

Sunflower pan genome – gene content

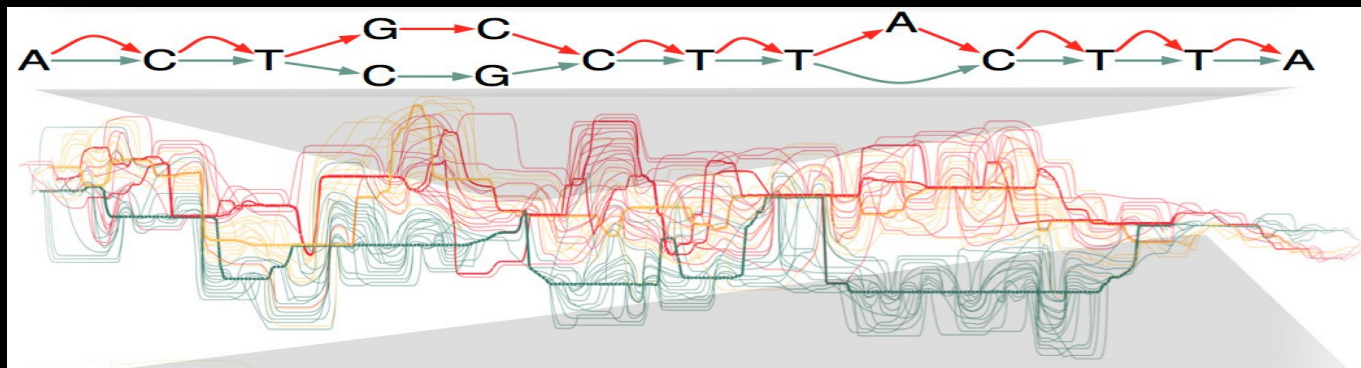


Sariel Hübner

PAG - 2016

What is a PanGenome?

- The full repertoire of genes in a clade or closely related individuals.
- While a genepool of closely related individuals has a substantial variation in gene content the current perspective is limited to a reference genome representing one specific individual.
- A species reference genome should be represented as a network, however the underlying computational and mathematical complications pose analytical limitation to the linear level.

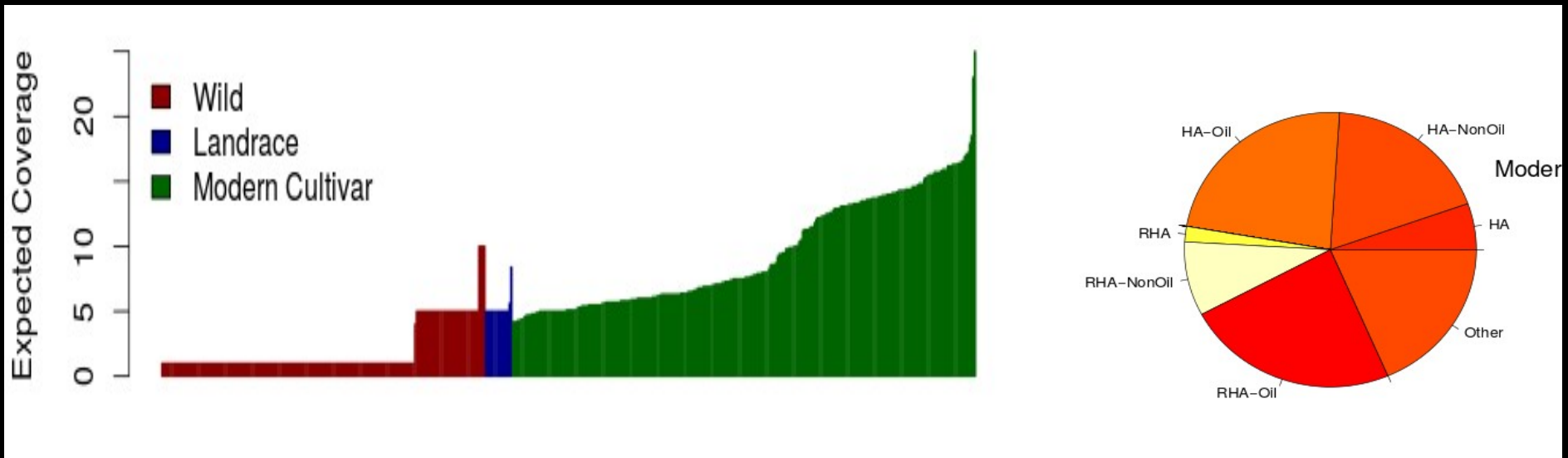


How to build a PanGenome?

- Bacteria: comparison of complete assembled genomes to identify common/unique between strains
- Transcriptome variation: transcripts that are common/unique between individuals
- PanGenome in plants: the soybean pangenome for 7 individuals (Li *et al.* 2014 *NatBiotech*)

The cultivated sunflower PanGenome

- The SAM population: 285 accessions representing most of the variation in cultivated sunflower
- WGSS of 5-25X



Our approach

- Align sequences of each accession to the reference genome
- Extract all unmapped and poorly mapped reads
- *de-novo* assemble sequences separately for each accession (low K-mer in Ray)
- Filter bad/represented assembled sequences
- Annotate

Align to
Ha412.v1.1

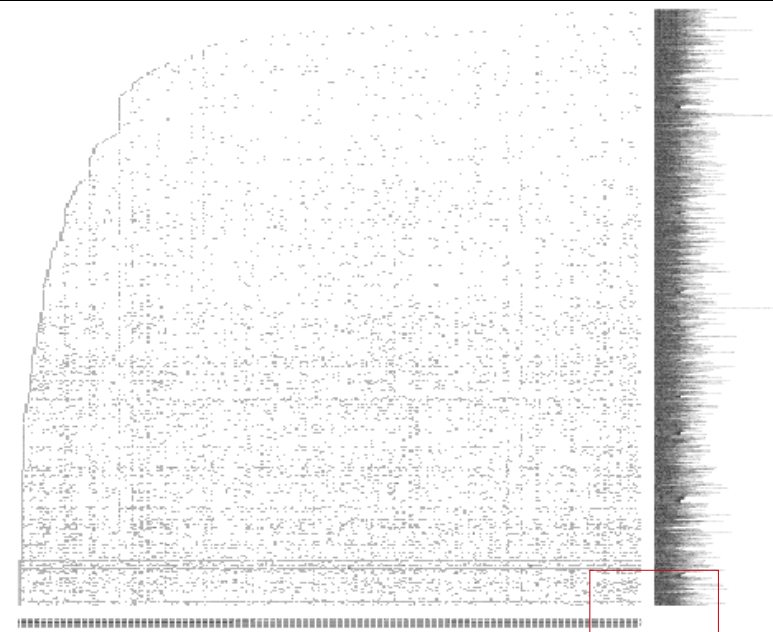
Extract relevant
reads

de-novo
assembly

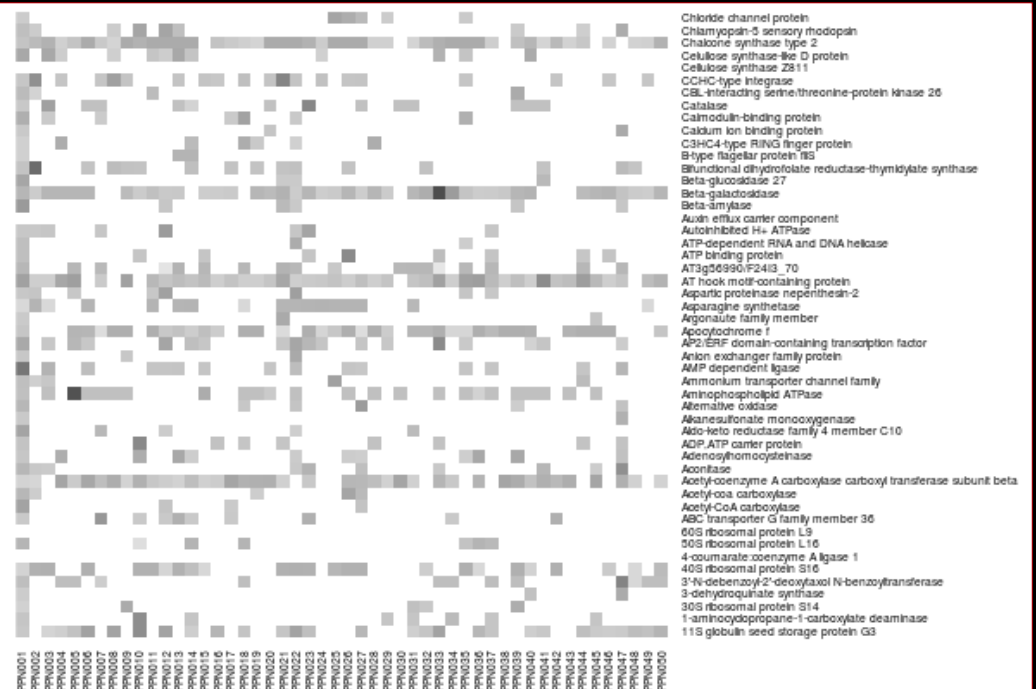
Filter

Annotation

The cultivated sunflower Pan-Genome

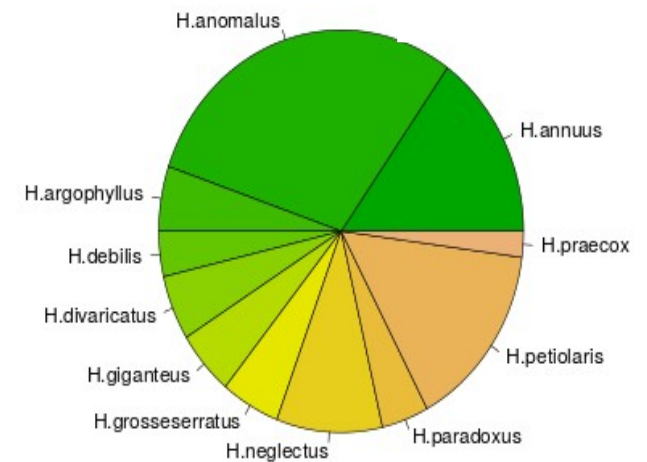
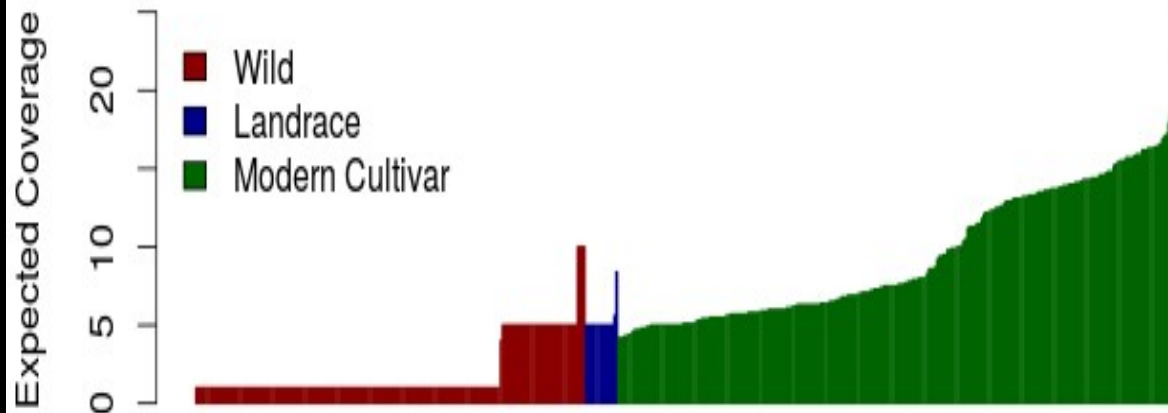


- 21081 sequences
- 13988 annotations
- 3257 gene clusters

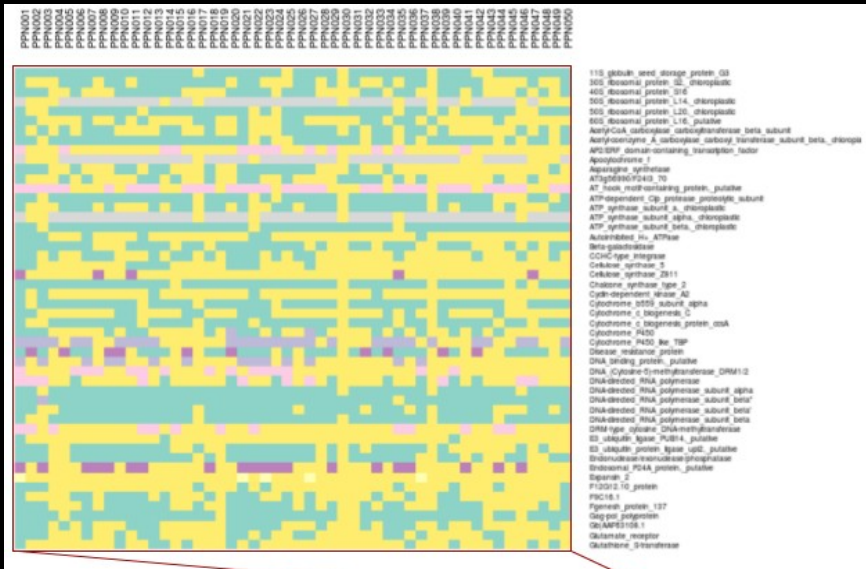


Identifying the wild source of introgressions into the cultivated sunflower

- Sunflower wild relatives: 193 accessions representing 11 species.
- WGSS of 1-5X

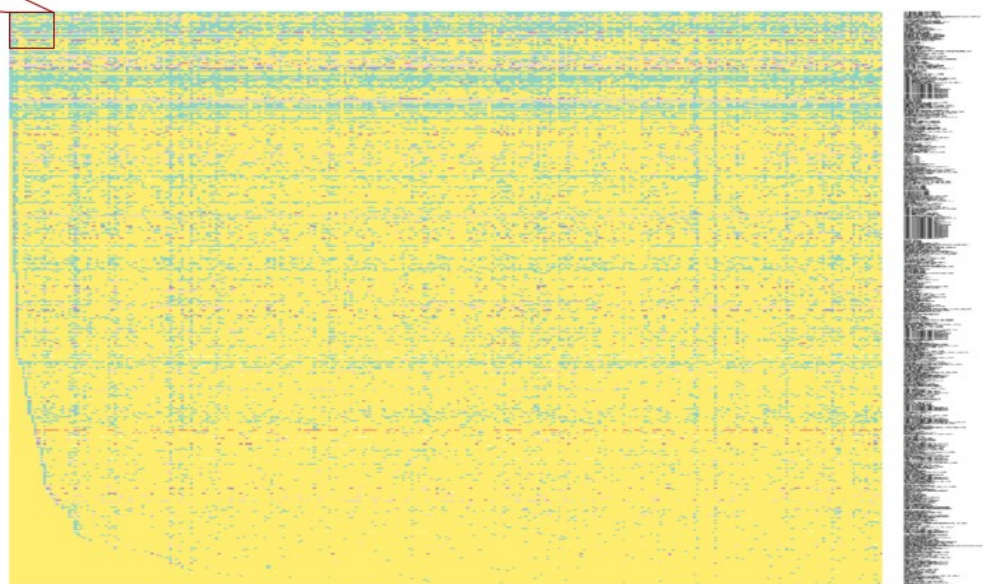
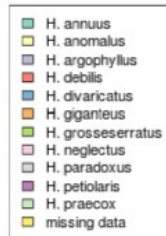


Identifying the wild source of introgressions into the cultivated sunflower

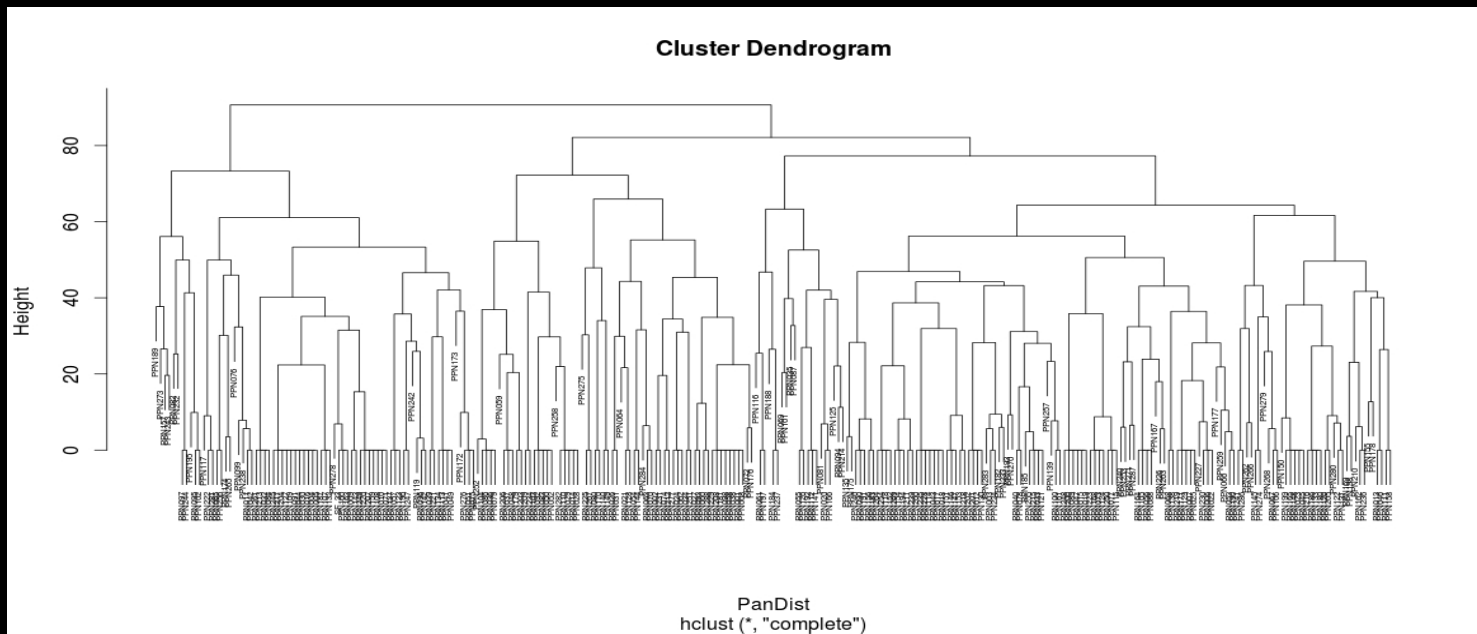
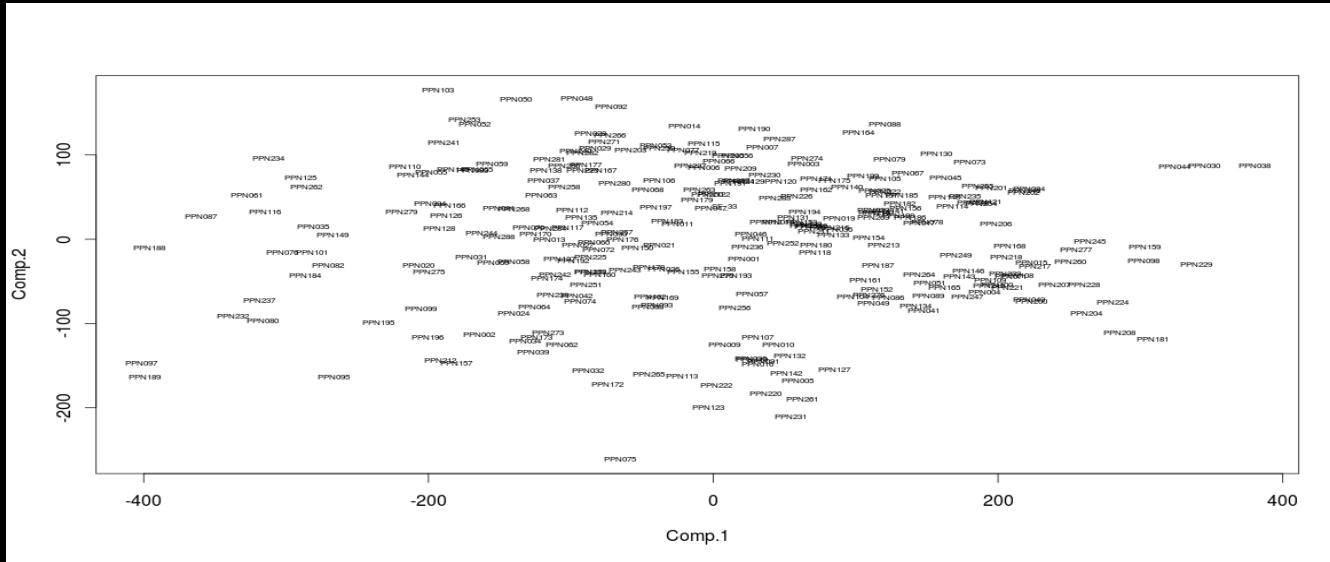


831 genes:

- *H. annuus* 617
- *H. neglectus* 67
- *H. anomalus* 45
- *H. petiolaris* 44
- *H. argophylus* 40
- *H. paradoxus* 12
- *H. debilis* 3
- *H. divaricatus* 2
- *H. praecox* 1

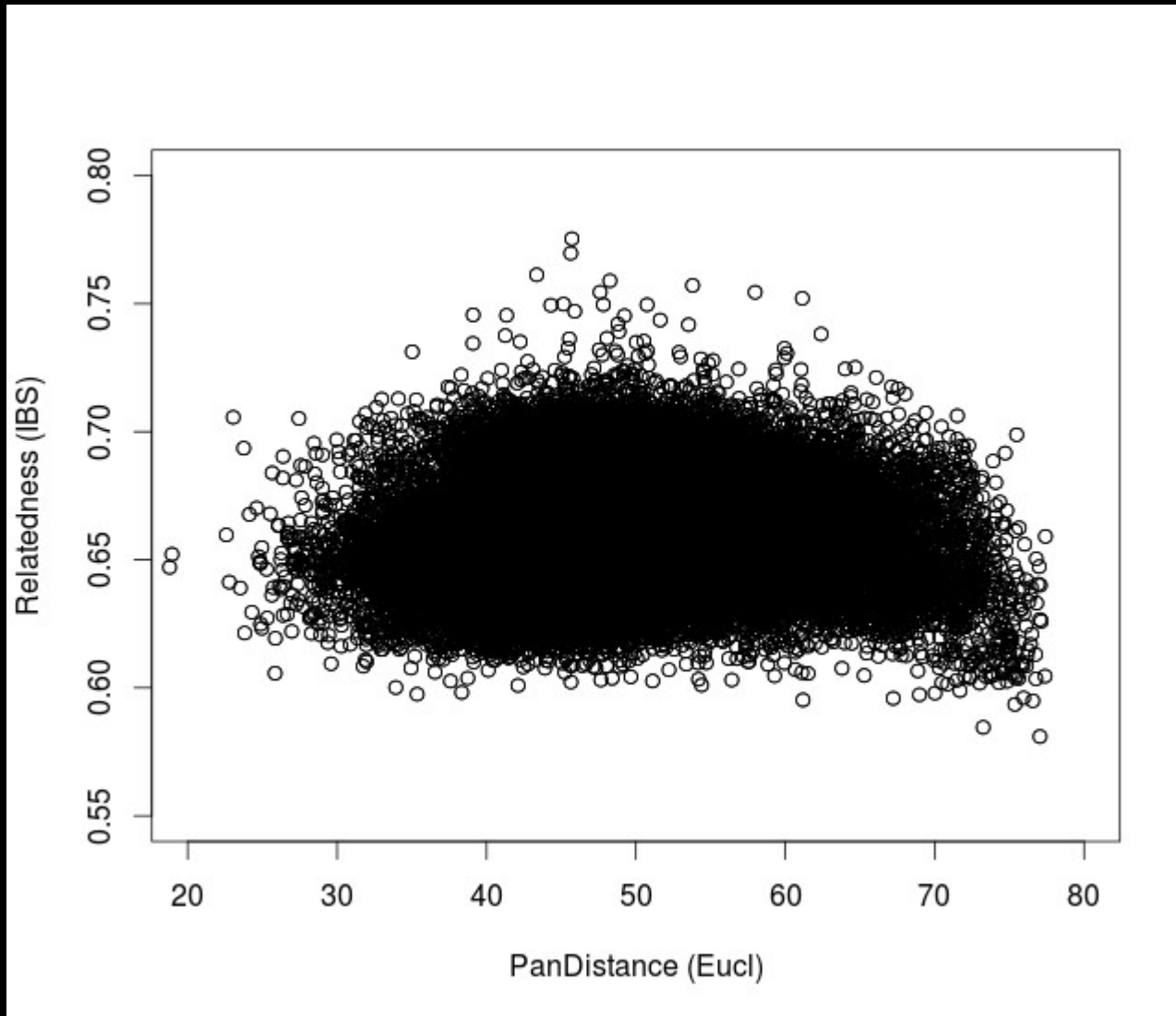


Clustering the cultivated sunflower based on introgressions

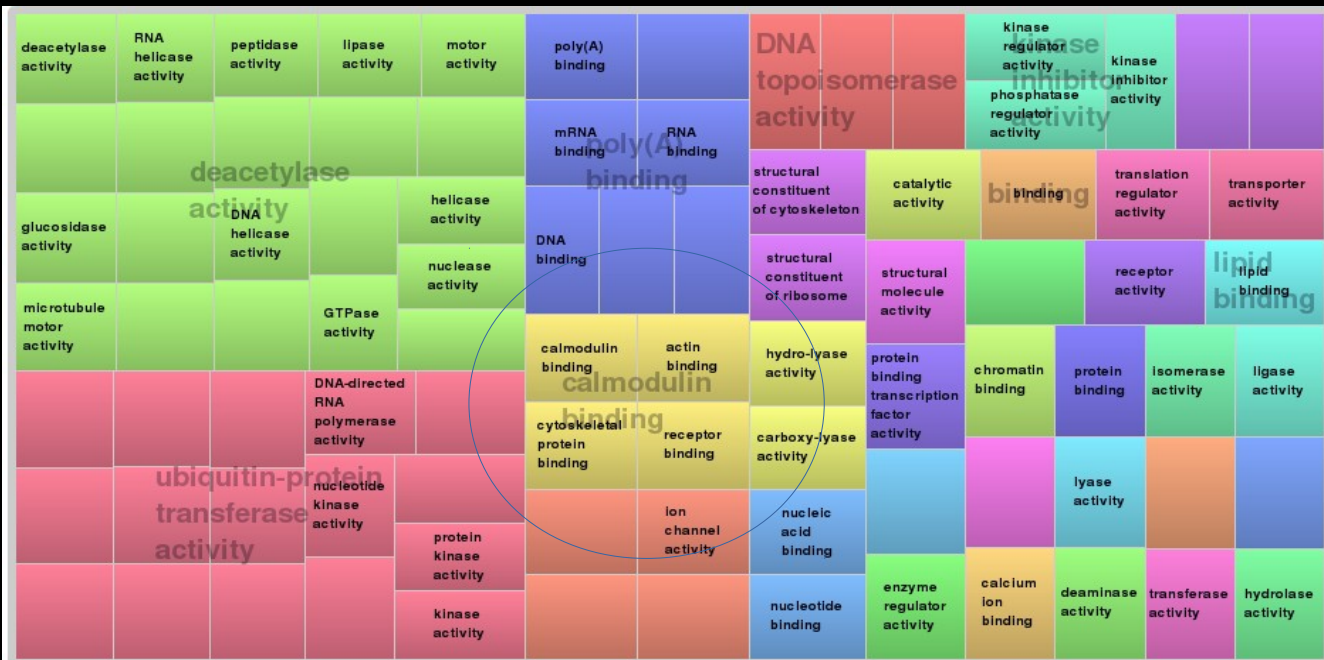
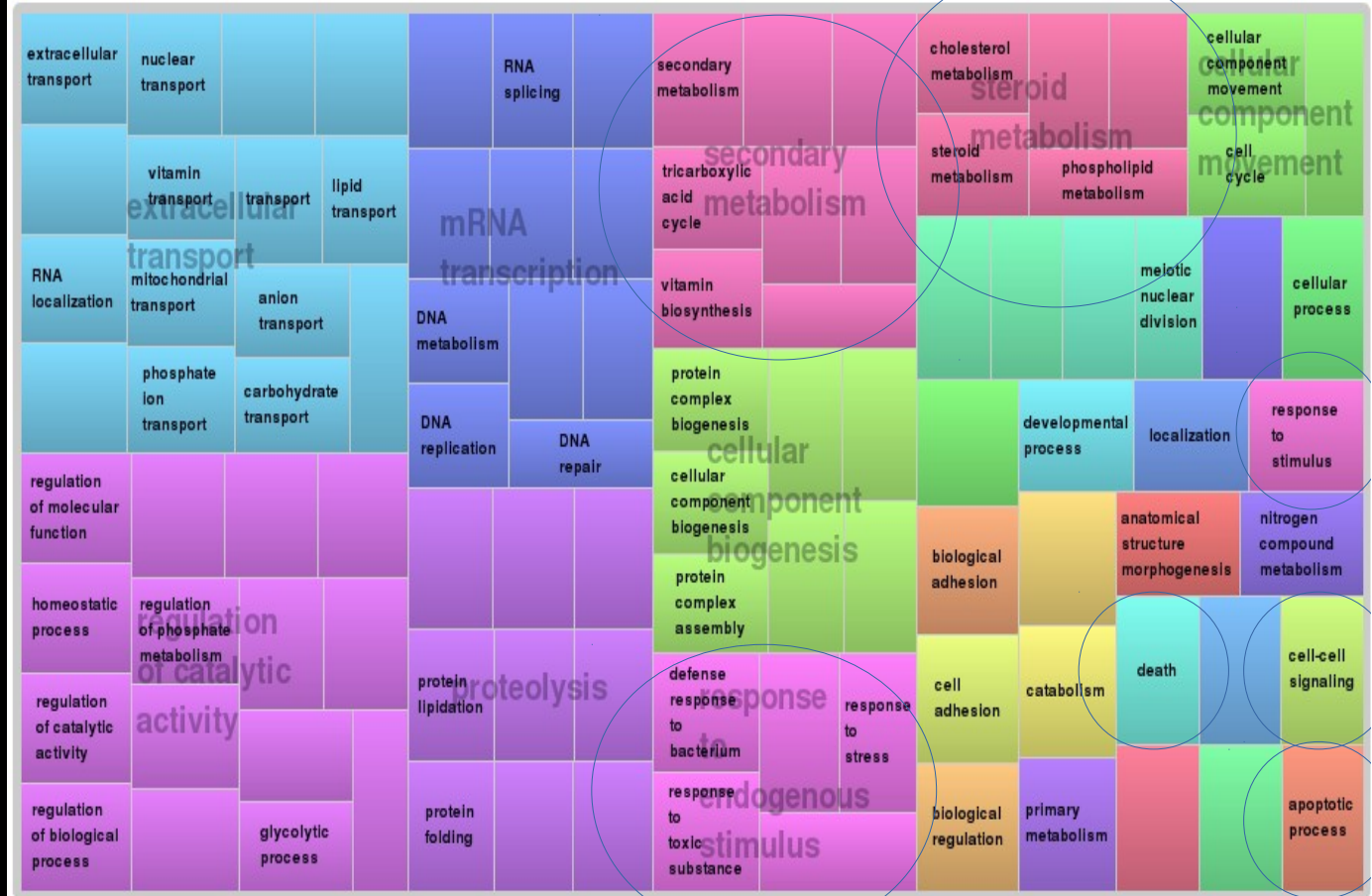


PanGenome is not pedigree derived

Mantel test (100rep, p-value = 0.13)



Response to Stress



Adaptive introgressions



Thanks!

Loren Rieseberg

Emanuel Zigler