BMT/15/16
ORIGINAL: English
DATE: May 20, 2016

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA PROFILING IN PARTICULAR

Fifteenth Session

Moscow, Russian Federation, May 24 to 27, 2016

TOWARDS DURABLE DNA DATABASES TO SUPPORT DUS TESTING

Document prepared by an expert from the Netherlands

Disclaimer: this document does not represent UPOV policies or guidance

The Annex to this document contains a copy of a presentation “Towards durable DNA databases to support DUS testing” to be made at its fifteenth session of the Working Group on Biochemical and Molecular Techniques and DNA-Profiling in particular (BMT).

Hedwich Teunissen, Molecular Biologist, Naktuinbouw

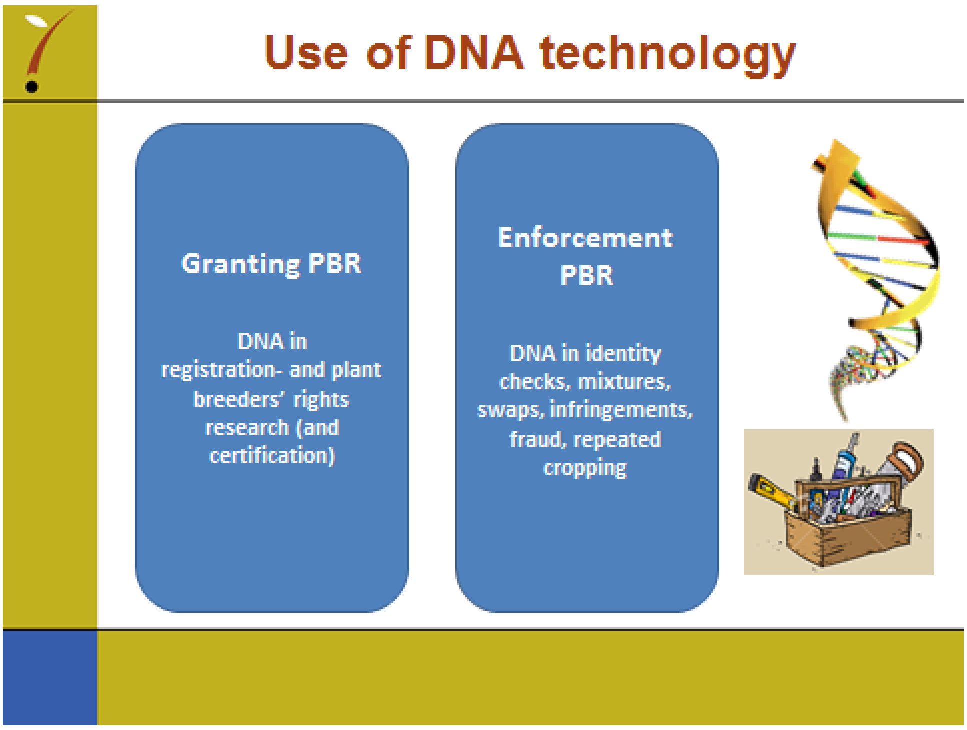
[Annex follows]

ANNEX

TOWARDS DURABLE DNA DATABASES TO SUPPORT DUS TESTING

Towards durable DNA databases to support DUS testing

Hedwich Teunissen, Naktuinbouw R&D
UPOV-BMT/15 – Moscow - May 2016



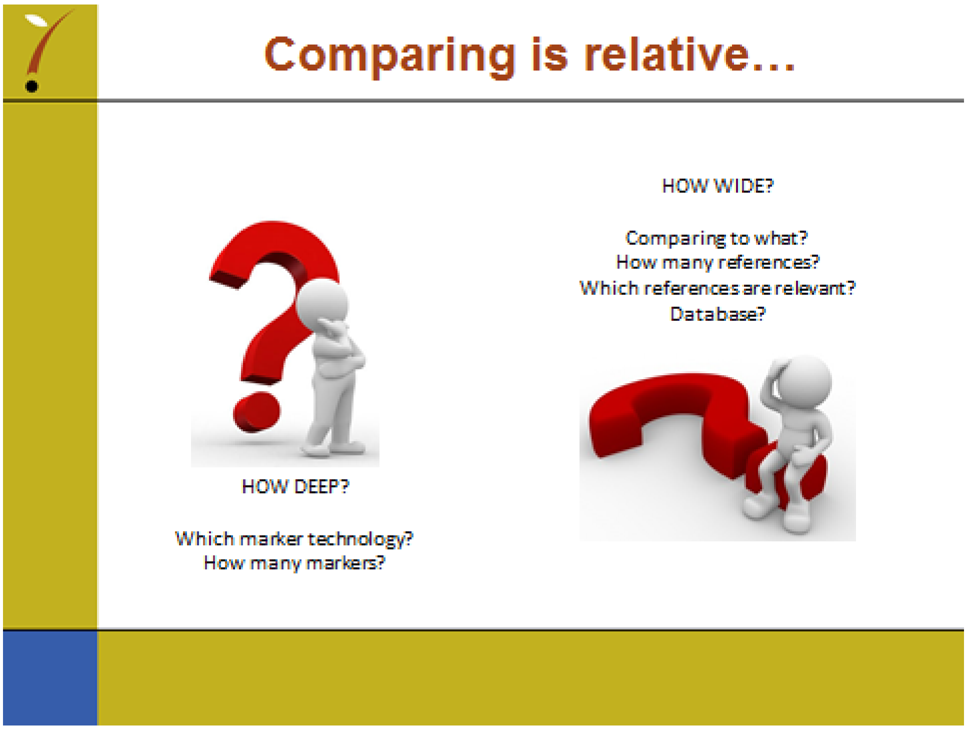
Use of DNA technology

Granting PBR

DNA in registration- and plant breeders' rights research (and certification)

Enforcement PBR

DNA in identity checks, mixtures, swaps, infringements, fraud, repeated cropping



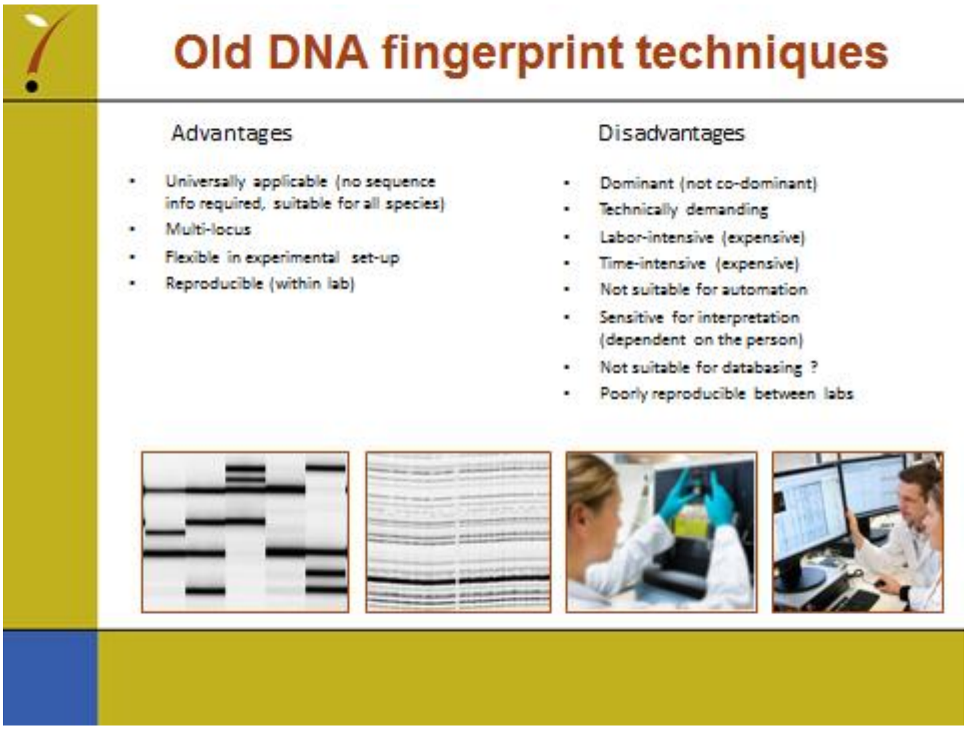
Comparing is relative…

HOW WIDE?

Comparing to what?
How many references?
Which references are relevant?
Database?

HOW DEEP?

Which marker technology?
How many markers?



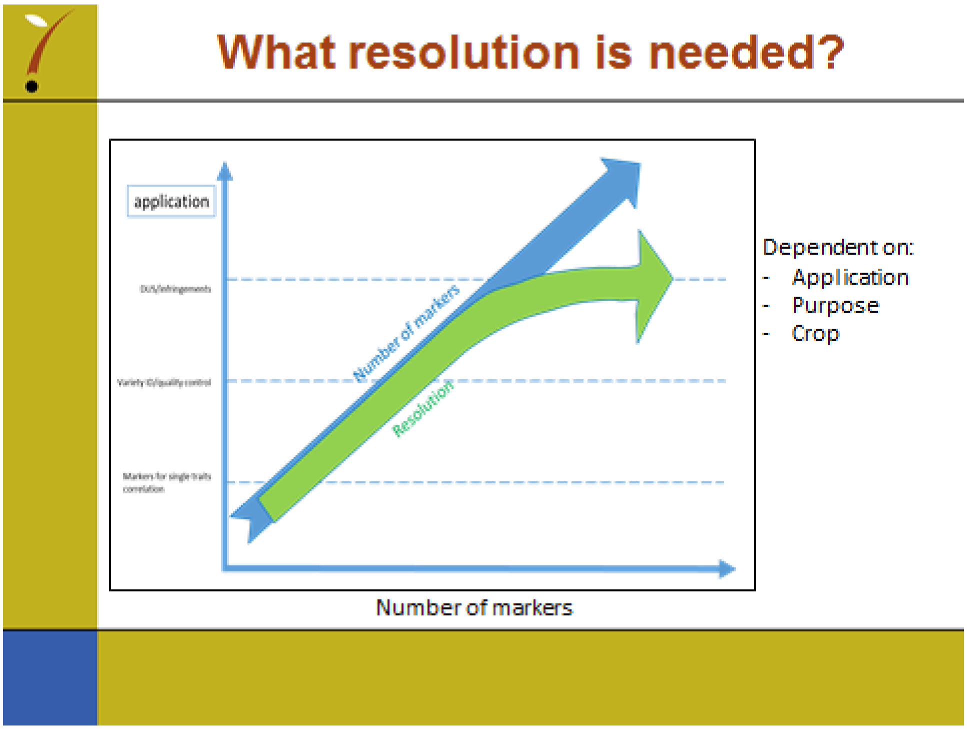
Old DNA fingerprint techniques

Advantages

- Universally applicable (no sequence info required, suitable for all species)
- Multi-locus
- Flexible in experimental set-up
- Reproducible (within lab)

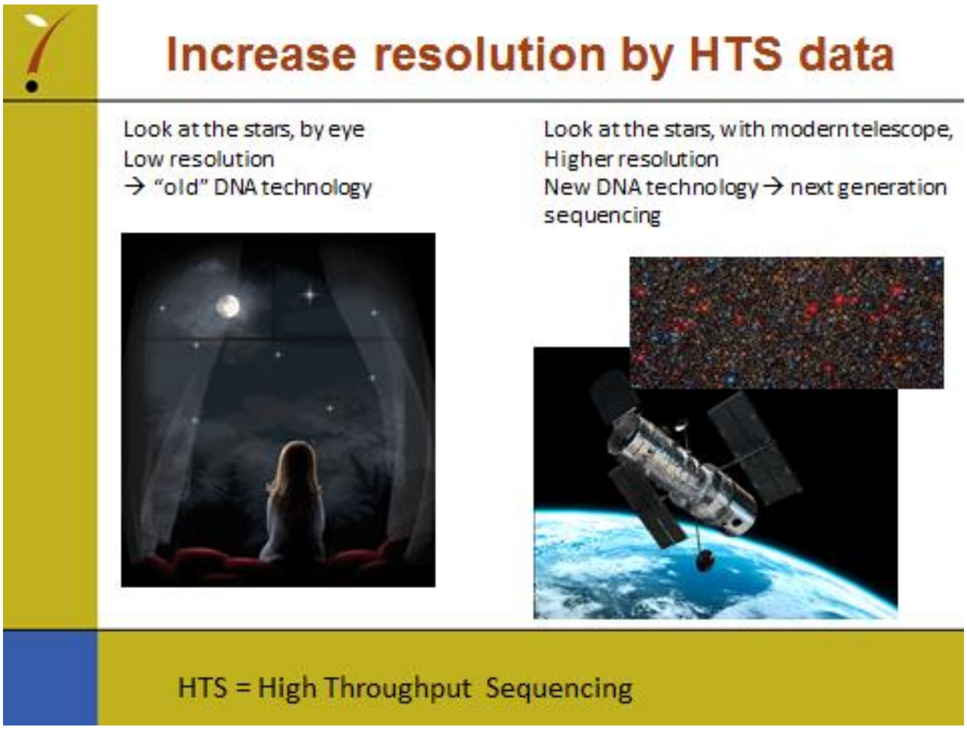
Disadvantages

- Dominant (not co-dominant)
- Technically demanding
- Labor-intensive (expensive)
- Time-intensive (expensive)
- Not suitable for automation
- Sensitive for interpretation (dependent on the person)
- Not suitable for databasing ?
- Poorly reproducible between labs



What resolution is needed?

Dependent on:
- Application
- Purpose
- Crop



Increase resolution by HTS data

Look at the stars, by eye
Low resolution
→ "old" DNA technology

Look at the stars, with modern telescope,
Higher resolution
New DNA technology → next generation sequencing

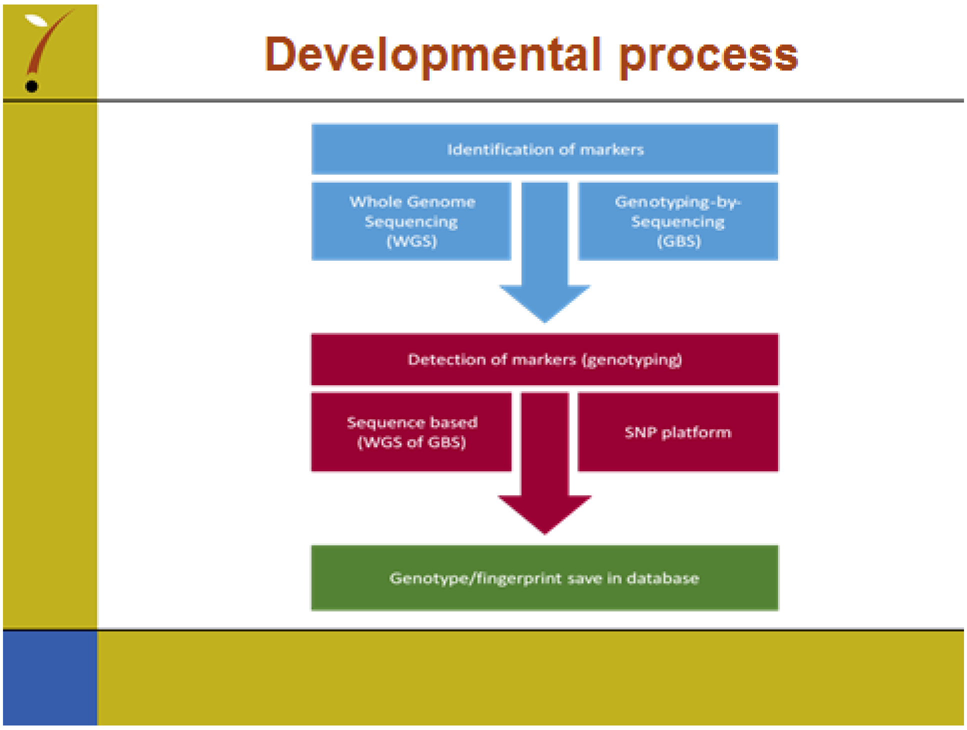
HTS = High Throughput Sequencing

Why now?

Sanger Sequencing
Human genome: 3 x10^9 bp
Time: (1990-2000) 10 years
Costs: $400.000.000

High-throughput (re)-sequencing
Lettuce genome: 3 x10^9 bp
Time: (2016) few months
Costs: $2000

Data analysis not included



Developmental process

Identification of markers

Whole Genome Sequencing (WGS)

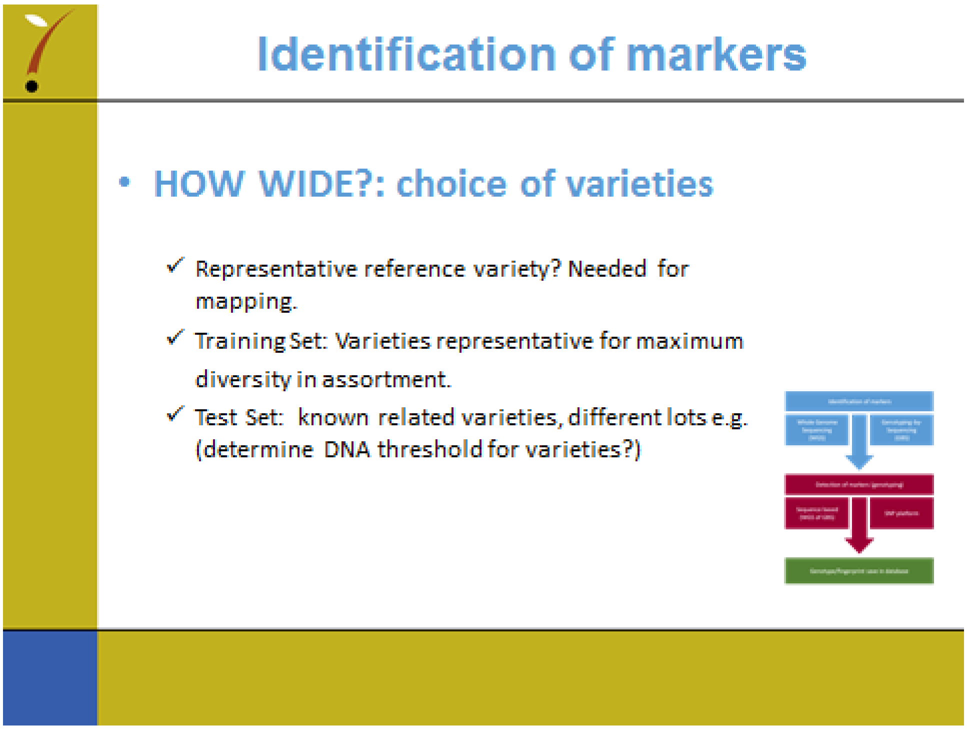
Genotyping-by-Sequencing (GBS)

Detection of markers (genotyping)

Sequence based (WGS of GBS)

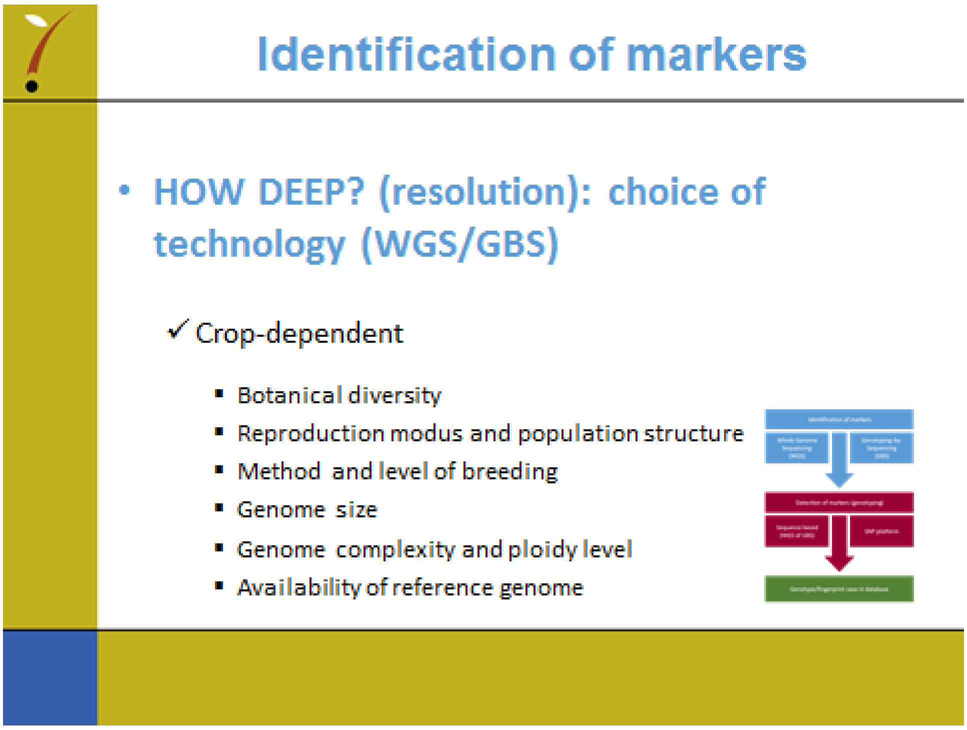
SNP platform

Genotype/fingerprint save in database



Identification of markers

- HOW WIDE?: choice of varieties
  - ✓ Representative reference variety? Needed for mapping.
  - ✓ Training Set: Varieties representative for maximum diversity in assortment.
  - ✓ Test Set: known related varieties, different lots e.g. (determine DNA threshold for varieties?)

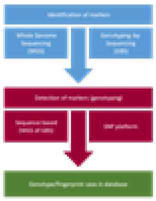


Identification of markers

- HOW DEEP? (resolution): choice of technology (WGS/GBS)
  - ✓ Crop-dependent
    - Botanical diversity
    - Reproduction modus and population structure
    - Method and level of breeding
    - Genome size
    - Genome complexity and ploidy level
    - Availability of reference genome

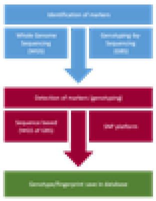
Identification of markers

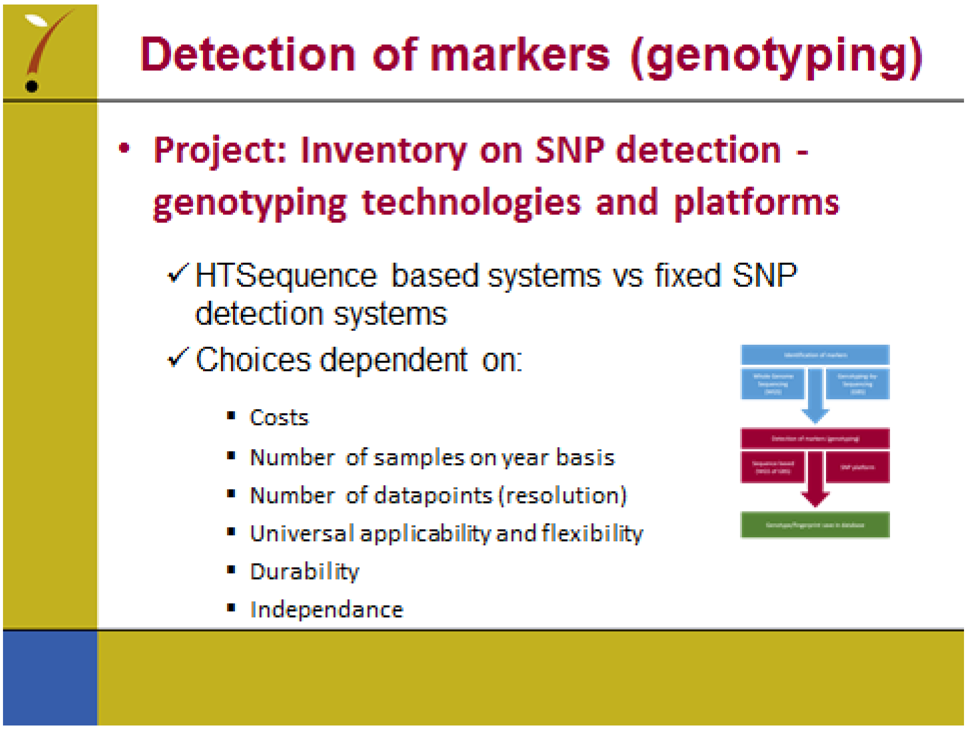
- **Project: compare GBS and WGS**
  - ✓ Deliverables: list of criteria
    - Botanical diversity
    - Reproduction modus and population structure
    - Method and level of breeding
    - Genome size
    - Genome complexity and ploidy level
    - Availability of reference genome



Identification of markers

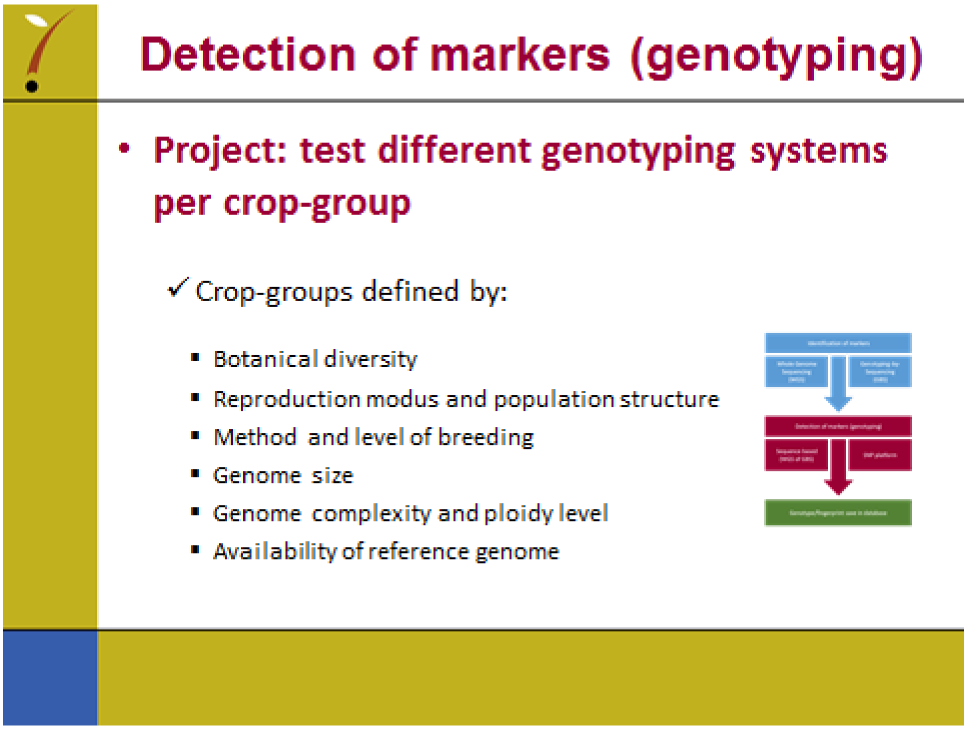
- **Project: make an inventory of information already available**
  - ✓ Public domain
  - ✓ Breeding companies – possibilities to share information





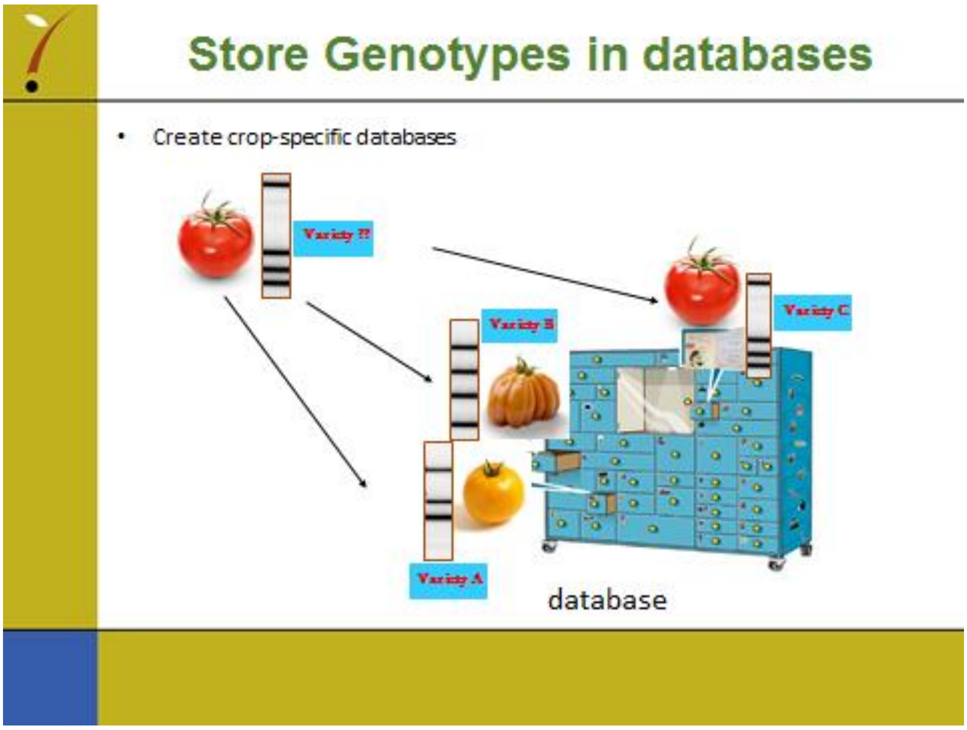
Detection of markers (genotyping)

- **Project: Inventory on SNP detection - genotyping technologies and platforms**
  - ✓ HTSequence based systems vs fixed SNP detection systems
  - ✓ Choices dependent on:
    - Costs
    - Number of samples on year basis
    - Number of datapoints (resolution)
    - Universal applicability and flexibility
    - Durability
    - Independance



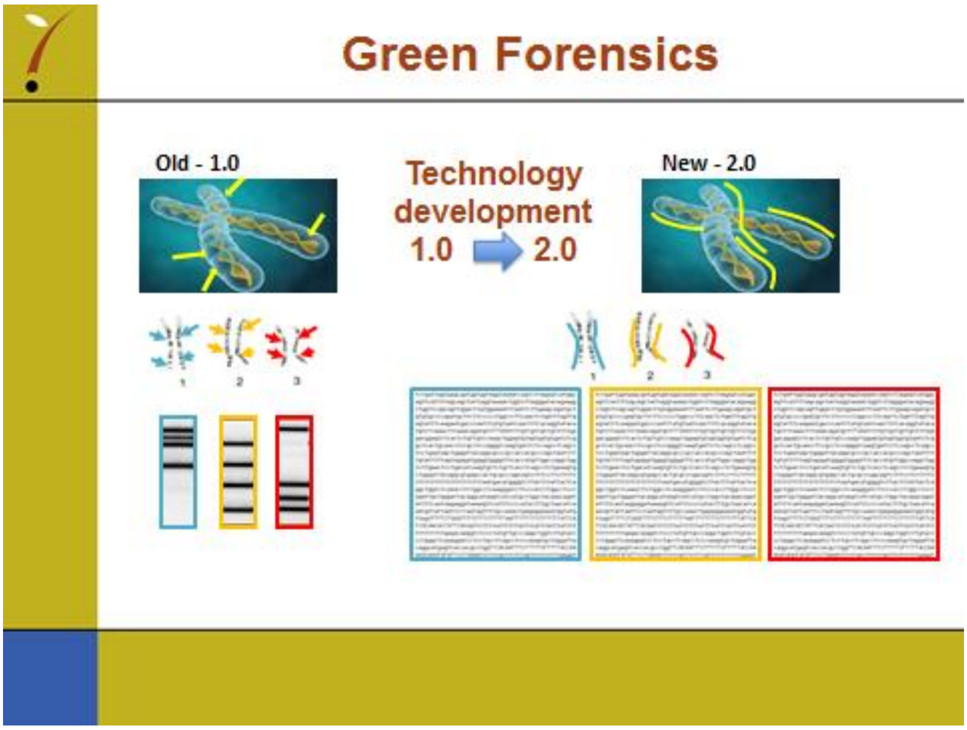
Detection of markers (genotyping)

- **Project: test different genotyping systems per crop-group**
  - ✓ Crop-groups defined by:
    - Botanical diversity
    - Reproduction modus and population structure
    - Method and level of breeding
    - Genome size
    - Genome complexity and ploidy level
    - Availability of reference genome

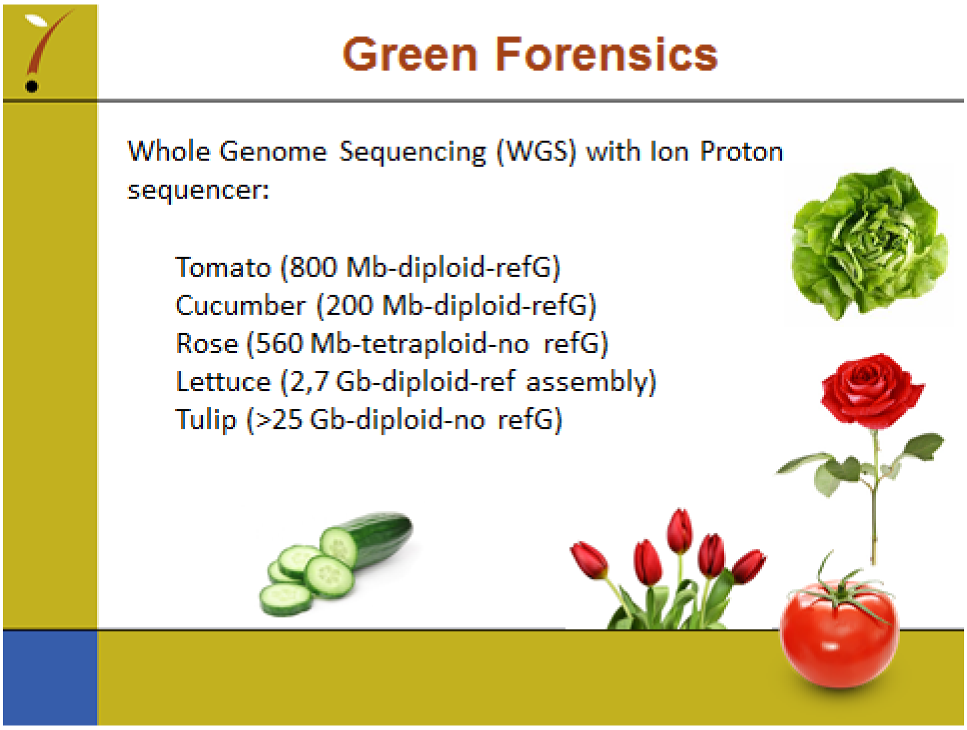


Store Genotypes in databases

- Create crop-specific databases



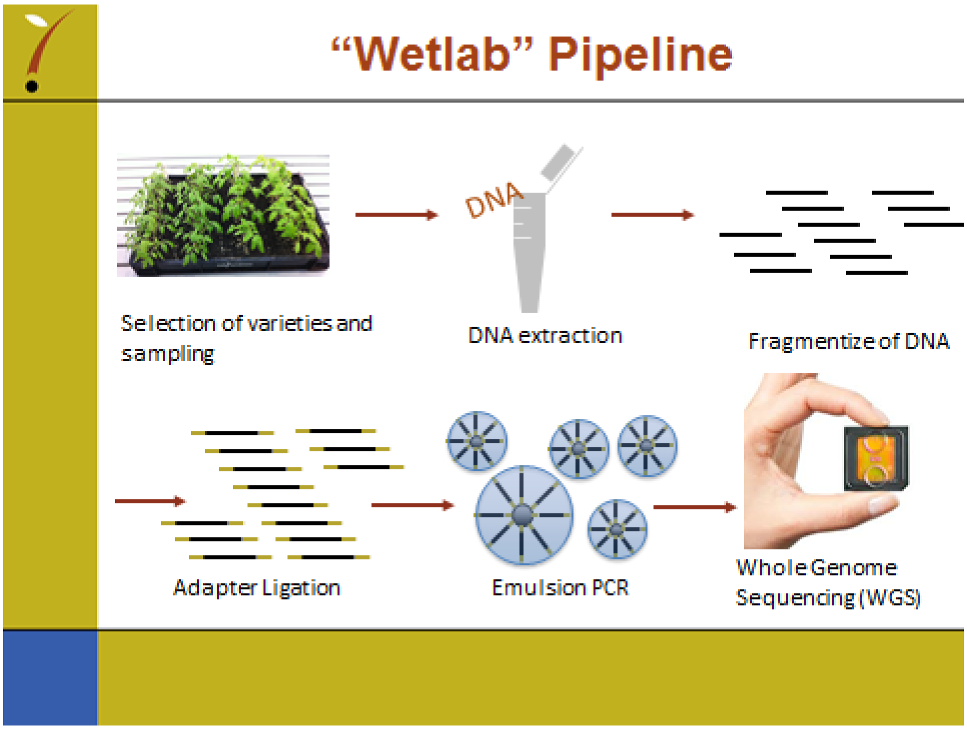
Green Forensics



Green Forensics

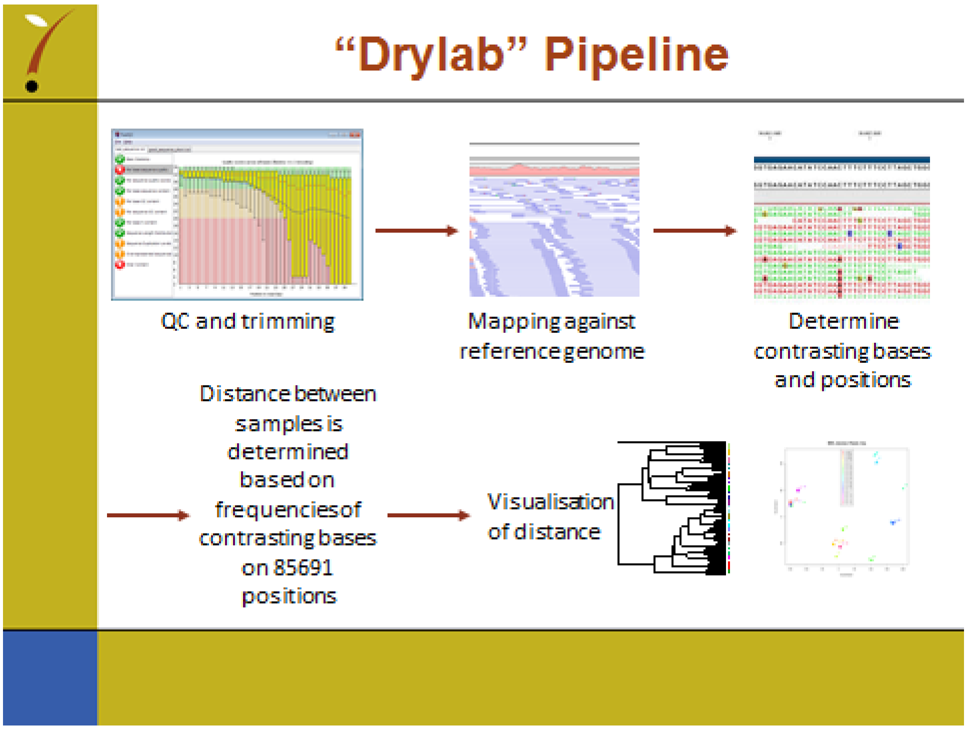
Whole Genome Sequencing (WGS) with Ion Proton sequencer:

Tomato (800 Mb-diploid-refG)
Cucumber (200 Mb-diploid-refG)
Rose (560 Mb-tetraploid-no refG)
Lettuce (2,7 Gb-diploid-ref assembly)
Tulip (>25 Gb-diploid-no refG)



“Wetlab” Pipeline

Selection of varieties and sampling → DNA extraction → Fragmentize of DNA → Adapter Ligation → Emulsion PCR → Whole Genome Sequencing (WGS)



“Drylab” Pipeline

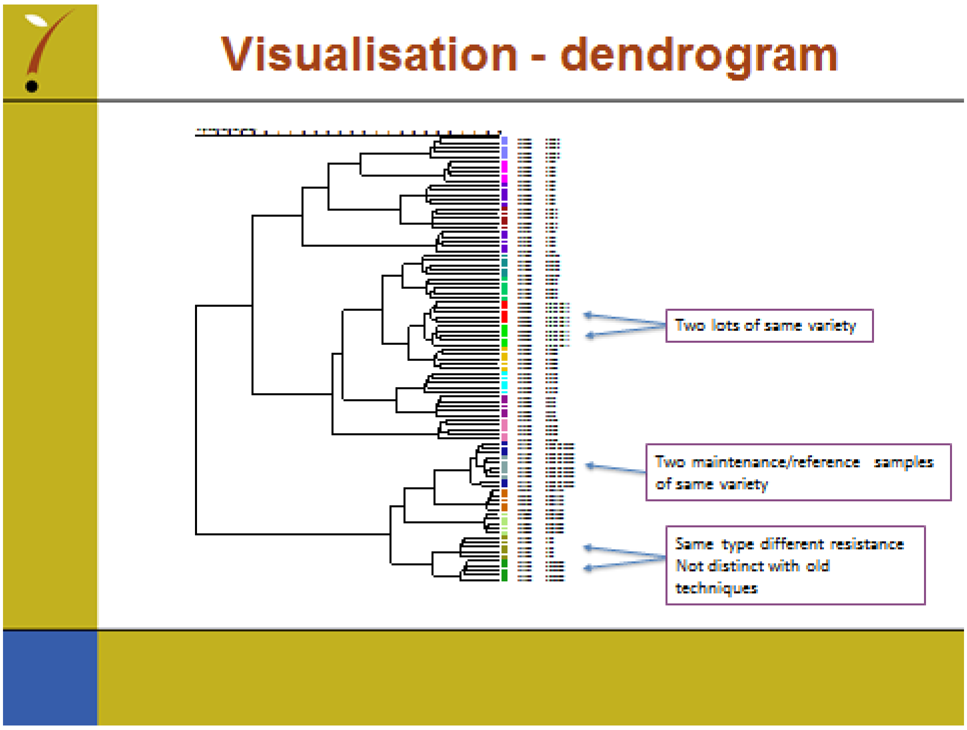
QC and trimming

Mapping against reference genome

Determine contrasting bases and positions

Distance between samples is determined based on frequencies of contrasting bases on 85691 positions

Visualisation of distance

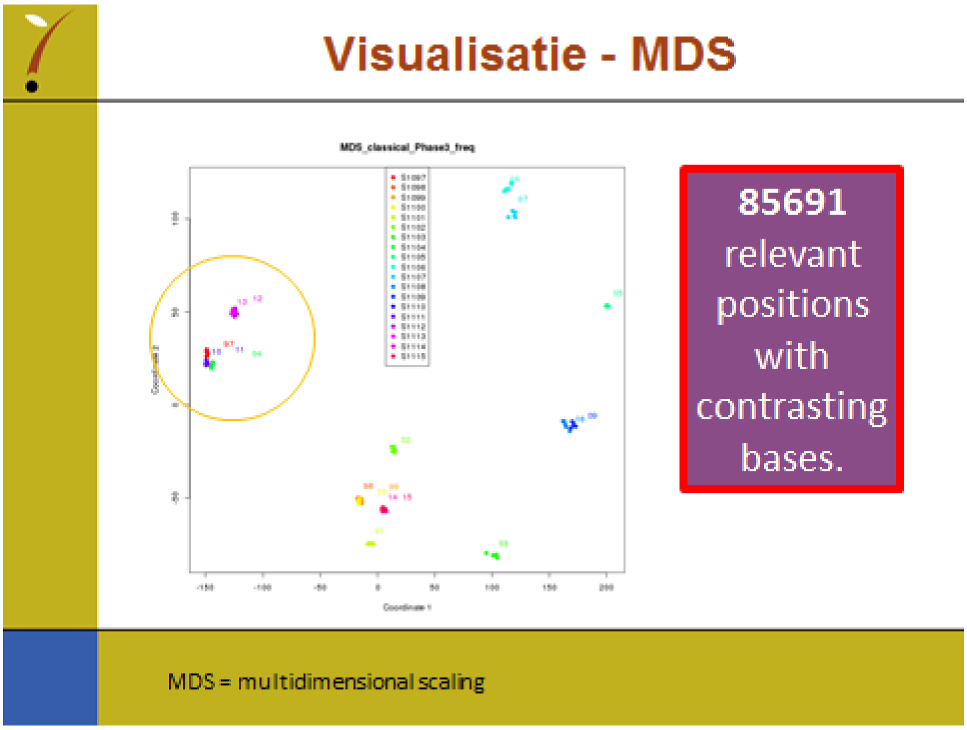


Visualisation - dendrogram

Two lots of same variety

Two maintenance/reference samples of same variety

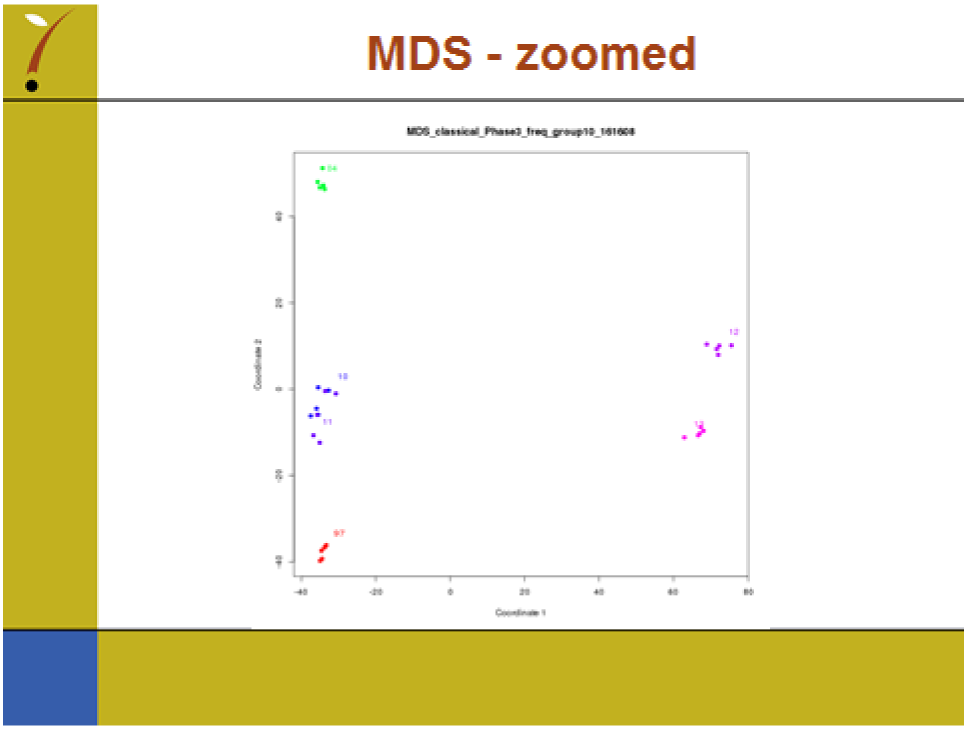
Same type different resistance
Not distinct with old techniques



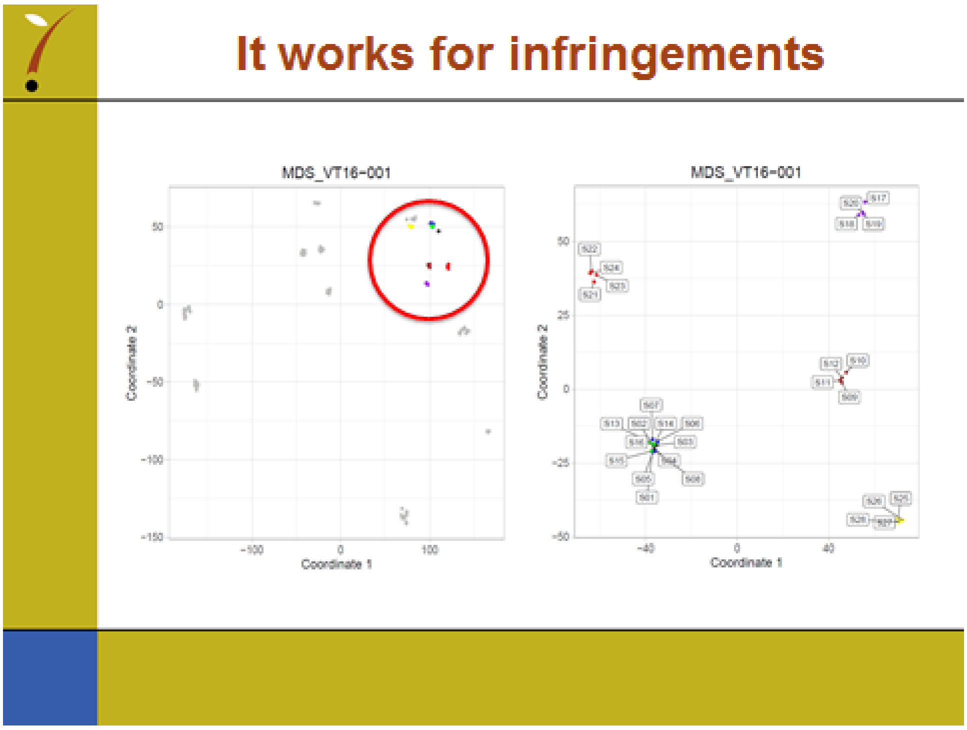
Visualisatie - MDS

85691 relevant positions with contrasting bases.

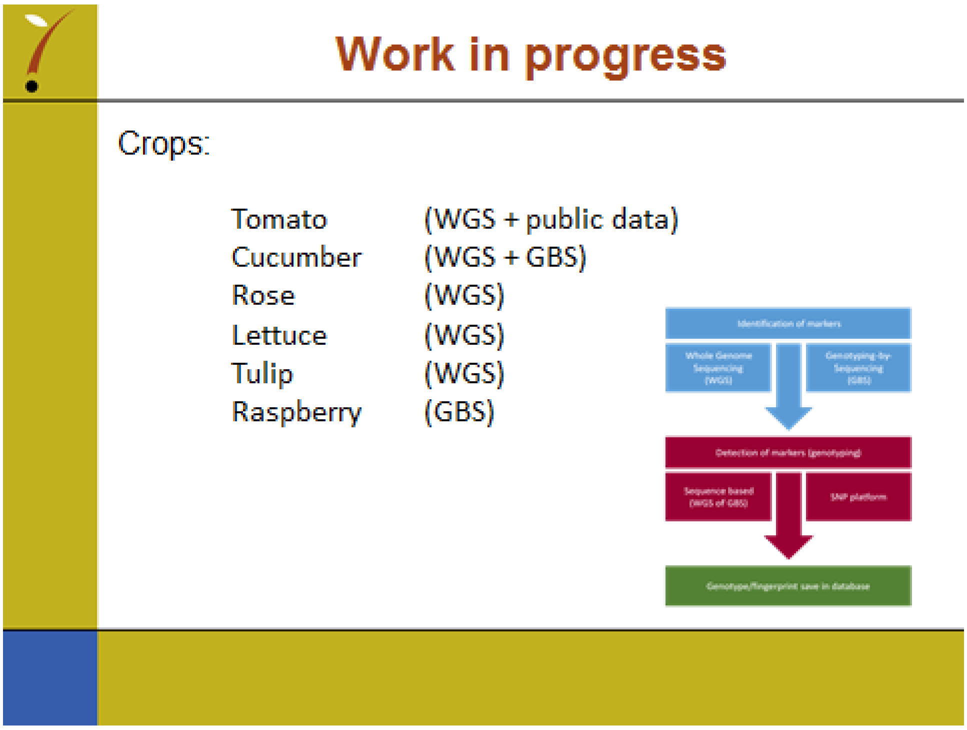
MDS = multidimensional scaling



MDS - zoomed



It works for infringements



Work in progress

Crops:

Tomato	(WGS + public data)
Cucumber	(WGS + GBS)
Rose	(WGS)
Lettuce	(WGS)
Tulip	(WGS)
Raspberry	(GBS)

Identification of markers

Whole Genome Sequencing (WGS)

Genotyping-by-Sequencing (GBS)

Detection of markers (genotyping)

Sequence based (WGS of GBS)

SNP platform

Genotype/fingerprint save in database

Quality in Horticulture

[End of Annex and of document]