

**BMT/15/23****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**



GREEN FORENSICS: WHOLE GENOME SEQUENCING APPROACH FOR PBR ENFORCEMENT

*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 "Green Forensics: Whole Genome Sequencing approach for PBR enforcement" 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]



---

## Green Forensics: *Whole Genome Sequencing approach for PBR enforcement*

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




## Fraud in horticulture

1. **Illegal propagation** of seeds or plants and commercialize under different name.


12,5% of all horticultural products is propagated illegally.

This percentage can even increase dependent on:

- manner of propagation
- location



2. **Repeated cropping**  
Misuse the name of a good and popular variety to sell plants or seeds of inferior quality using this name. (fake logos and packages).



## Search for fraud

### Crime Plant Investigation

RAI, AIB en Breeders' Trust  
Police C.I.D. or F.B.I. of plantworld



Still of Ted Darrah:  
Crime Scene Investigator



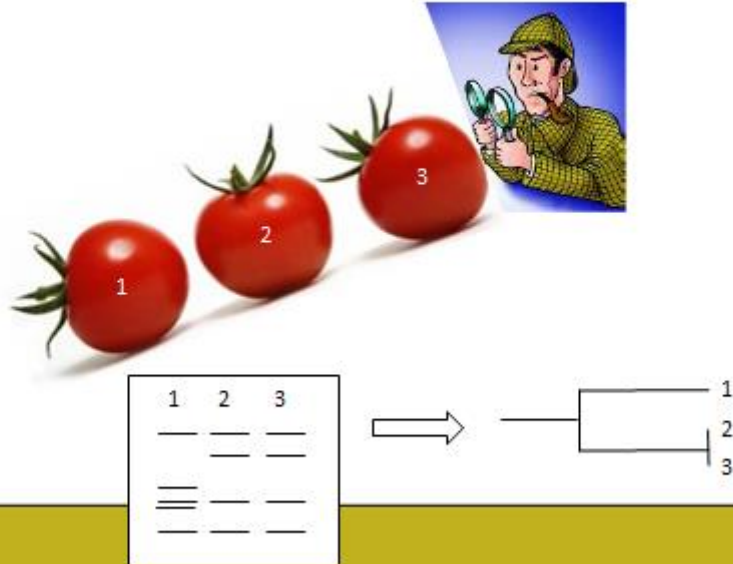
## Variety Tracer

### Genetic analyses for PBR enforcement

The 'Sherlock Holmes' concept  
in infringements



## Limitation of current technology



## Green Forensics



## From Variety Tracer 1.0 to 2.0

Replace current DNA fingerprint technology by new technology based on High Throughput Sequencing (HTS)

- more information = better resolution
- Higher discriminative power
- Co-dominant in stead of dominant
- Better representation of the total genome (total DNA)
- More reliable
- Generally applicable
- Cost effective

Variety Tracer 1.0



Variety Tracer 2.0



## Green Forensics

### Selection of crops

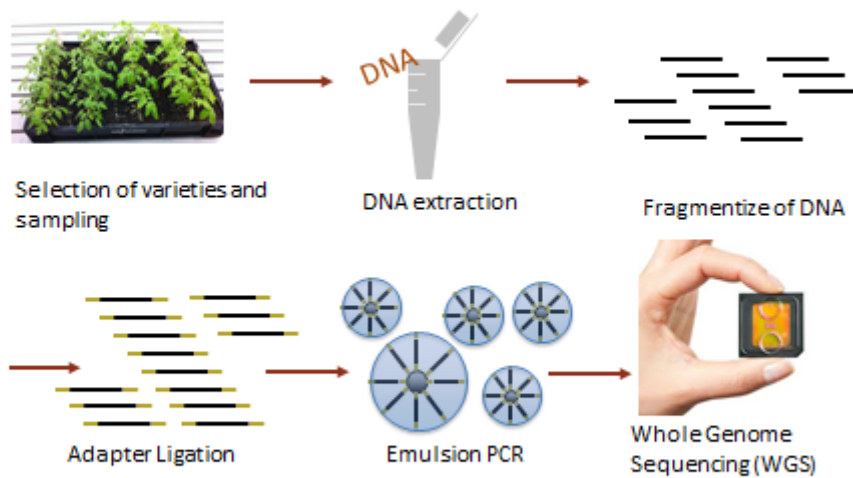
	Tomato	800 Mb	diploid	refG	proof of principle
	Cucumber	200 Mb	diploid	refG	(=Chinese long)
	Rose	560 MB	tetraploid	no refG	
	Lettuce	2.7 Gb	diploid	ref-Assembly	
	Tulip	>25 Gb	diploid	no refG	



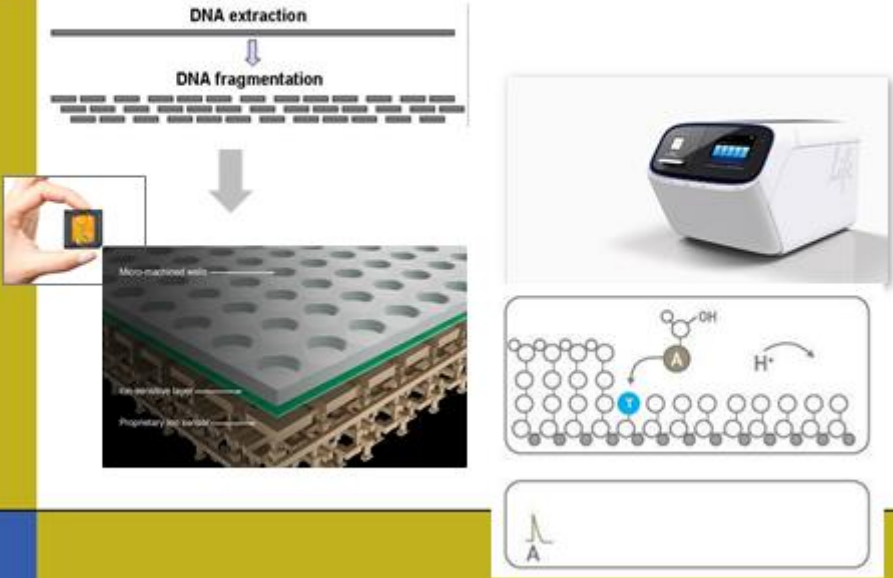
## Approach - step by step

- **Selection of varieties, growing and DNA extraction**
- **Sequencing**
- Mapping against Reference genome
- Determine Informative positions
- Construction of database
  - 19 accessions/varieties; 5 replica's (training set)
- Visualization – built visualization tool MDS
- Testing and validation
  - Different types; genetically closely related varieties; different Lots of same variety; mutants; seedlings...e.g.

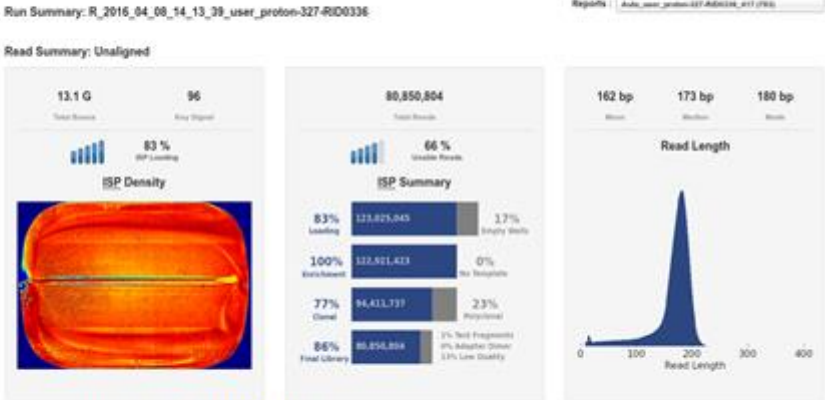
## “Wetlab” Pipeline



# Sequencing: Ion Torrent

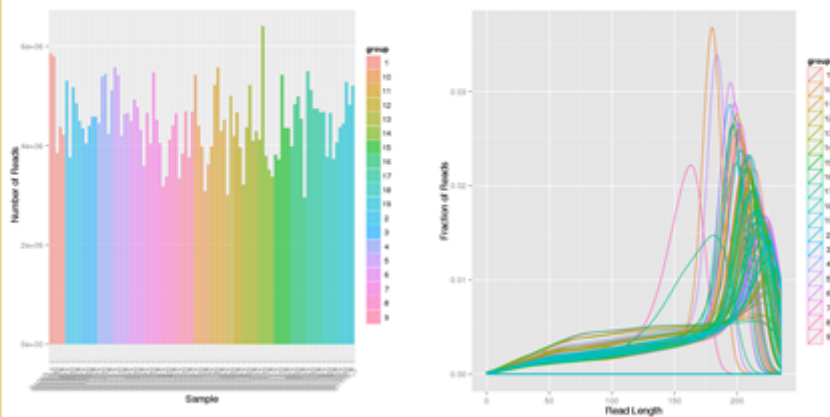


# Ion Torrent - output



## Data QC and trimming

Check sequence data for artefacts



## Approach - step by step

- Selection of varieties, growing and DNA extraction
- Sequencing
- **Mapping against Reference genome**
- Determine Informative positions
- Construction of database
  - 19 accessions/varieties; 5 replica's (training set)
- Visualization – built visualization tool MDS
- Testing and validation
  - Different types; genetically closely related varieties; different Lots of same variety; mutants; seedlings...e.g.





## Organizing the sequence reads



Sequence data analysis is like: 'do puzzles'



## With or without Example on the cover

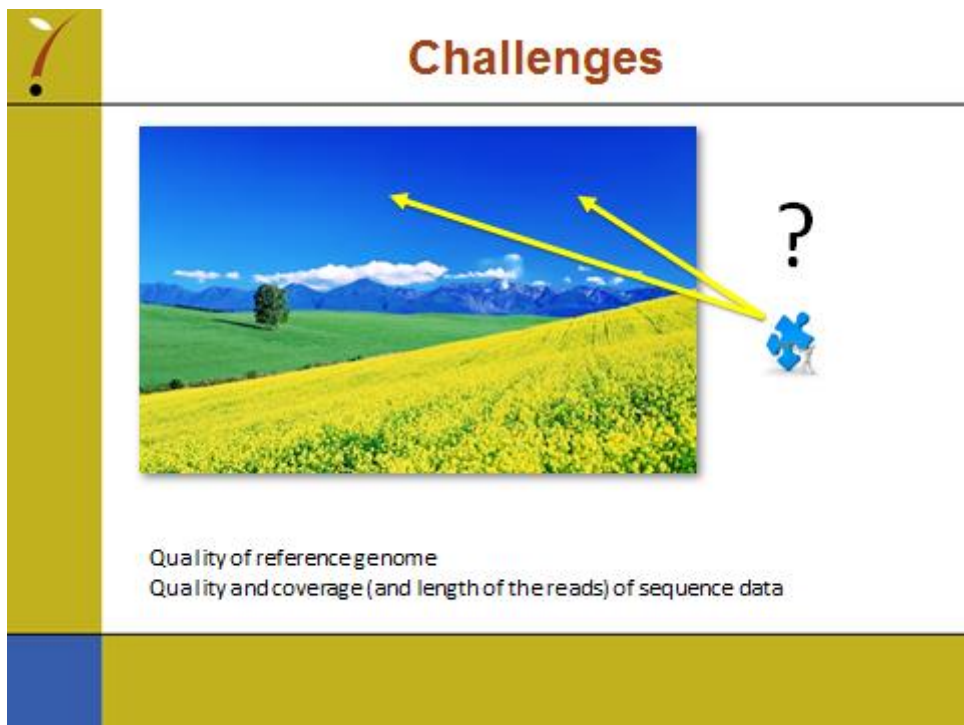
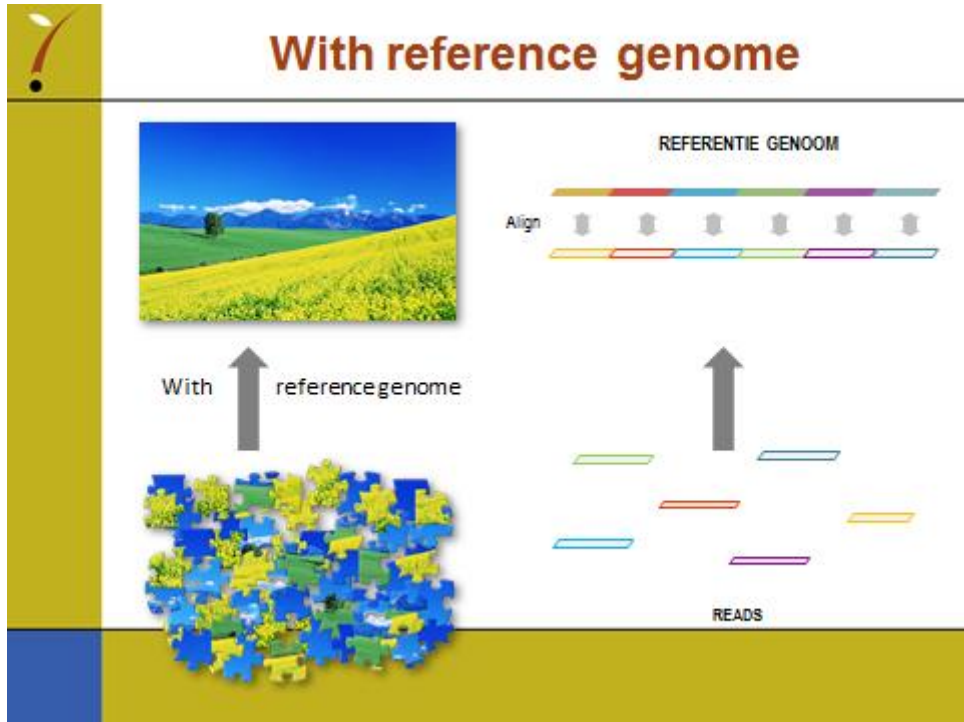


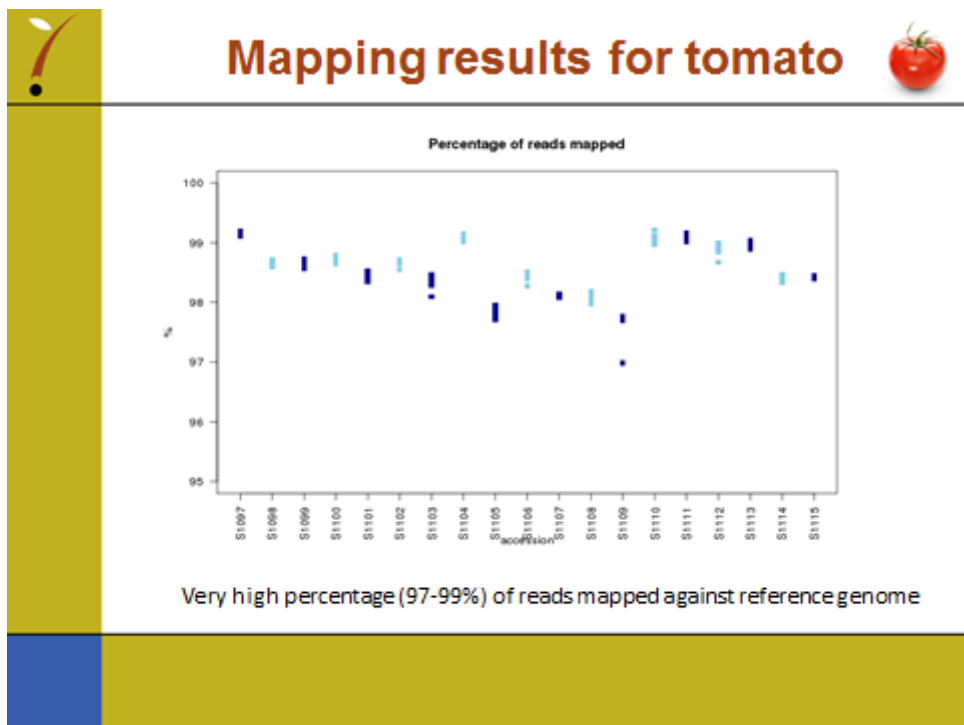
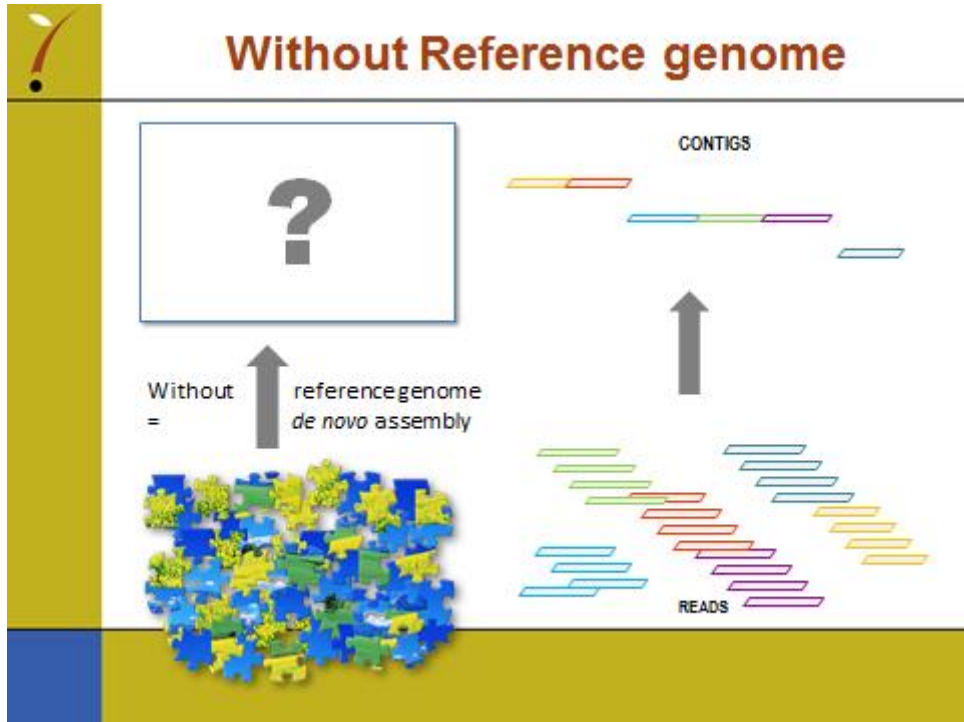
With ↑ reference genome



Without = ↑ reference genome  
*de novo* assembly



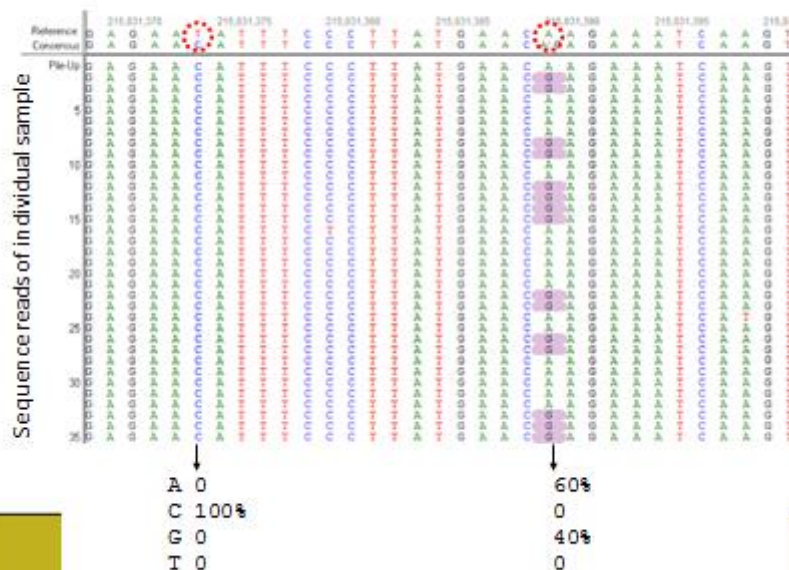




## Approach - step by step

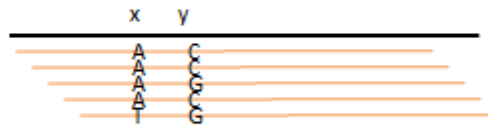
- Selection of varieties, growing and DNA extraction
- Sequencing
- Mapping against Reference genome
- **Determine Informative positions**
- Construction of database
  - 19 accessions/varieties; 5 replica's (training set)
- Visualization – built visualization tool MDS
- Testing and validation
  - Different types; genetically closely related varieties; different Lots of same variety; mutants; seedlings...e.g.

## Alignments



## Base position analysis

- At each position the read coverage is determined and the occurrence of each nucleotide as a fraction of read depth (fraction  $N$ ; frequency) is determined.



pos	A	C	G	T
-				
x	0.8	0	0	0.2
-				
y	0	0.6	0.4	0

## Base position analysis

- There are 3 possible genotypes per position (in diploid).
- A genotype is called only if **all** samples (replica's) of a variety have the same genotype!
- A 'contrast' is defined as a combination of opposing genotypes.
- Base positions with contrasting genotypes between two varieties are selected.

Fraction $N$	Genotype	Description	Genotype 1	Genotype 2
0 - 0.2	-	Low abundance	+	-
0.4 - 0.6	H	Heterozygous	+	H
0.8 - 1	+	High abundance	-	+
other	$N$	Not determined	-	H
			H	+
			H	-

## Tomato: 171 pairwise combinations

Number of positions with contrasting genotypes between variety pairs

	Sl097	Sl098	Sl099	Sl100	Sl101	Sl102	Sl103	Sl104	Sl105	Sl106	Sl107	Sl108	Sl109	Sl110	Sl111	Sl112	Sl113	Sl114	Sl115
Sl097	0	542	510	704	1293	734	6090	370	5579	3589	6109	4727	6198	241	111	343	314	742	866
Sl098		0	9	64	303	218	188	347	2201	1480	2175	369	1946	366	355	257	374	105	155
Sl099			0	73	423	198	171	294	1653	1143	1651	452	1303	288	278	207	353	95	64
Sl100				0	859	216	161	446	2611	1744	2794	649	2246	466	384	400	482	185	162
Sl101					0	597	859	809	3713	3002	3016	713	889	949	800	877	1341	431	364
Sl102						0	59	405	1460	1276	1669	337	1348	300	423	323	323	205	257
Sl103							0	3321	1480	1295	1693	218	1087	3649	4024	3679	3647	235	261
Sl104								0	4867	1699	3297	2366	3320	139	119	234	355	406	347
Sl105									0	285	833	116	169	3679	6088	4541	7300	1418	1977
Sl106										0	873	188	899	2340	2365	1386	2099	1070	1557
Sl107											0	575	702	4014	4176	1787	2809	1566	2227
Sl108												0	51	3196	3472	3326	4606	404	356
Sl109													0	3926	4213	4075	3564	313	668
Sl110														0	7	216	180	390	348
Sl111															0	183	134	375	443
Sl112																0	16	335	374
Sl113																	0	322	306
Sl114																		0	8
Sl115																			0

In total 85,691 positions with contrasting genotypes between varieties were selected including 2,351 heterozygous base positions (1.5%)

## genetic distance

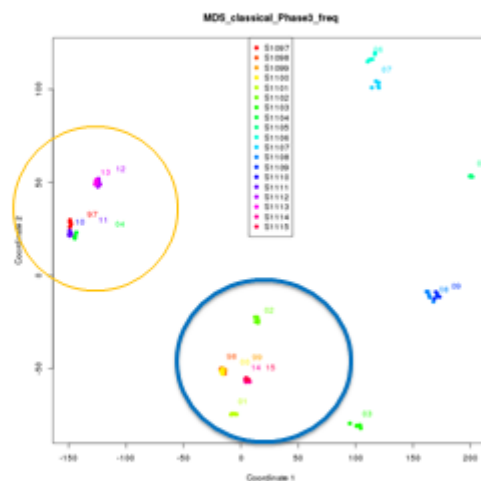
- Genetic distance is calculated based on the original frequencies (fraction N) of the contrasting base pairs on 85691 positions
- >160.000 contrasting base pairs in this dataset



## Approach - step by step

- Selection of varieties, growing and DNA extraction
- Sequencing
- Mapping against Reference genome
- Determine Informative positions
- **Construction of database**
  - 19 accessions/varieties; 5 replica's (training set)
- **Visualization – built visualization tool: MDS**
- Testing and validation
  - Different types; genetically closely related varieties; different Lots of same variety; mutants; seedlings...e.g.

## Visualization - MDS

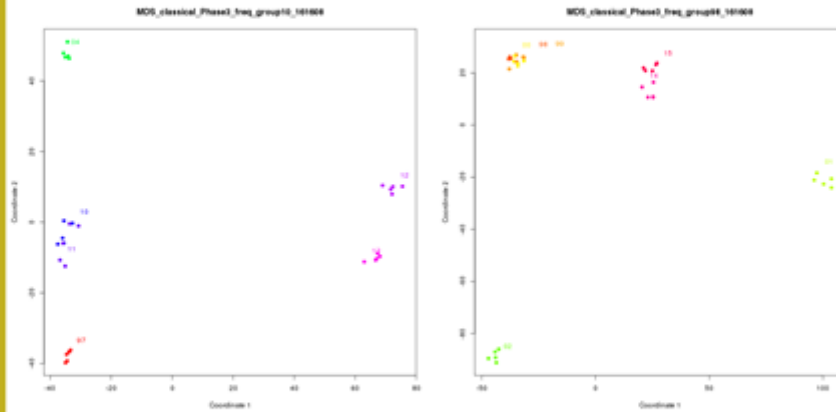


85691  
relevant  
positions

>160000  
contrasting  
bases.

MDS = multidimensional scaling

## Re-analyzing the clusters



## Approach - step by step

- Selection of varieties, growing and DNA extraction
- Sequencing
- Mapping against Reference genome
- Determine Informative positions
- Construction of database
  - 19 accessions/varieties; 5 replica's (training set)
- Visualization – built visualization tool: MDS
- **Testing and validation**
  - **Different types; genetically closely related varieties; different Lots of same variety; mutants; seedlings...e.g.**

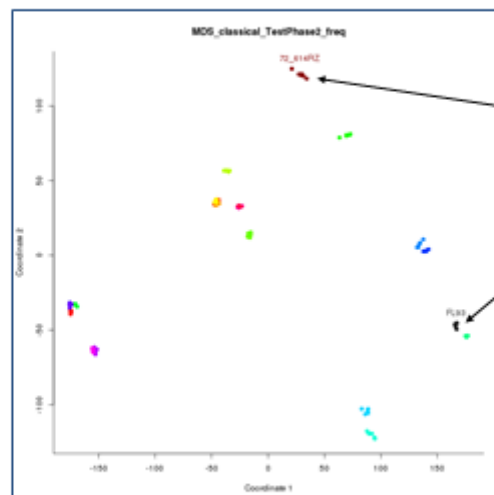


## Testing

1. Cherry and Fresh market tomato (different types) and several maintenance samples.
2. Tomato growing contest (Harrowgate)
3. Real infringement case

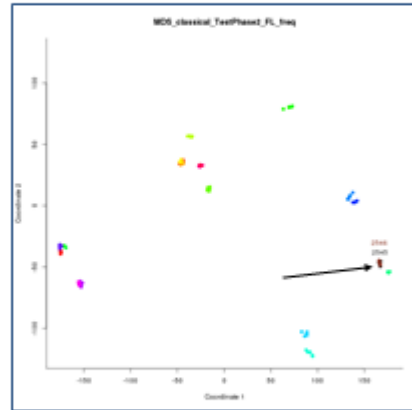
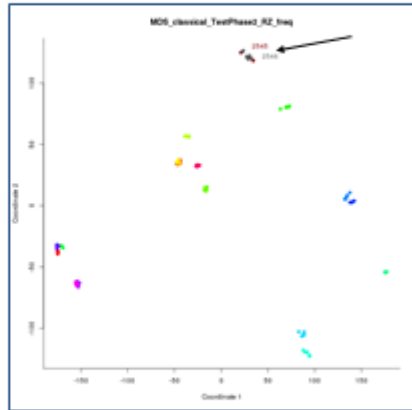


## Cherry and Fresh market tomato





## Cherry and Fresh market tomato



## Tomato growing contest



 **NEWS** Sign in

## Harrogate

### Now even veg shows are testing for cheats: Tomatoes in contest will have DNA examined to ensure they are correct variety

- Harrogate Autumn Flower Show is using DNA technology to ensure a £1,000 giant tomato prize is not won by a cheating grower
- The show is running a Gigantomo class, with a £1,000 top prize at stake
- Organiser will use Dutch specialists to DNA test plants to make sure the entries are what they claim to be
- Winner could scoop a further £5k if heaviest fruit sets a new world record

By DAILY MAIL REPORTER  
PUBLISHED: 00:58 GMT, 8 September 2015 | UPDATED: 03:33 GMT, 8 September 2015

 Share     

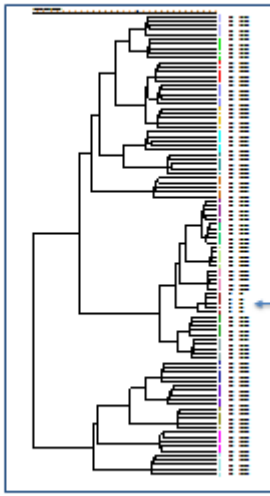
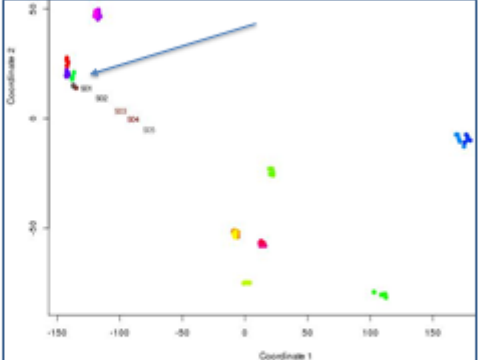
You would think a set of scales and an expert eye would be enough to judge entries in a giant vegetable contest.

But the competition has grown so fierce in one picture: DNA testing to weed out any imposters.

Adjudicators at the Harrogate Autumn Flower Show will now be using DNA testing to ensure that entries to their Gigantomo tomato class really are what they claim to be.



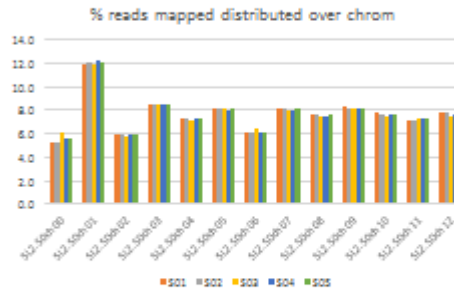
## Identity confirmed – no cheating



## Unexpected discovery

Low mapping percentage for one sample??

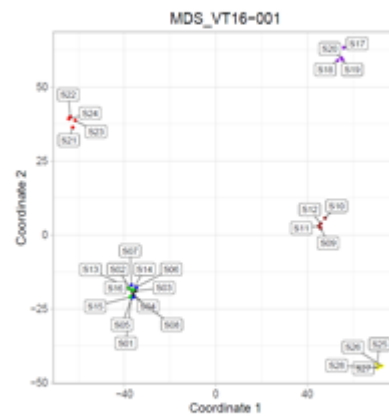
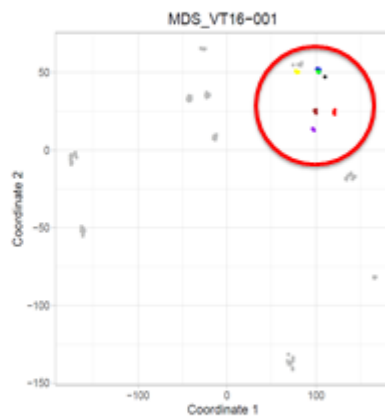
Sample	Nr reads	Nr reads mapped	% mapped
S01	14,891,765	14,891,276	99.7
S02	19,500,545	19,026,261	97.6
S03	16,092,446	6,221,944	38.7
S04	16,251,261	16,075,639	98.9
S05	17,655,012	17,558,599	99.4








DNA of *Pseudomonas fluorescens*: **dope??**



## It works for infringements



## Current status

	Tomato	<i>proof of principle</i>	<input checked="" type="checkbox"/>
	Cucumber	<i>ref-G (=Chinese Long)</i>	<input checked="" type="checkbox"/>
	Rose	<i>tetraploid no ref-G</i>	<input checked="" type="checkbox"/>
	Lettuce	<i>2.7 Gb diploid</i>	<input type="checkbox"/>
	Tulp	<i>&gt;25 Gb diploid no ref-G</i>	<input type="checkbox"/>

## Acknowledgements

### University of Amsterdam (MAD)

- Timo Breit
- Rick Orij
- Oskar Bruning-Gawehns
- **Inez Terpstra**
- Genevieve Gerard
- Wim de Leeuw
- Selina van Leeuwen
- Marina van Olst



### Naktuinbouw

- Michel Ebskamp
- **Fleur Gawehns**
- Menno Hoekstra
- Daniel Deinum
- Crop DUS specialists



# ***Quality in Horticulture***

[End of Annex and of document]