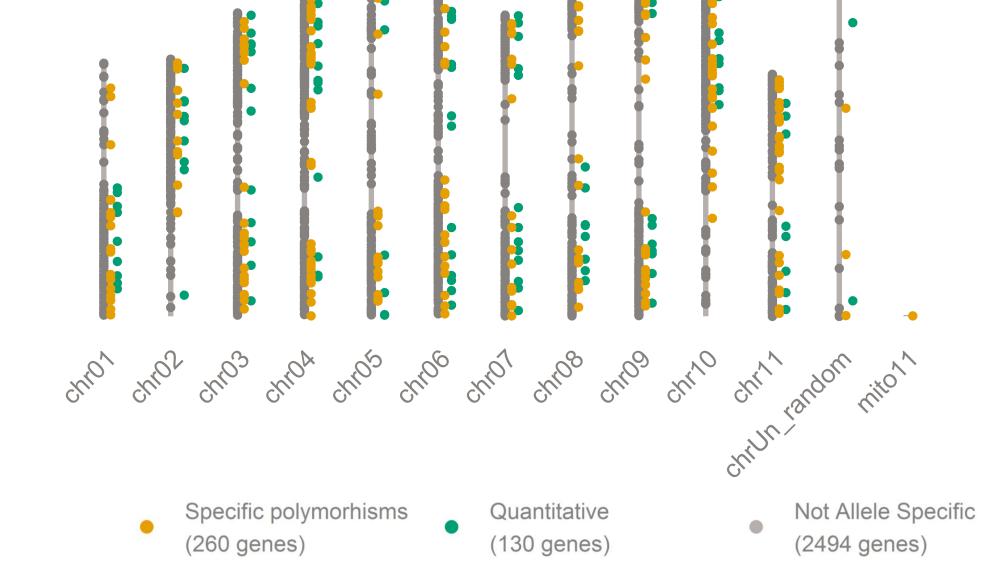The phenotype is controlled by different protein isoform(s) / transcript(s) / gene copy(s). This integrative workflow allows to unravel genetic diversity in polyploid (non-) model crops at the gene variant level.

We identified 2754 proteins
260 identified SAAP
130 with differential peptide abundance

Some of the allele specific transcript levels show deviations from what is expected based on the genomic constitution
27 identified polymorphisms show 100% biased expression levels.

Specific alleles are enriched in genes related to general osmotic stress responses, respiration, ROS scavenging, and HSP.

Identified specific loci

Specific polymorhisms (260 genes) — Quantitative (130 genes) — Not Allele Specific (2494 genes)