Creating Genomic Infrastructure for Mango breeding

Yuval Cohen
Department of Fruit Tree Sciences
Volcani Research Center, ARO, Israel

X International mango symposium, Punta Cana, June, 2013

Acknowledgements

- *Dr. Amir Sherman*
- *Dr. Ron Ophir*
- *Mor Rubinstein*
- *Michal Sharabi*
- *Dr. Ada Rozen*
- *Ravit Eshed*
- *David Saada*
- *Miri Benita*
- *Dr. Mazal Ish-Shalom*

- *Funding: Israeli Ministry of Agriculture*

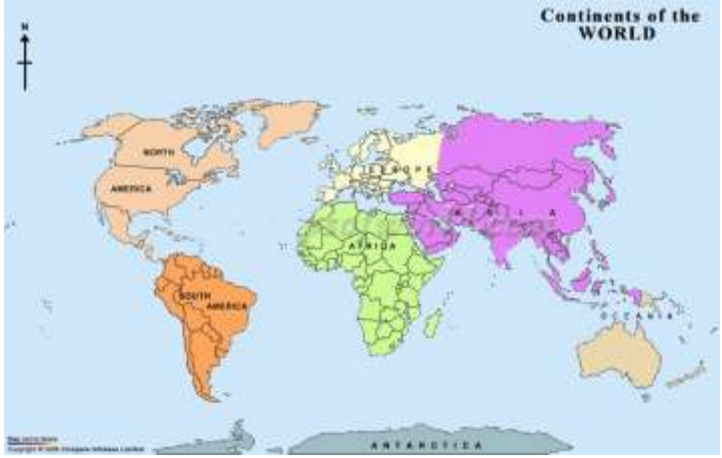
Introduction

- Specific obstacles in mango breeding:
  - Long juvenile stage
  - Limited markers and knowledge on heritability
  - Very low fruit setting.
- Mango breeding is primarily based on screening and selection.
- Tools to enhance the breeding process are required.

The project aim is integration of genetic and genomic approaches for fruit trees and specifically for mango breeding

Breeding, maps and resolution

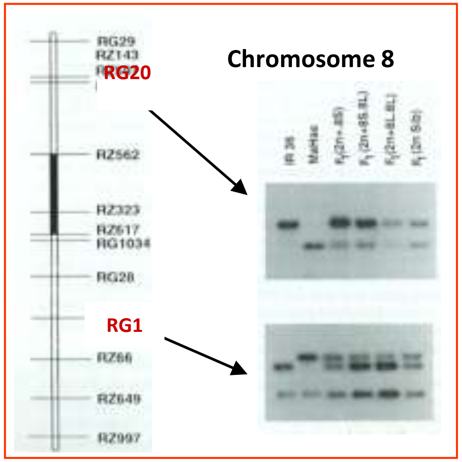
World map from the 16th century



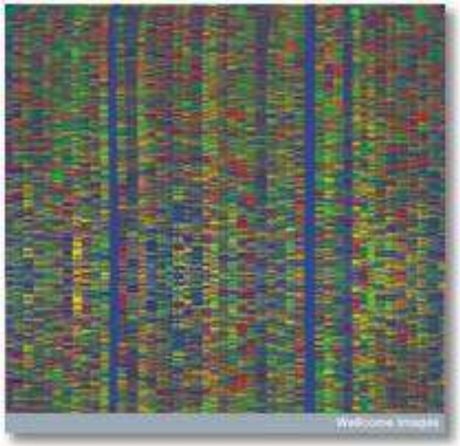
Earth continental maps

A satellite view of Earth

Phenotype and important traits



Genetic maps based on molecular markers



genome sequencing

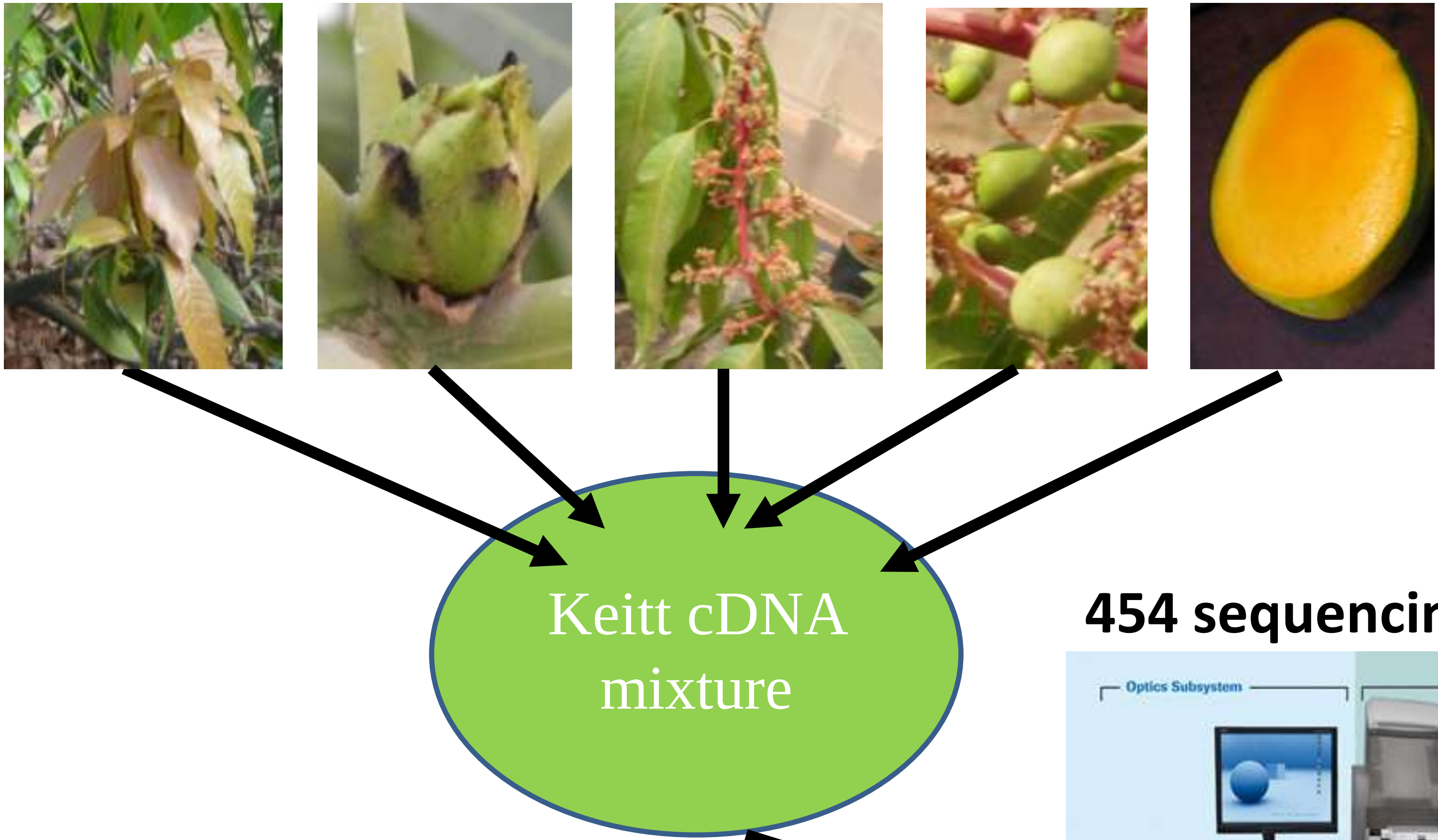
The project main goals:

- Development of a genomic infrastructure for mango using high throughput sequencing
- Identification of molecular markers associated with important horticultural traits
- Use of the identified markers as a main and early selection tool in our mango breeding project

Generating markers for horticultural traits in mango

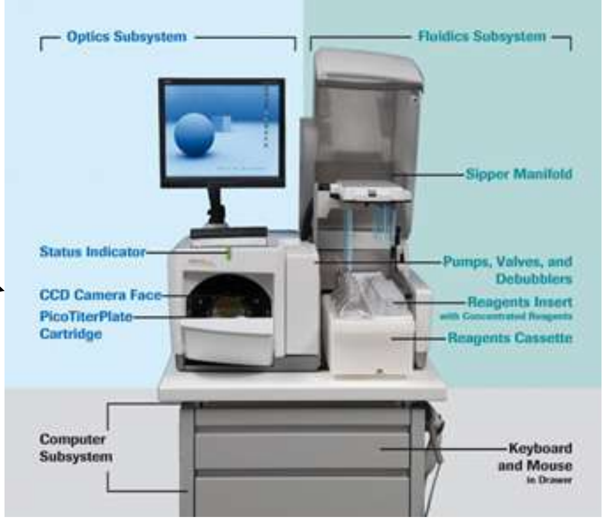
Trait	Keitt	Tommy Atkins
Peal Color	Green	Red
Fibers	Almost none	Some fibers in flesh
Size	Large	Moderate
Harvesting Season	Early	Late

Characterization of the mango transcriptome



Keitt cDNA mixture

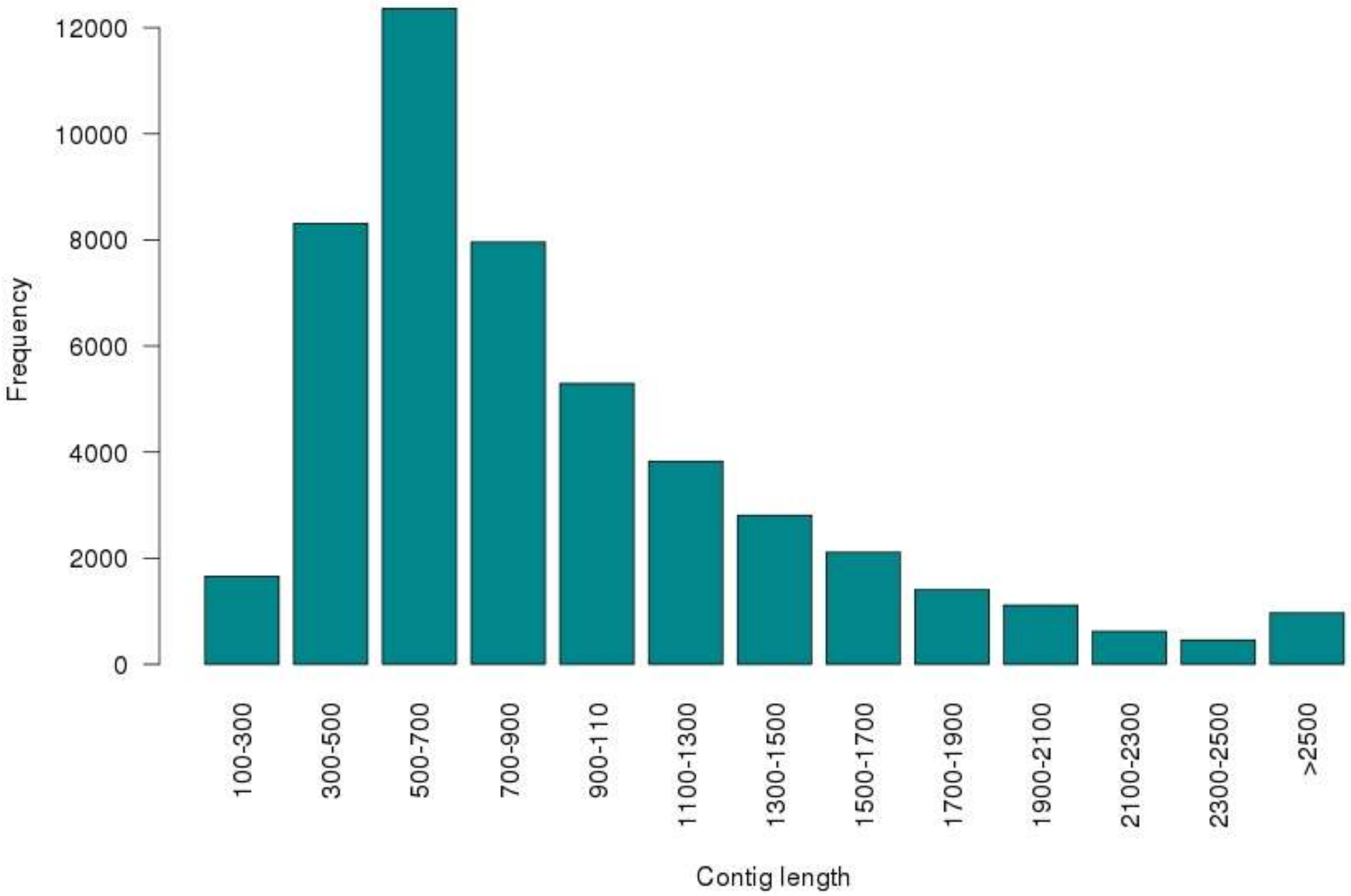
454 sequencing



High throughput sequencing statistics

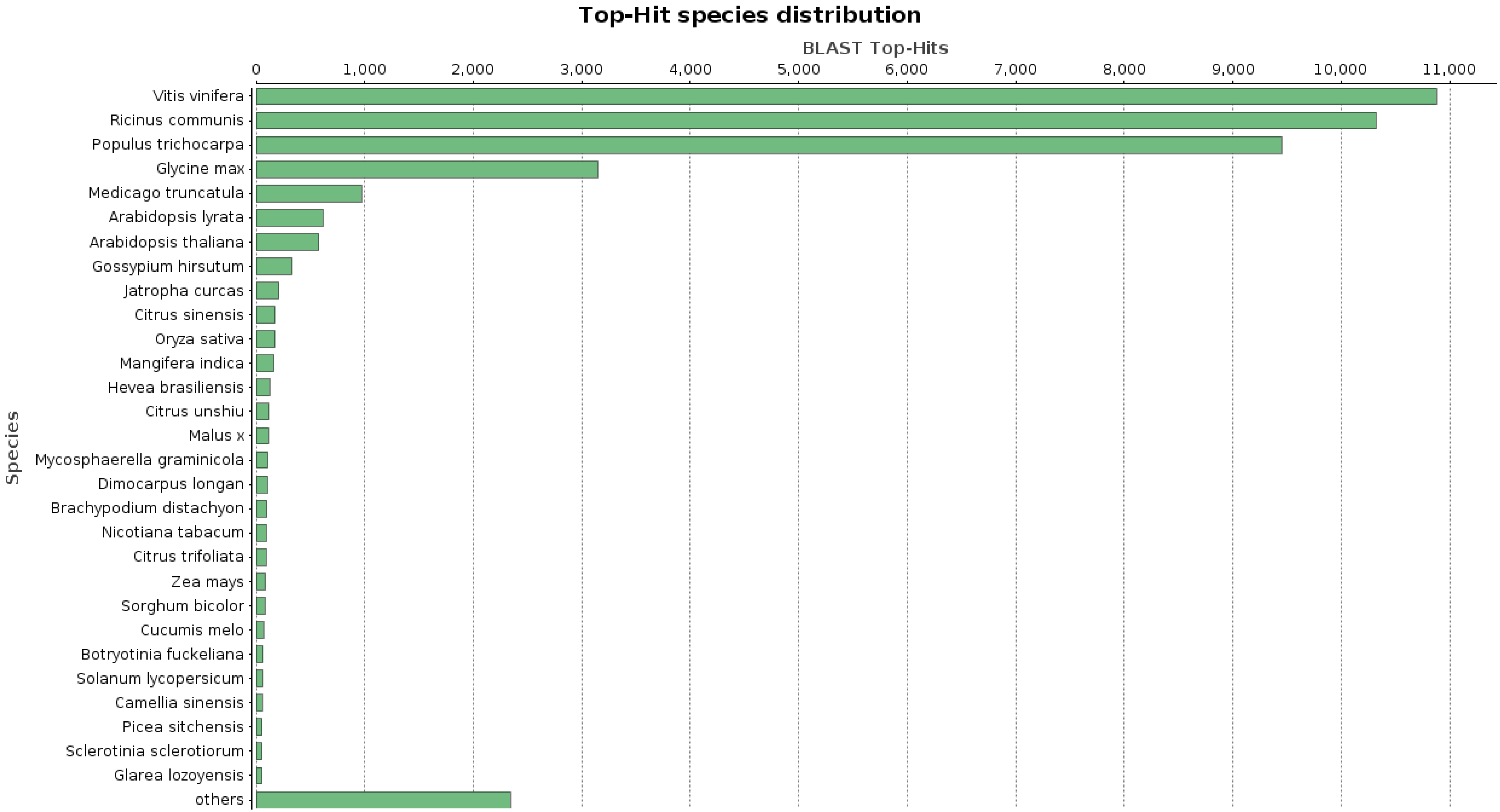
	454 sequencing of Keitt RNA	After CAP3 analysis
Number of Contigs	60,997 contigs	49,717 contigs
Reads/bps	1,329,313 reads	44,845,498 bps
Average coverage	X 17 coverage	

Length Distribution of contigs

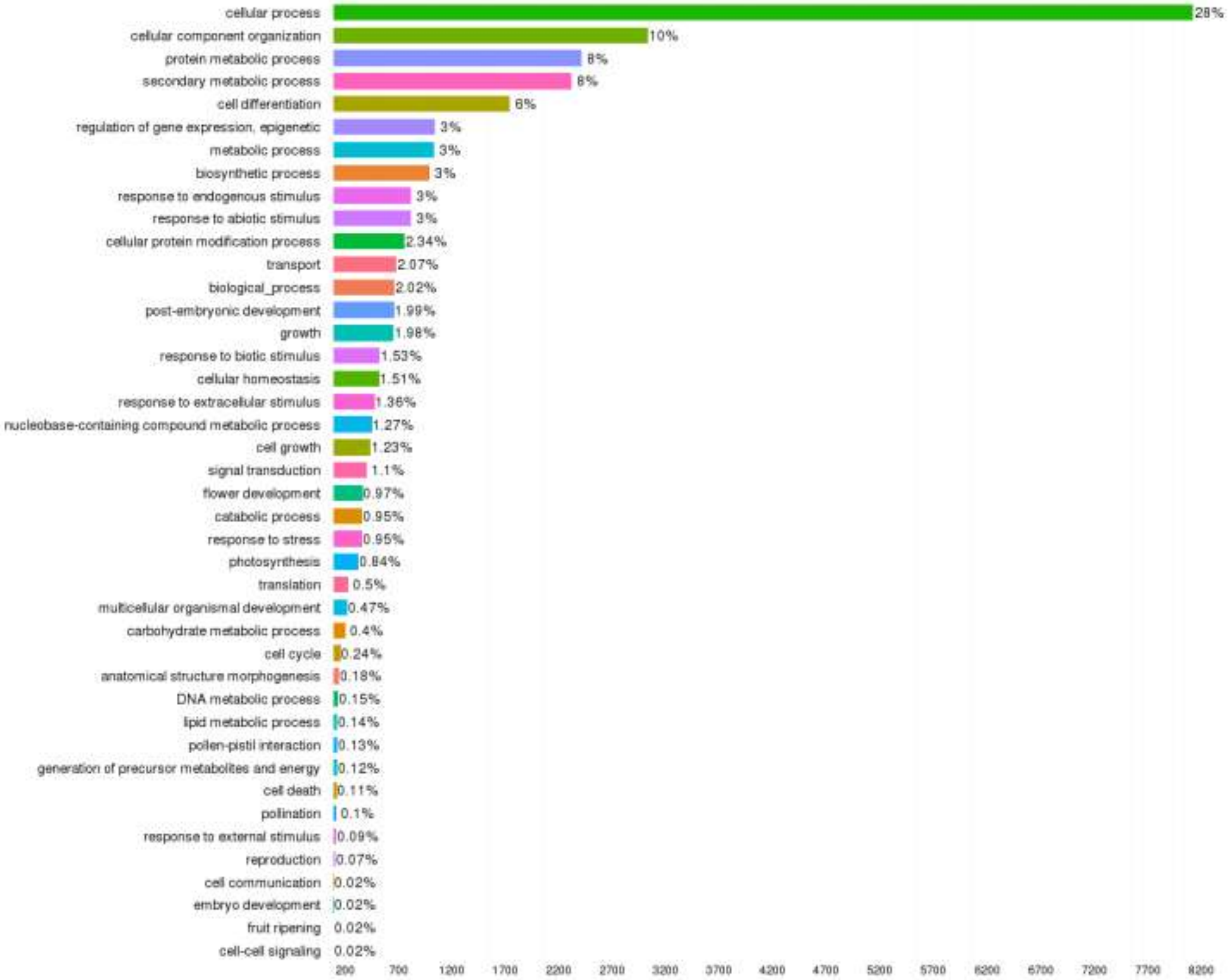


Annotating the Mango Transcriptome

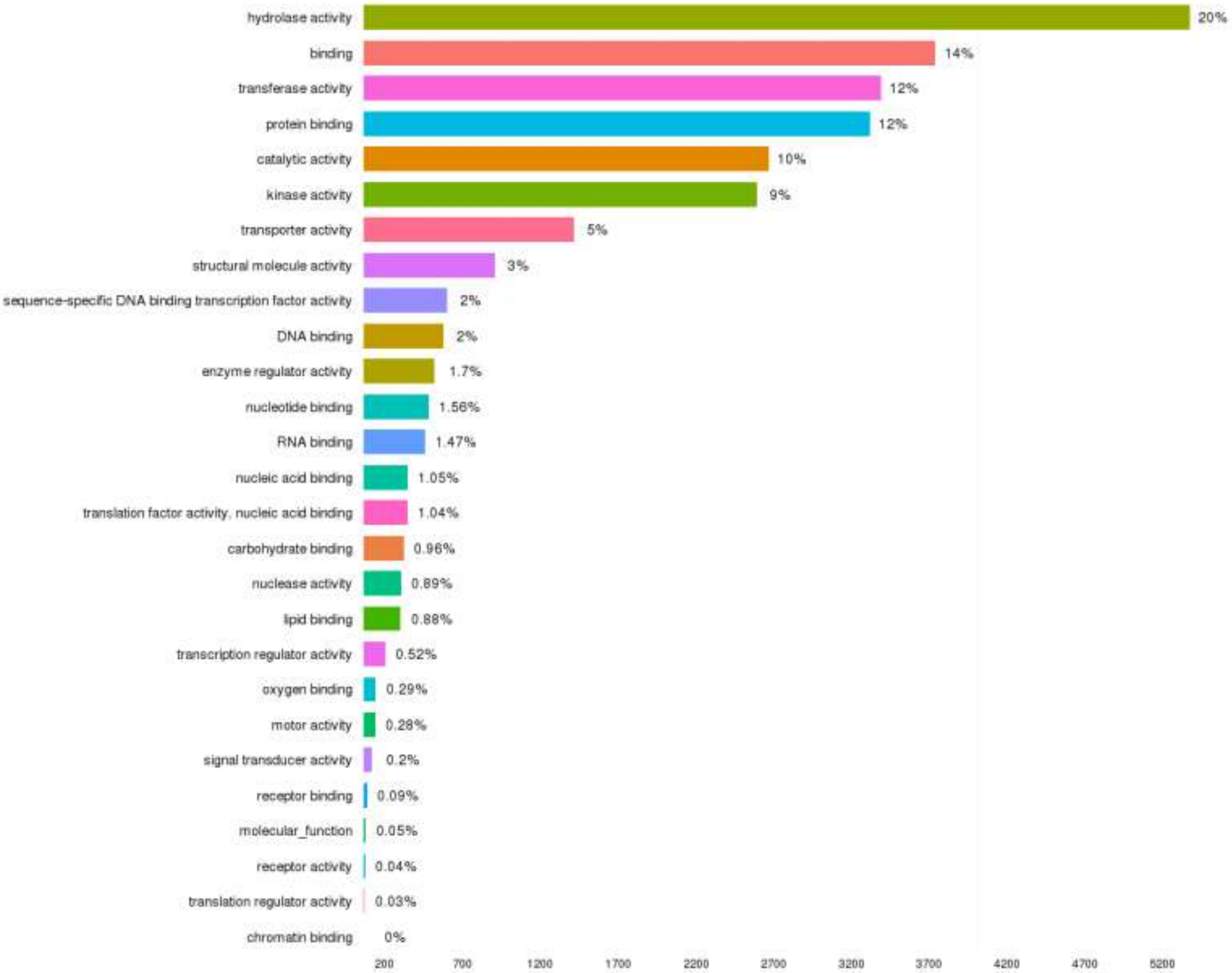
- 42,000 ORFs were found (out of the 49,000 contigs)
- 28,000 were annotated and were mapped using Gene Ontology


Top-Hit species distribution
BLAST Top-Hits
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
11,000
Species
Vitis vinifera
Ricinus communis
Populus trichocarpa
Glycine max
Medicago truncatula
Arabidopsis lyrata
Arabidopsis thaliana
Gossypium hirsutum
Jatropha curcas
Citrus sinensis
Oryza sativa
Mangifera indica
Hevea brasiliensis
Citrus unshiu
Malus x
Mycosphaerella graminicola
Dimocarpus longan
Brachypodium distachyon
Nicotiana tabacum
Citrus trifoliata
Zea mays
Sorghum bicolor
Cucumis melo
Botryotinia fuckeliana
Solanum lycopersicum
Camellia sinensis
Picea sitchensis
Sclerotinia sclerotiorum
Glarea lozoyensis
others

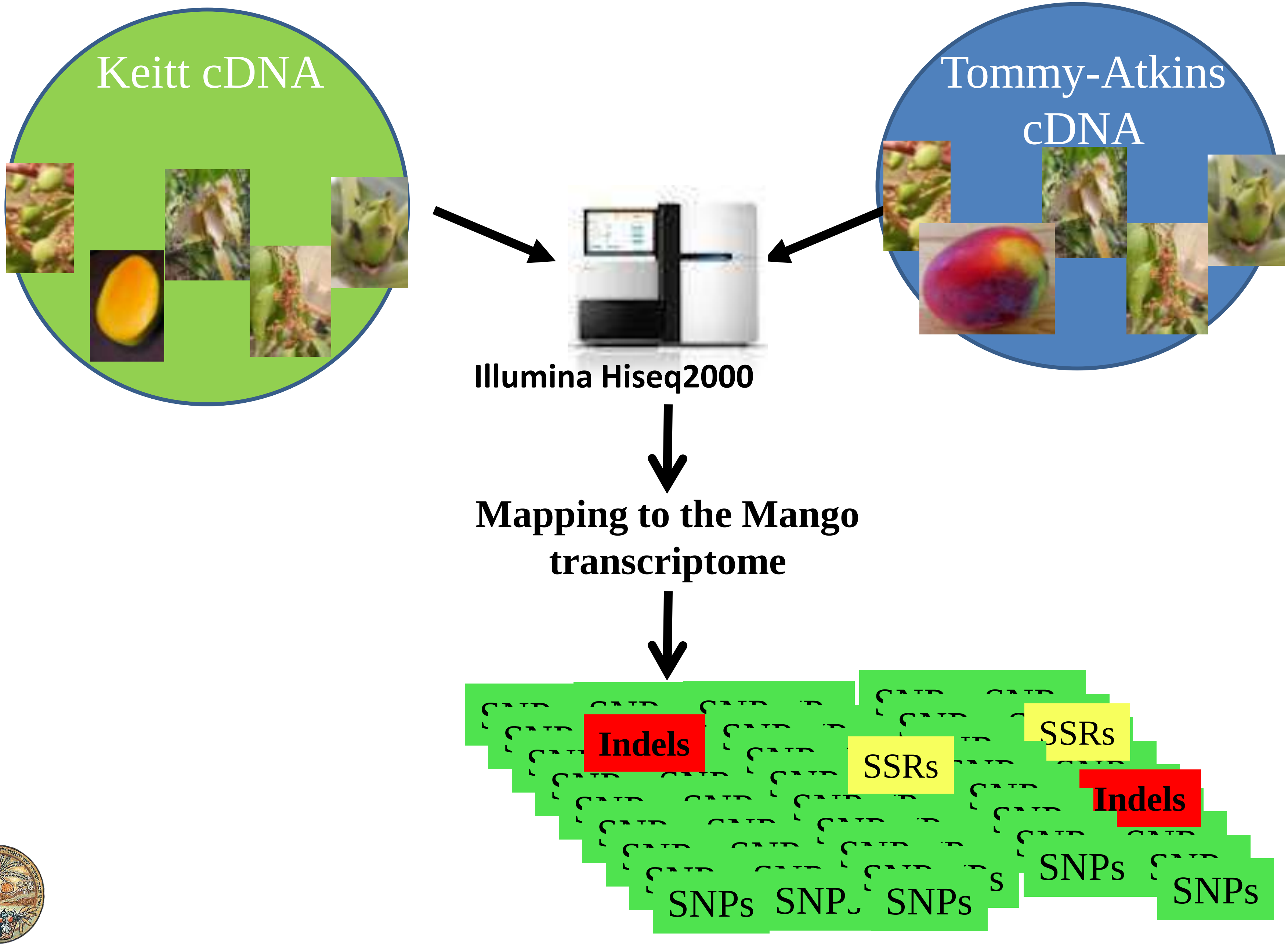
Biological Processes of Mango Transcriptome Annotations



Molecular function of annotated genes



High throughput sequencing for mango SNP discovery



Keitt cDNA

Tommy-Atkins cDNA

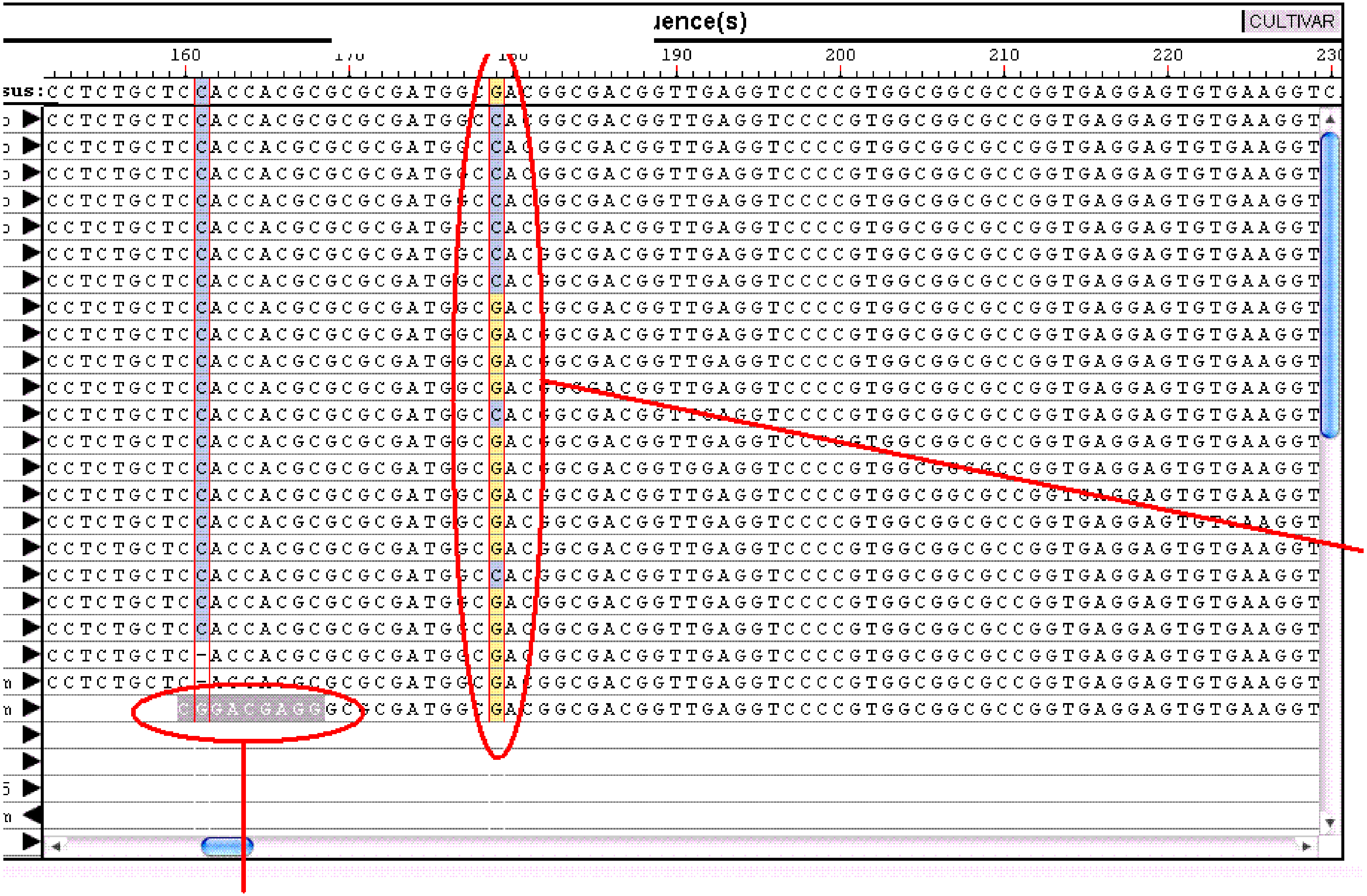
Illumina Hiseq2000

Mapping to the Mango transcriptome

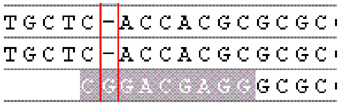
Mapping of Keitt and Tomy Atkins transcriptomes using Illumina HiSeq2000 sequencing

cDNA MIX	Number of reads	Percentage mapped reads	Coverage
Keitt	70,163,264	83.14%	X 129
Tommy-Atkins	69,070,018	79.23%	X 121

SNP identification using high throughput sequencing



```
sus:CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGTC
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTGGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCCACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTCCACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTC-ACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
    CCTCTGCTC-ACCACGCGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
             CGGACGAGGGCGCGATGGCGACGGCGACGGTTGAGGTCCCCGTGGCGGCGCCGGTGAGGAGTGTGAAGGT
```



```
TGCTC-ACCACGCGCGC
TGCTC-ACCACGCGCGC
     CGGACGAGGGCGC
```

Identified Keitt – Tommy Atkins SNPs

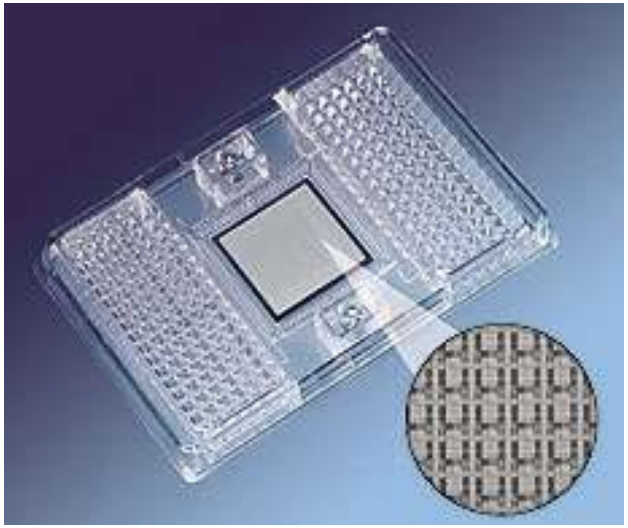
genotype categories	Number of SNPs
Both Homozygous	9,390
Homozygous in Keitt and heterozygous in Tommy	45,861
Heterozygous in Keitt and homozygous in Tommy	36,088
Heterozygous in both	38,511
Total identified SNPs	**129,850**

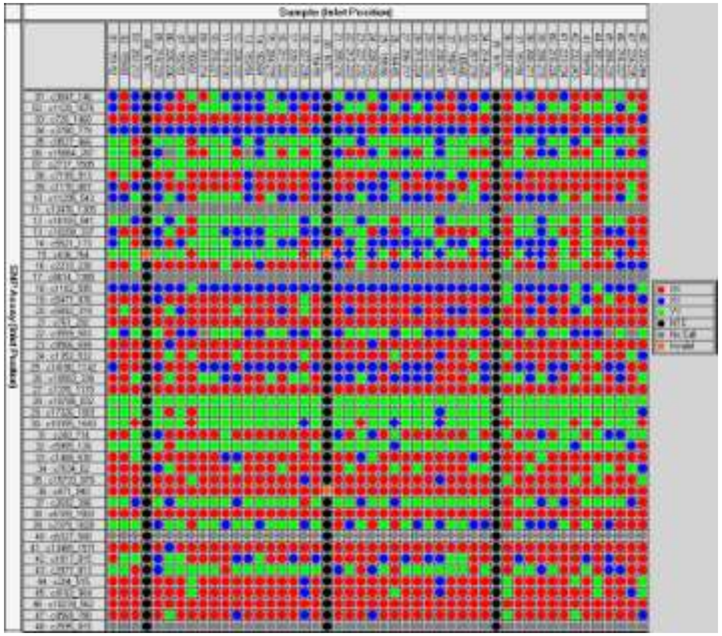
Germplasm genotyping – Fluidigm platform

Germplasm collection

480 SNP assays

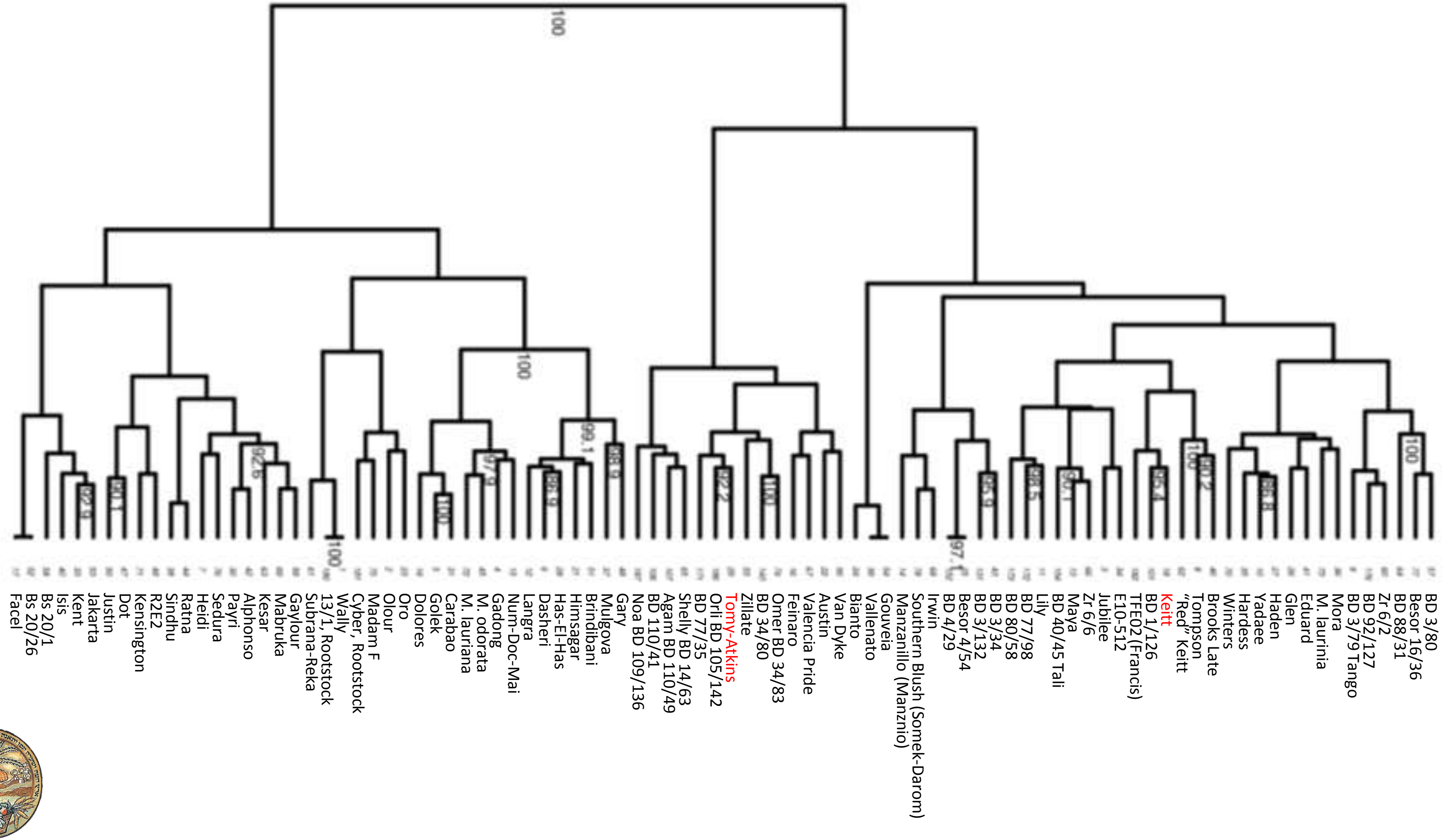
FLUIDIGM EP



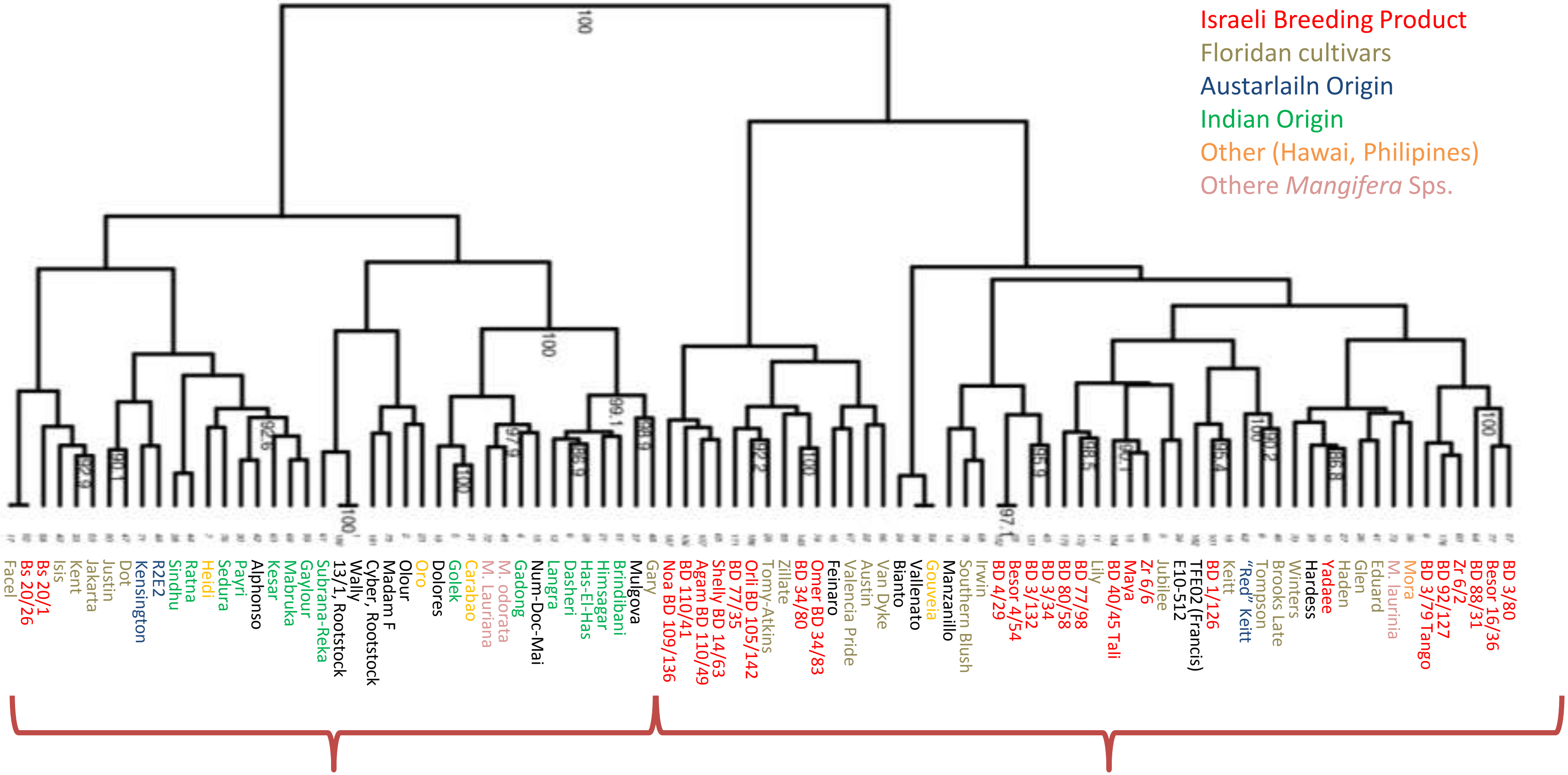


SNP call map

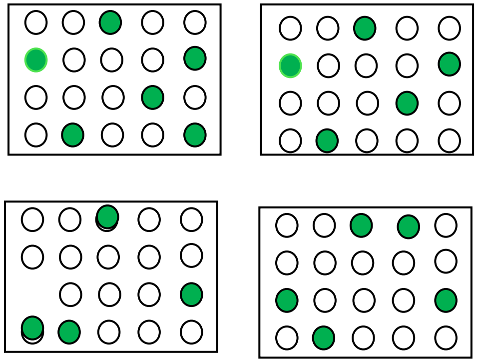
A dendrogram of SNPs variation in mango cultivars and accessions



A dendrogram of SNps variation in mango cultivars and accessions



Identification of markers for horticultural traits using the germ line collection



Identification of specific markers in the germ line collection

Phenotypic characterization of the collection

Discovery of genetic linkages between markers and traits

Use of marker assisted selection in mango breeding

Use of molecular markers in mango breeding

- Creation and identification of progenies (self/cross) from specific parents.
- Generation of specific F1 populations by molecular means.
- Generation of molecular TAGs for mango cultivars
- Use of molecular markers for early identification of key traits in mango breeding (color, fiberness, aroma, resistance to pathogens)

Further analysis will include a larger germplasm and analysis on F2 populations to create a physical genetic map

Acknowledgements

- *Dr. Amir Sherman*
- *Dr. Ron Ophir*
- *Mor Rubinstein*
- *Michal Sharabi*
- *Dr. Ada Rozen*
- *Ravit Eshed*
- *David Saada*
- *Miri Benita*
- *Dr. Mazal Ish-Shalom*

- *Funding: Israeli Ministry of Agriculture*