UPOV

International Union for the Protection of New Varieties of Plants

E

Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular

BMT/16/18 Rev.

Sixteenth Session
La Rochelle, France, November 7 to 10, 2017

Original: English
Date: October 17, 2017

ASSIGNMENT TESTS FOR GENOTYPE CLASSIFICATION

Document prepared by experts from the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO)

Disclaimer: this document does not represent UPOV policies or guidance

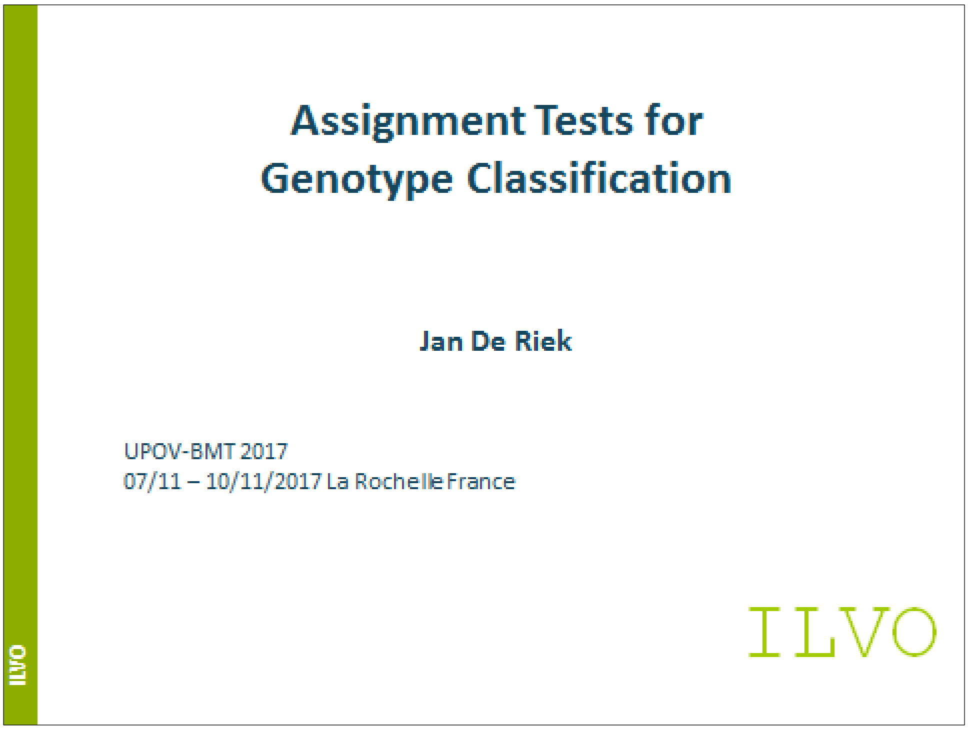
The Annex to this document contains a copy of a presentation "Assignment Tests for Genotype Classification" to be made at its sixteenth session of the Working Group on Biochemical and Molecular Techniques and DNA-Profiling in particular (BMT).

[Annex follows]

ANNEX

ASSIGNMENT TESTS FOR GENOTYPE CLASSIFICATION

Presentation prepared by experts from the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO)

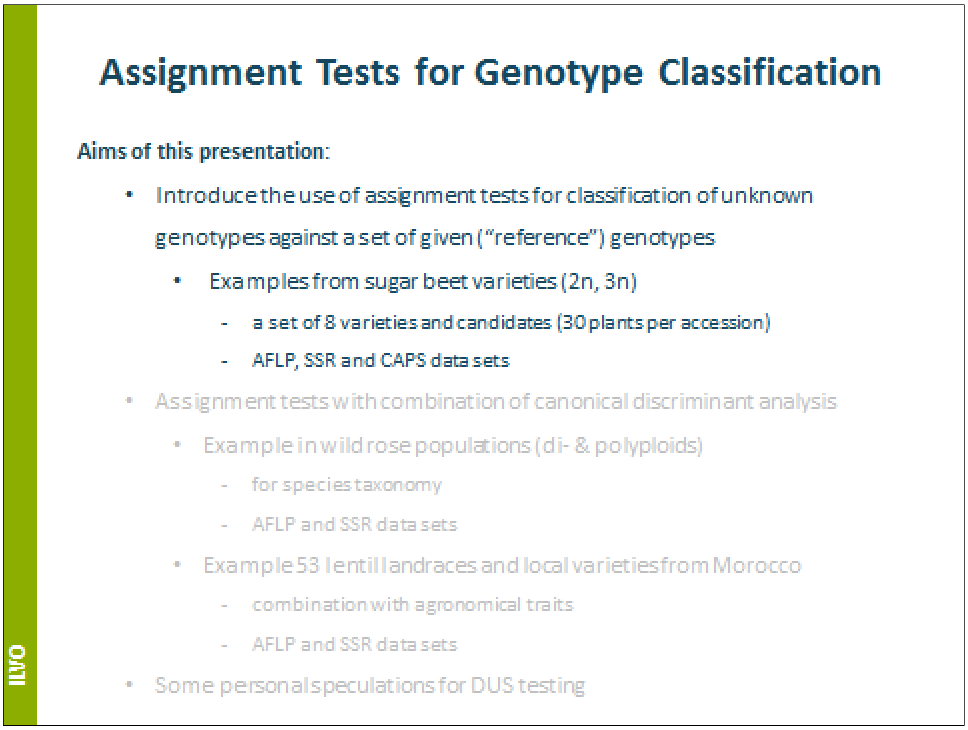


Assignment Tests for Genotype Classification

Jan De Riek

UPOV-BMT 2017
07/11 – 10/11/2017 La Rochelle France

ILVO



Assignment Tests for Genotype Classification

Aims of this presentation:

- Introduce the use of assignment tests for classification of unknown genotypes against a set of given ("reference") genotypes
  - Examples from sugar beet varieties (2n, 3n)
    - a set of 8 varieties and candidates (30 plants per accession)
    - AFLP, SSR and CAPS data sets
- Assignment tests with combination of canonical discriminant analysis
  - Example in wild rose populations (di- & polyploids)
    - for species taxonomy
    - AFLP and SSR data sets
  - Example 53 lentil landraces and local varieties from Morocco
    - combination with agronomical traits
    - AFLP and SSR data sets
- Some personal speculations for DUS testing

Variety Identification in Sugar Beet Varieties
Used Methods

High genetic variation within sugar beet varieties hampers reliable classification procedures independent of the type of marker technique applied. AFLPs, SSRs and CAPSs marker datasets in 8 sugar beet varieties were subjected to

(i) supervised classifiers:

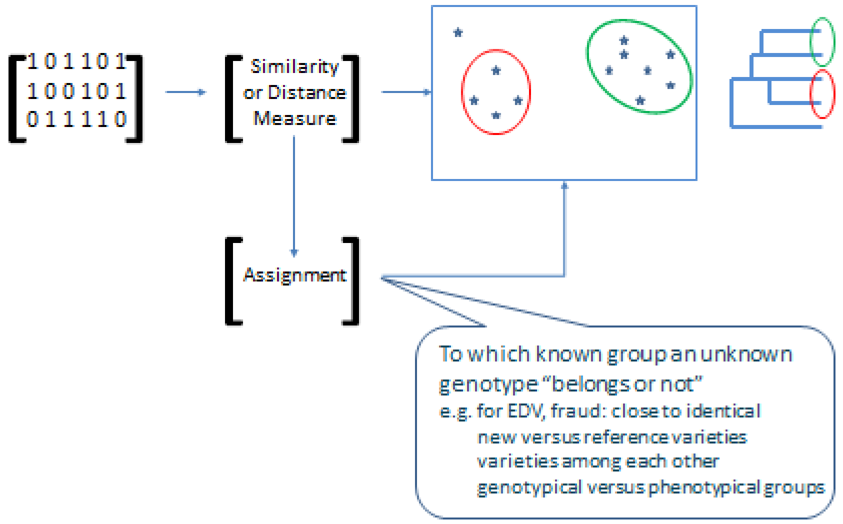
- methods in which individual assignments are made to predefined classes

(ii) unsupervised classifiers:

- defined afterwards on the similarity in marker composition from sampled individuals

(De Riek et al. 2007; Crop Science)

Non-supervised classification



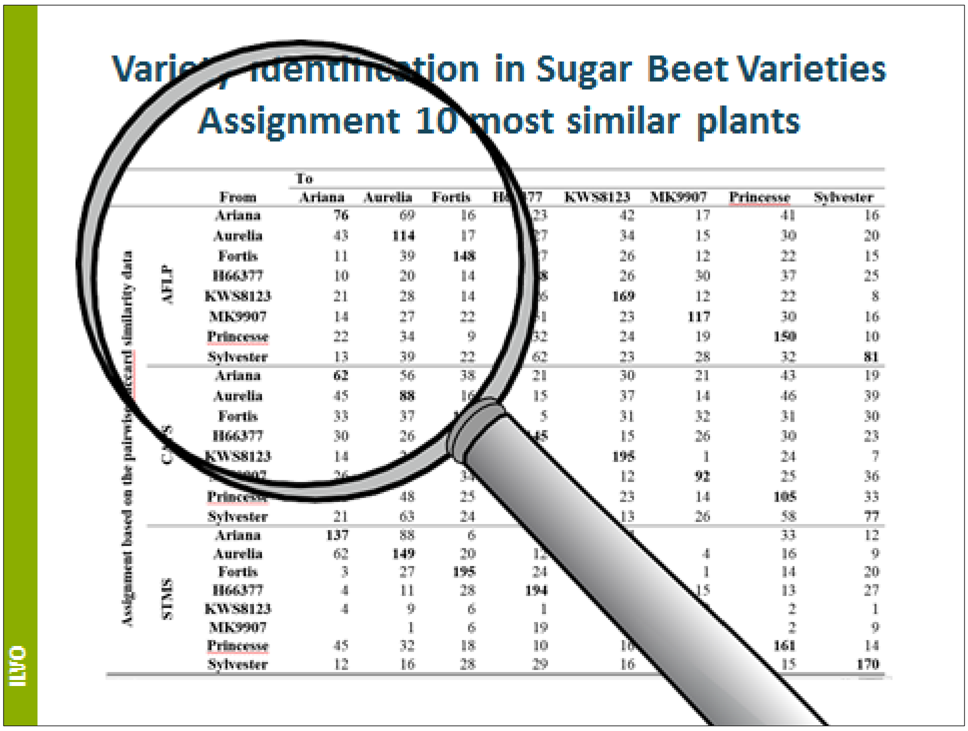
$$\begin{bmatrix} 1&0&1&1&0&1 \\ 1&0&0&1&0&1 \\ 0&1&1&1&1&0 \end{bmatrix}$$

[Similarity or Distance Measure]

[Assignment]

To which known group an unknown genotype "belongs or not"
e.g. for EDV, fraud: close to identical
- new versus reference varieties
- varieties among each other
- genotypical versus phenotypical groups

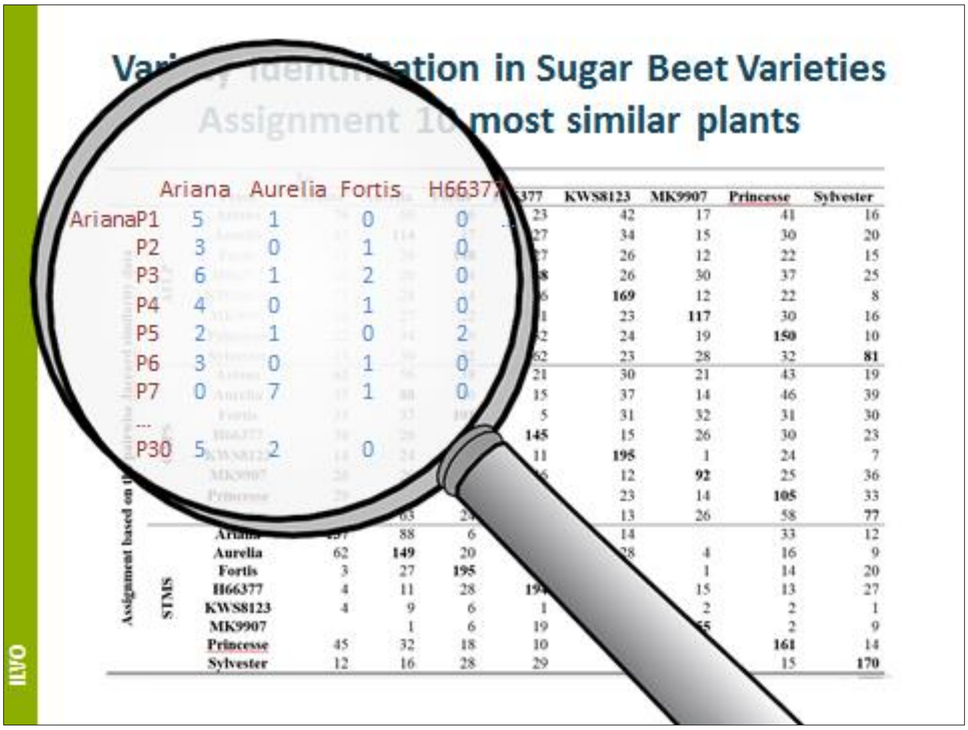
Supervised classification



Variety Identification in Sugar Beet Varieties
Assignment 10 most similar plants

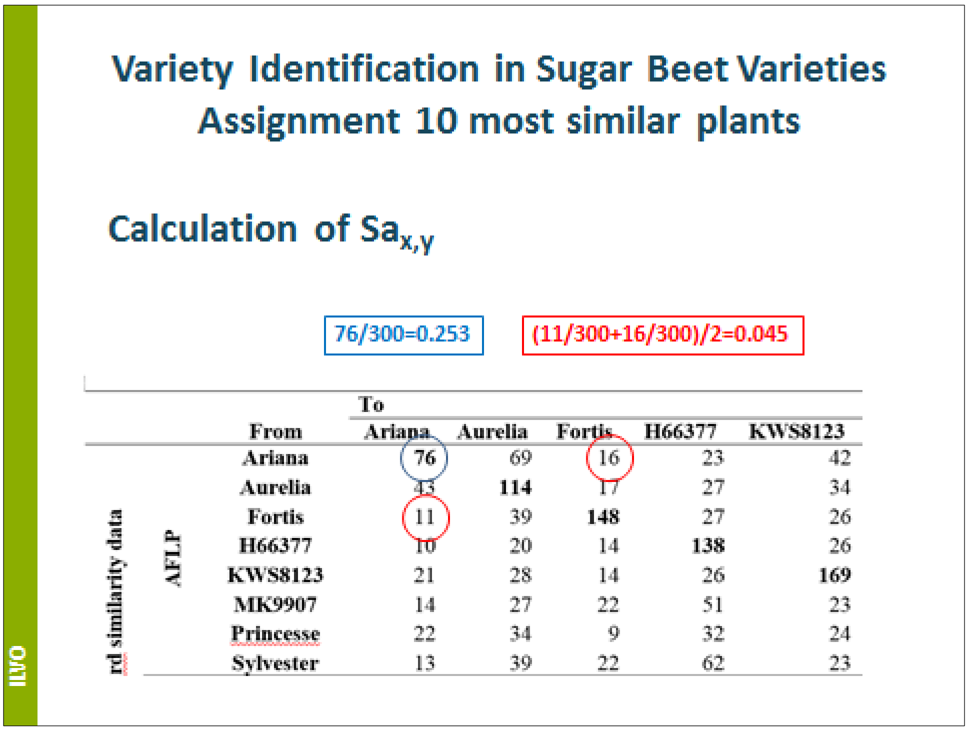
Assignment based on the pairwise Jaccard similarity data

	From	To Ariana	Aurelia	Fortis	H66377	KWS8123	MK9907	Princesse	Sylvester
AFLP	Ariana	**76**	69	16	23	42	17	41	16
	Aurelia	43	**114**	17	27	34	15	30	20
	Fortis	11	39	**148**	[illegible]	26	12	22	15
	H66377	10	20	14	[illegible]	26	30	37	25
	KWS8123	21	28	14	[illegible]	**169**	12	22	8
	MK9907	14	27	22	[illegible]	23	**117**	30	16
	Princesse	22	34	9	32	24	19	**150**	10
	Sylvester	13	39	22	62	23	28	32	**81**
CAPS	Ariana	**62**	56	38	21	30	21	43	19
	Aurelia	45	**88**	[illegible]	15	37	14	46	39
	Fortis	33	37	[illegible]	5	31	32	31	30
	H66377	30	26	[illegible]	[illegible]	15	26	30	23
	KWS8123	14	[illegible]	[illegible]	[illegible]	**195**	1	24	7
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	12	**92**	25	36
	Princesse	[illegible]	48	25	[illegible]	23	14	**105**	33
	Sylvester	21	63	24	[illegible]	13	26	58	**77**
STMS	Ariana	**137**	88	6	[illegible]	[illegible]	[illegible]	33	12
	Aurelia	62	**149**	20	[illegible]	[illegible]	4	16	9
	Fortis	3	27	**195**	24	[illegible]	1	14	20
	H66377	4	11	28	**194**	[illegible]	[illegible]	13	27
	KWS8123	4	9	6	1	[illegible]	[illegible]	2	1
	MK9907		1	6	19	[illegible]	[illegible]	2	9
	Princesse	45	32	18	10	[illegible]	[illegible]	**161**	14
	Sylvester	12	16	28	29	16	[illegible]	15	**170**



Variety Identification in Sugar Beet Varieties
Assignment 10 most similar plants

	Ariana	Aurelia	Fortis	H6637
ArianaP1	5	1	0	0
P2	3	0	1	0
P3	6	1	2	0
P4	4	0	1	0
P5	2	1	0	2
P6	3	0	1	0
P7	0	7	1	0
...				
P30	5	2	0	



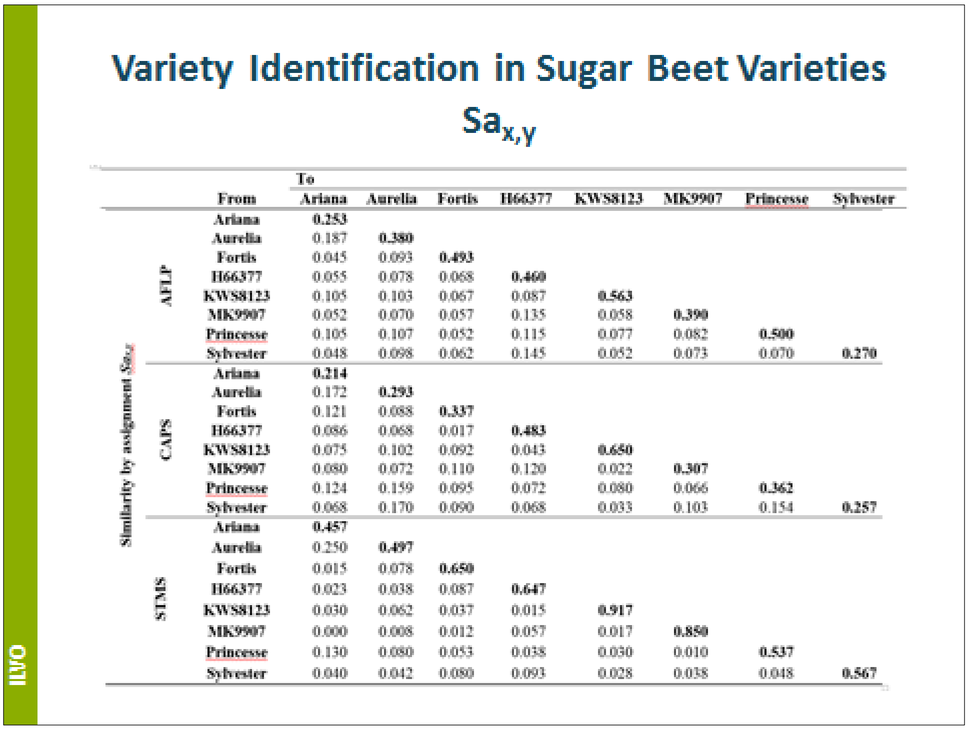
Variety Identification in Sugar Beet Varieties Assignment 10 most similar plants

Calculation of $Sa_{x,y}$

76/300=0.253

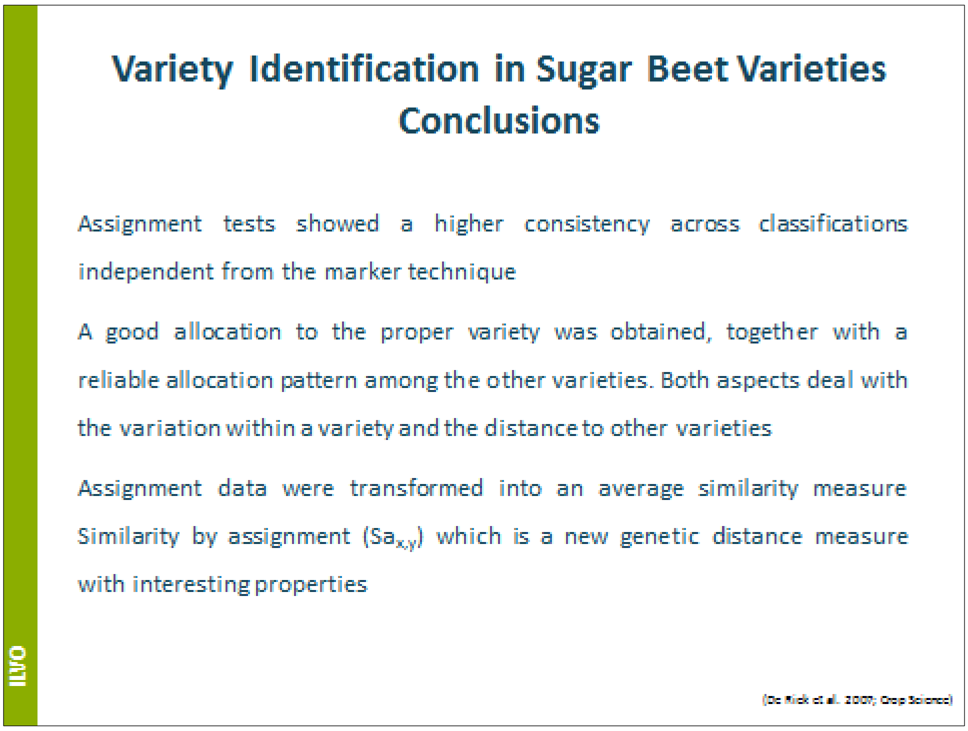
(11/300+16/300)/2=0.045

rd similarity data		To				
	From	**Ariana**	**Aurelia**	**Fortis**	**H66377**	**KWS8123**
AFLP	**Ariana**	**76**	69	16	23	42
	Aurelia	43	**114**	17	27	34
	Fortis	11	39	**148**	27	26
	H66377	10	20	14	**138**	26
	KWS8123	21	28	14	26	**169**
	MK9907	14	27	22	51	23
	Princesse	22	34	9	32	24
	Sylvester	13	39	22	62	23



Variety Identification in Sugar Beet Varieties $Sa_{x,y}$

Similarity by assignment $Sa_{x,y}$		To							
	From	**Ariana**	**Aurelia**	**Fortis**	**H66377**	**KWS8123**	**MK9907**	**Princesse**	**Sylvester**
AFLP	**Ariana**	**0.253**							
	Aurelia	0.187	**0.380**						
	Fortis	0.045	0.093	**0.493**					
	H66377	0.055	0.078	0.068	**0.460**				
	KWS8123	0.105	0.103	0.067	0.087	**0.563**			
	MK9907	0.052	0.070	0.057	0.135	0.058	**0.390**		
	Princesse	0.105	0.107	0.052	0.115	0.077	0.082	**0.500**	
	Sylvester	0.048	0.098	0.062	0.145	0.052	0.073	0.070	**0.270**
CAPS	**Ariana**	**0.214**							
	Aurelia	0.172	**0.293**						
	Fortis	0.121	0.088	**0.337**					
	H66377	0.086	0.068	0.017	**0.483**				
	KWS8123	0.075	0.102	0.092	0.043	**0.650**			
	MK9907	0.080	0.072	0.110	0.120	0.022	**0.307**		
	Princesse	0.124	0.159	0.095	0.072	0.080	0.066	**0.362**	
	Sylvester	0.068	0.170	0.090	0.068	0.033	0.103	0.154	**0.257**
STMS	**Ariana**	**0.457**							
	Aurelia	0.250	**0.497**						
	Fortis	0.015	0.078	**0.650**					
	H66377	0.023	0.038	0.087	**0.647**				
	KWS8123	0.030	0.062	0.037	0.015	**0.917**			
	MK9907	0.000	0.008	0.012	0.057	0.017	**0.850**		
	Princesse	0.130	0.080	0.053	0.038	0.030	0.010	**0.537**	
	Sylvester	0.040	0.042	0.080	0.093	0.028	0.038	0.048	**0.567**



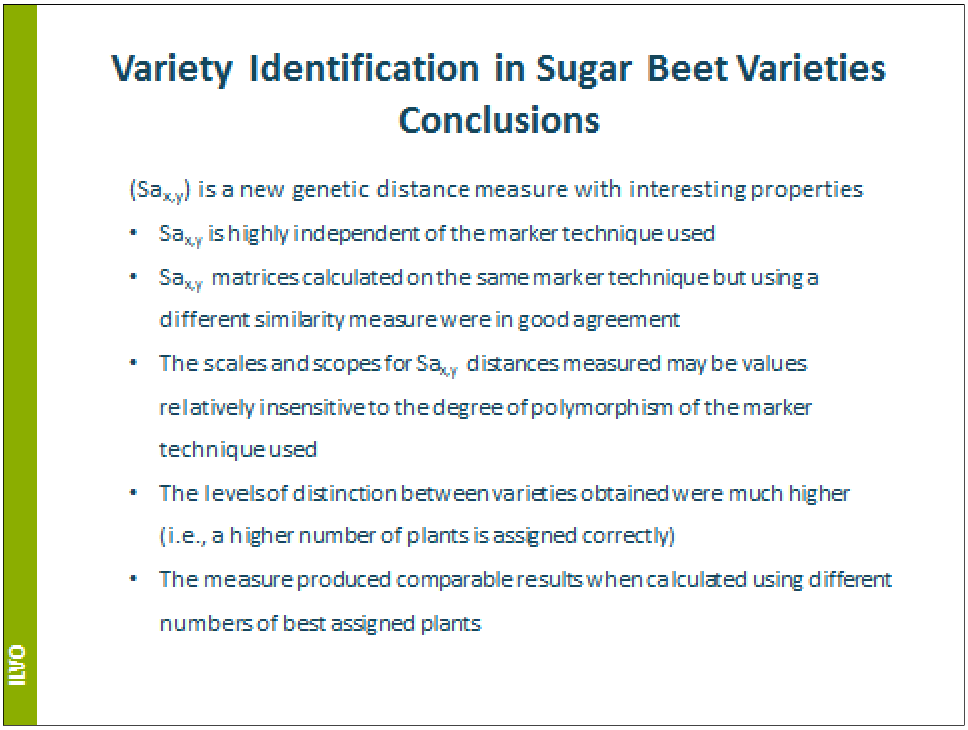
Variety Identification in Sugar Beet Varieties Conclusions

Assignment tests showed a higher consistency across classifications independent from the marker technique

A good allocation to the proper variety was obtained, together with a reliable allocation pattern among the other varieties. Both aspects deal with the variation within a variety and the distance to other varieties

Assignment data were transformed into an average similarity measure Similarity by assignment ($Sa_{x,y}$) which is a new genetic distance measure with interesting properties

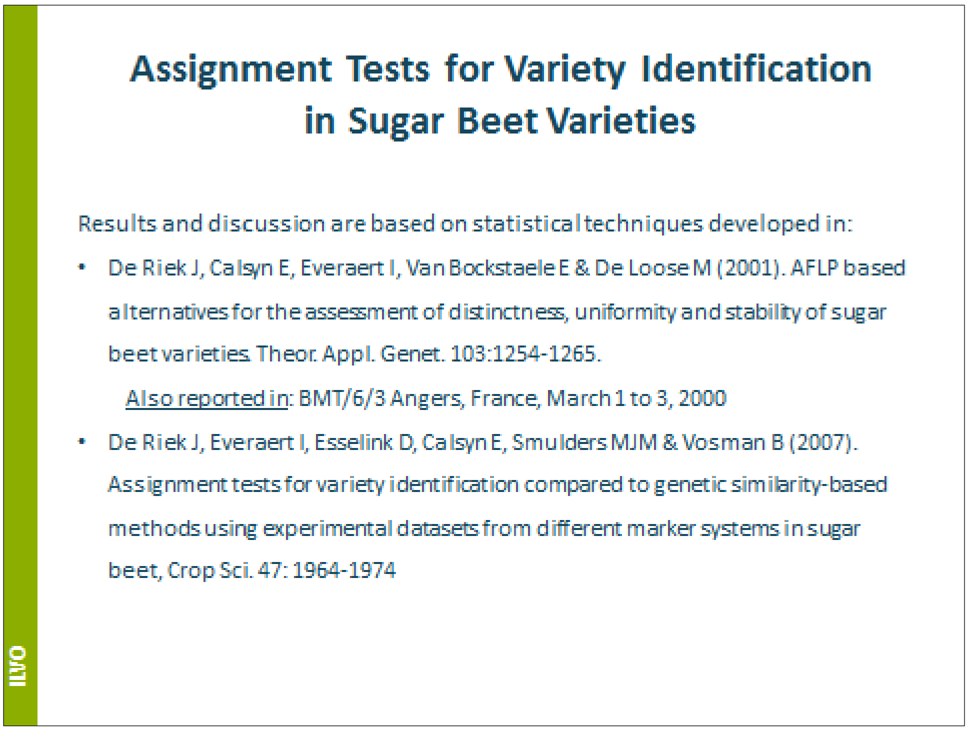
(De Riek et al. 2007; Crop Science)



Variety Identification in Sugar Beet Varieties Conclusions

($Sa_{x,y}$) is a new genetic distance measure with interesting properties

- $Sa_{x,y}$ is highly independent of the marker technique used
- $Sa_{x,y}$ matrices calculated on the same marker technique but using a different similarity measure were in good agreement
- The scales and scopes for $Sa_{x,y}$ distances measured may be values relatively insensitive to the degree of polymorphism of the marker technique used
- The levels of distinction between varieties obtained were much higher (i.e., a higher number of plants is assigned correctly)
- The measure produced comparable results when calculated using different numbers of best assigned plants

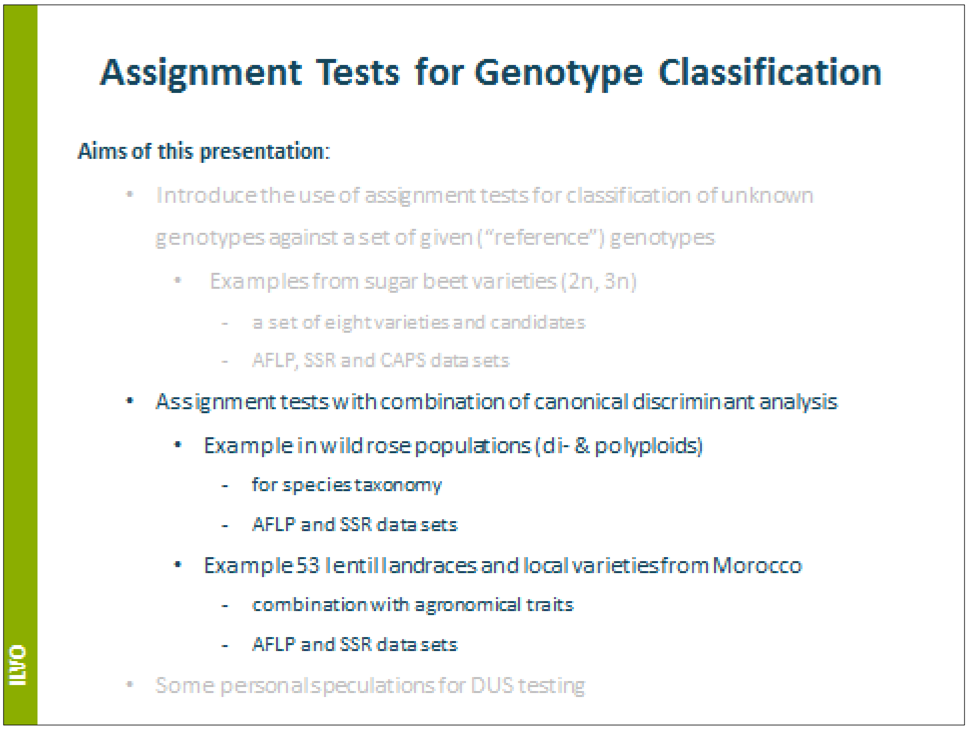


Assignment Tests for Variety Identification in Sugar Beet Varieties

Results and discussion are based on statistical techniques developed in:

- De Riek J, Calsyn E, Everaert I, Van Bockstaele E & De Loose M (2001). AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet. 103:1254-1265.

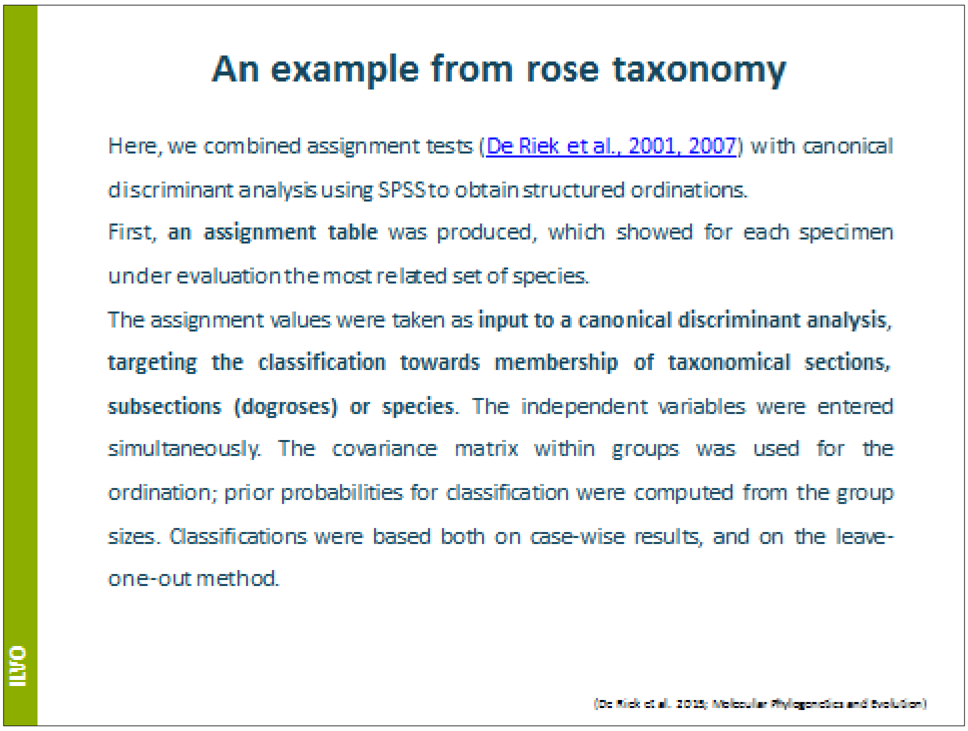
  Also reported in: BMT/6/3 Angers, France, March 1 to 3, 2000
- De Riek J, Everaert I, Esselink D, Calsyn E, Smulders MJM & Vosman B (2007). Assignment tests for variety identification compared to genetic similarity-based methods using experimental datasets from different marker systems in sugar beet, Crop Sci. 47: 1964-1974



Assignment Tests for Genotype Classification

Aims of this presentation:

- Introduce the use of assignment tests for classification of unknown genotypes against a set of given (“reference”) genotypes
  - Examples from sugar beet varieties (2n, 3n)
    - a set of eight varieties and candidates
    - AFLP, SSR and CAPS data sets
- Assignment tests with combination of canonical discriminant analysis
  - Example in wild rose populations (di- & polyploids)
    - for species taxonomy
    - AFLP and SSR data sets
  - Example 53 lentil landraces and local varieties from Morocco
    - combination with agronomical traits
    - AFLP and SSR data sets
- Some personal speculations for DUS testing



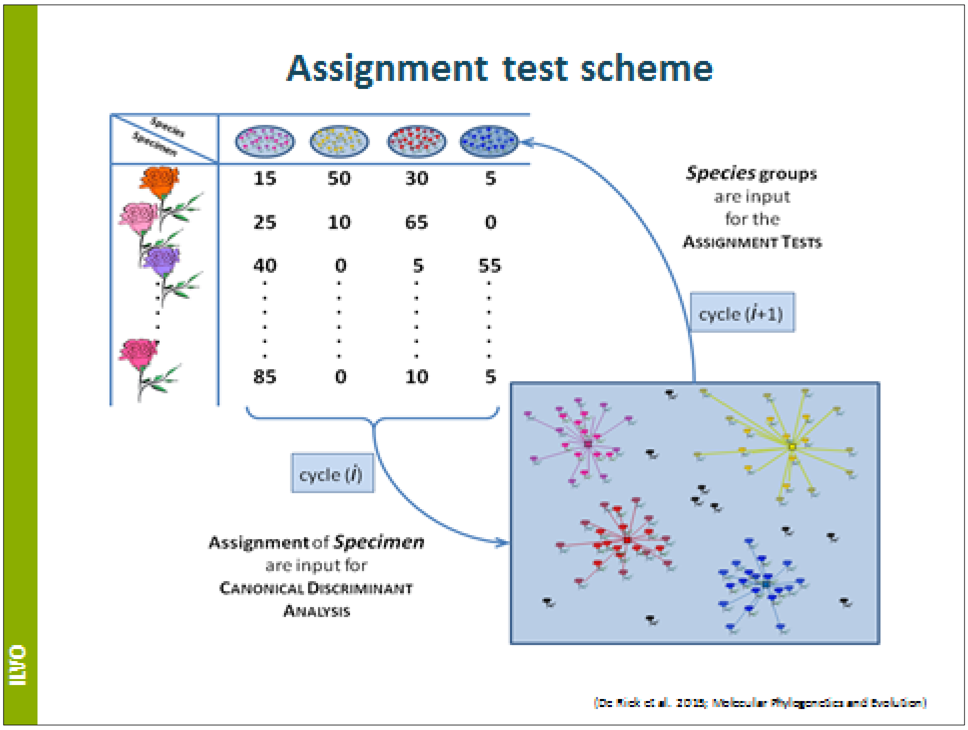
An example from rose taxonomy

Here, we combined assignment tests (De Riek et al., 2001, 2007) with canonical discriminant analysis using SPSS to obtain structured ordinations.

First, **an assignment table** was produced, which showed for each specimen under evaluation the most related set of species.

The assignment values were taken as **input to a canonical discriminant analysis, targeting the classification towards membership of taxonomical sections, subsections (dogroses) or species**. The independent variables were entered simultaneously. The covariance matrix within groups was used for the ordination; prior probabilities for classification were computed from the group sizes. Classifications were based both on case-wise results, and on the leave-one-out method.

(De Riek et al. 2013; Molecular Phylogenetics and Evolution)



Assignment test scheme

Species / Specimen				
	15	50	30	5
	25	10	65	0
	40	0	5	55
⋮	⋮	⋮	⋮	⋮
	85	0	10	5

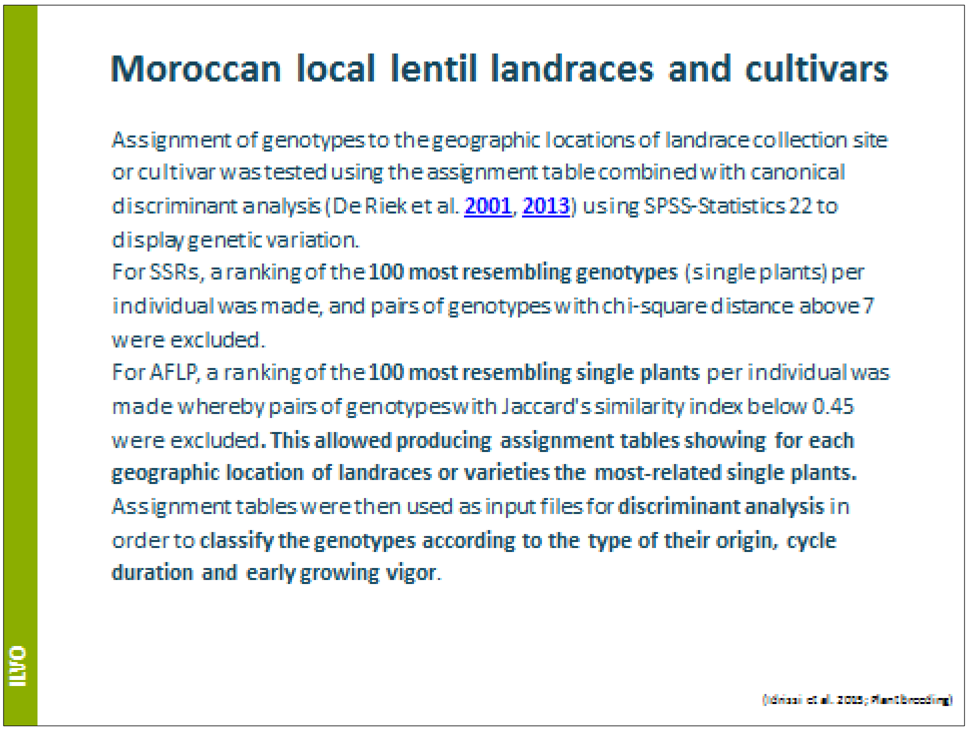
***Species* groups**
are input
for the
ASSIGNMENT TESTS

cycle (*i*+1)

cycle (*i*)

Assignment of *Specimen*
are input for
CANONICAL DISCRIMINANT ANALYSIS

(De Riek et al. 2013; Molecular Phylogenetics and Evolution)



Moroccan local lentil landraces and cultivars

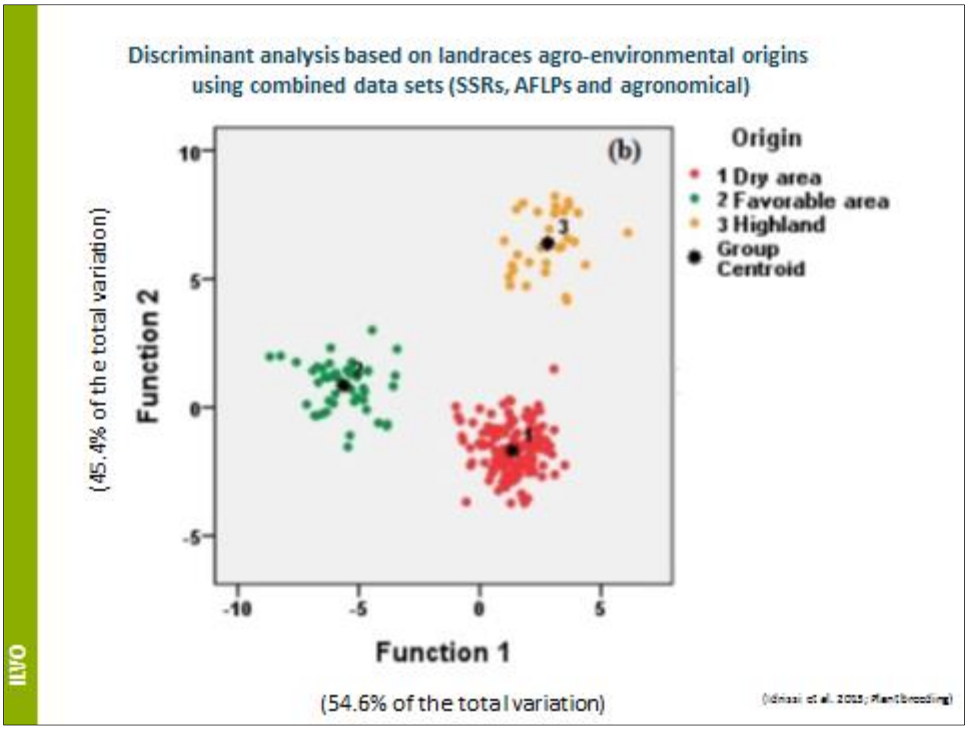
Assignment of genotypes to the geographic locations of landrace collection site or cultivar was tested using the assignment table combined with canonical discriminant analysis (De Riek et al. 2001, 2013) using SPSS-Statistics 22 to display genetic variation.

For SSRs, a ranking of the **100 most resembling genotypes** (single plants) per individual was made, and pairs of genotypes with chi-square distance above 7 were excluded.

For AFLP, a ranking of the **100 most resembling single plants** per individual was made whereby pairs of genotypes with Jaccard's similarity index below 0.45 were excluded. **This allowed producing assignment tables showing for each geographic location of landraces or varieties the most-related single plants.**

Assignment tables were then used as input files for **discriminant analysis** in order to **classify the genotypes according to the type of their origin, cycle duration and early growing vigor**.

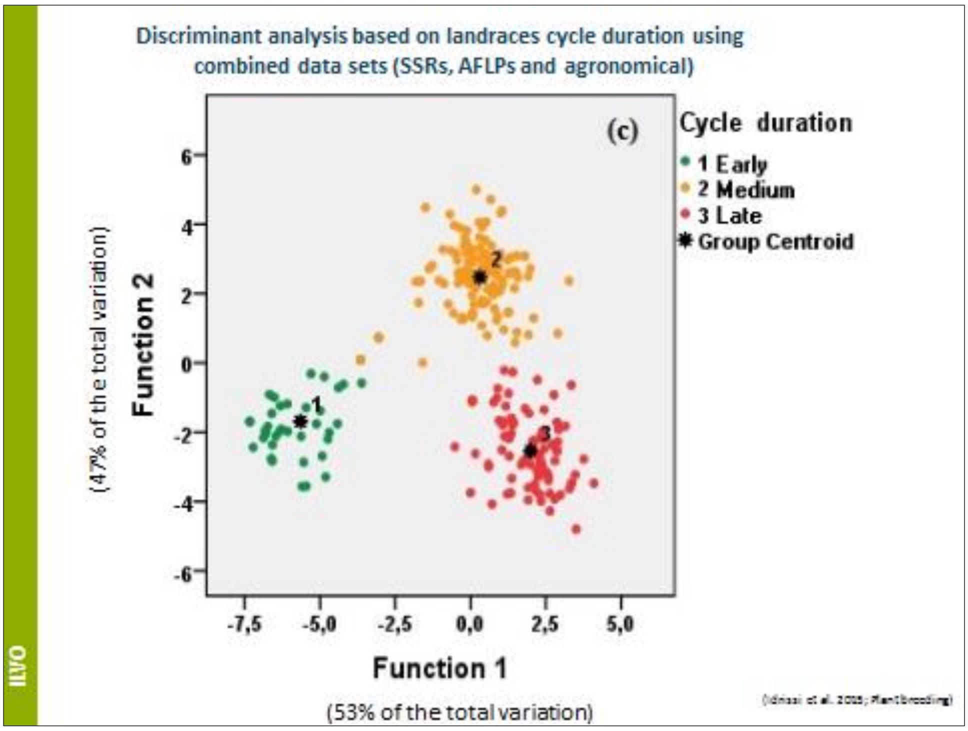
(Idrissi et al. 2015; Plant Breeding)



Discriminant analysis based on landraces agro-environmental origins using combined data sets (SSRs, AFLPs and agronomical)

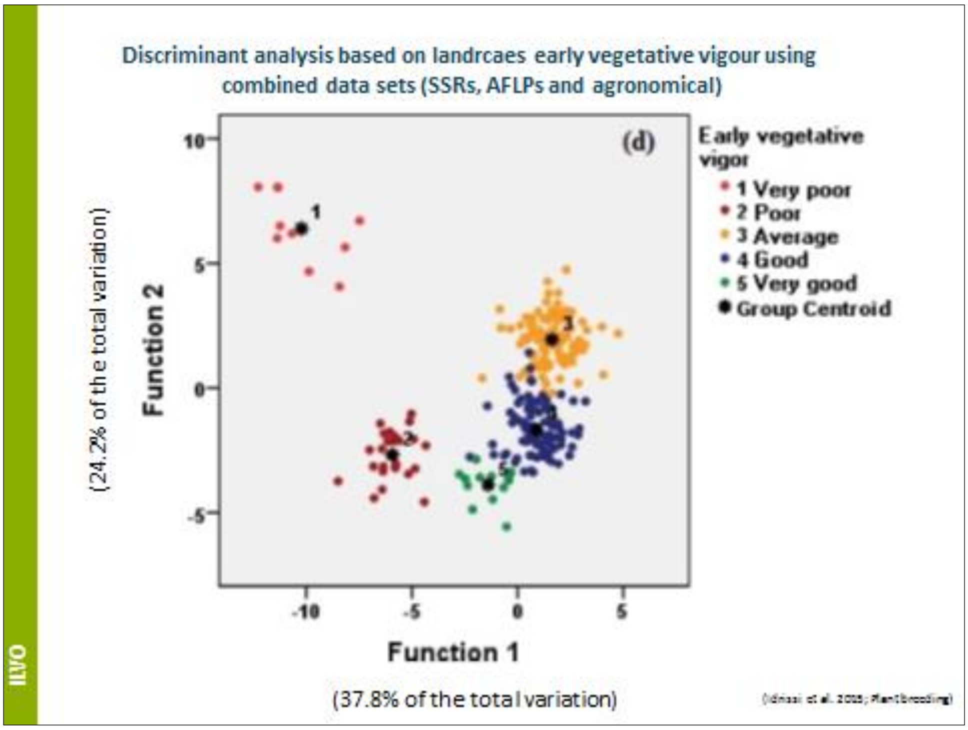
(54.6% of the total variation)

(Idrissi et al. 2015; Plant Breeding)



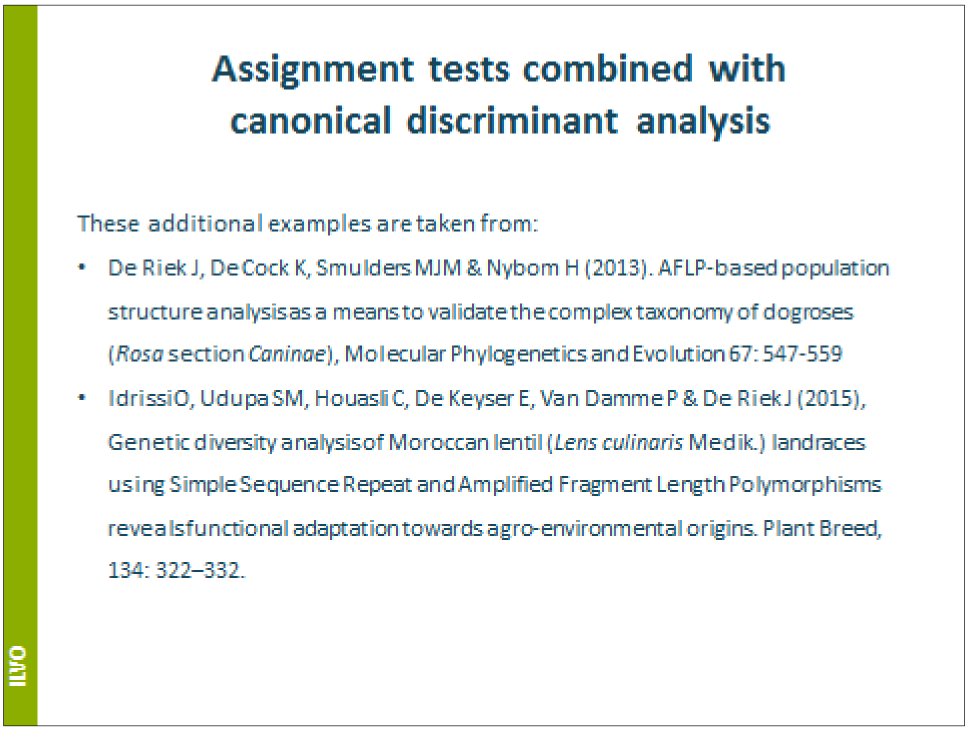
Discriminant analysis based on landraces cycle duration using combined data sets (SSRs, AFLPs and agronomical)

(Idrissi et al. 2015; Plantbreeding)



Discriminant analysis based on landrcaes early vegetative vigour using combined data sets (SSRs, AFLPs and agronomical)

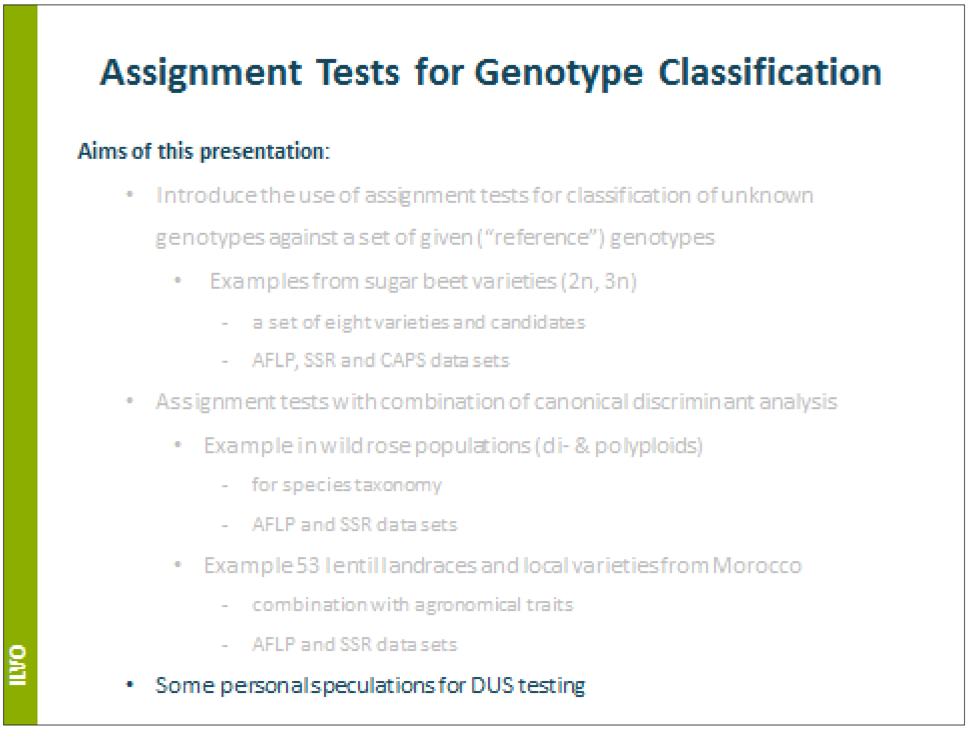
(Idrissi et al. 2015; Plantbreeding)



Assignment tests combined with canonical discriminant analysis

These additional examples are taken from:

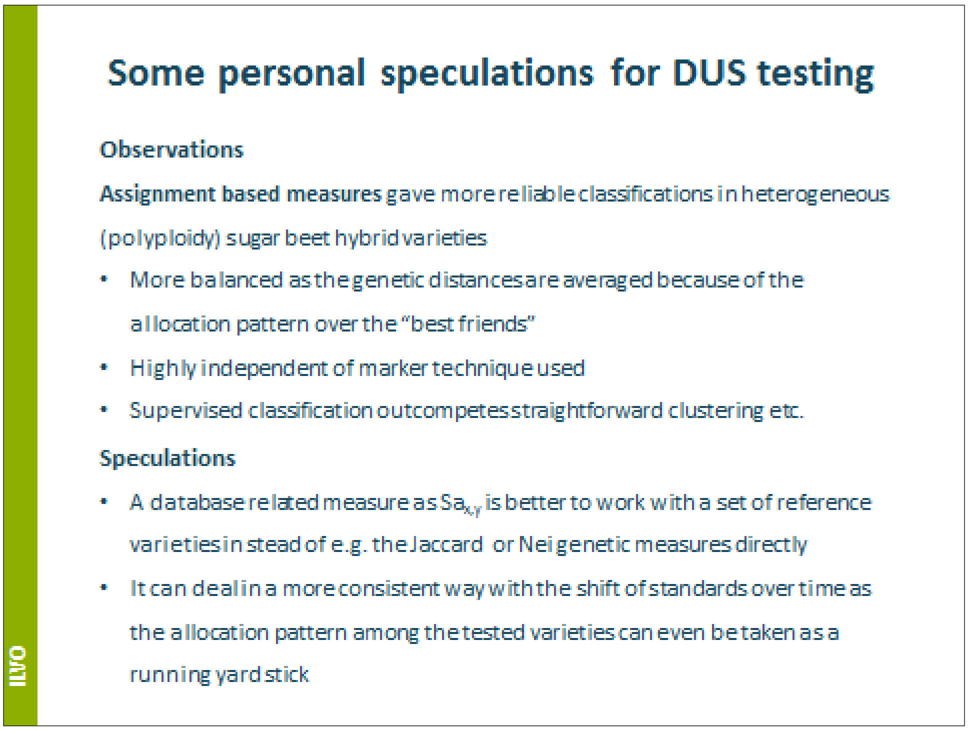
- De Riek J, De Cock K, Smulders MJM & Nybom H (2013). AFLP-based population structure analysis as a means to validate the complex taxonomy of dogroses (*Rosa* section *Caninae*), Molecular Phylogenetics and Evolution 67: 547-559
- Idrissi O, Udupa SM, Houasli C, De Keyser E, Van Damme P & De Riek J (2015), Genetic diversity analysis of Moroccan lentil (*Lens culinaris* Medik.) landraces using Simple Sequence Repeat and Amplified Fragment Length Polymorphisms reveals functional adaptation towards agro-environmental origins. Plant Breed, 134: 322–332.



Assignment Tests for Genotype Classification

Aims of this presentation:

- Introduce the use of assignment tests for classification of unknown genotypes against a set of given ("reference") genotypes
    - Examples from sugar beet varieties (2n, 3n)
        - a set of eight varieties and candidates
        - AFLP, SSR and CAPS data sets
- Assignment tests with combination of canonical discriminant analysis
    - Example in wild rose populations (di- & polyploids)
        - for species taxonomy
        - AFLP and SSR data sets
    - Example 53 lentil landraces and local varieties from Morocco
        - combination with agronomical traits
        - AFLP and SSR data sets
- Some personal speculations for DUS testing



Some personal speculations for DUS testing

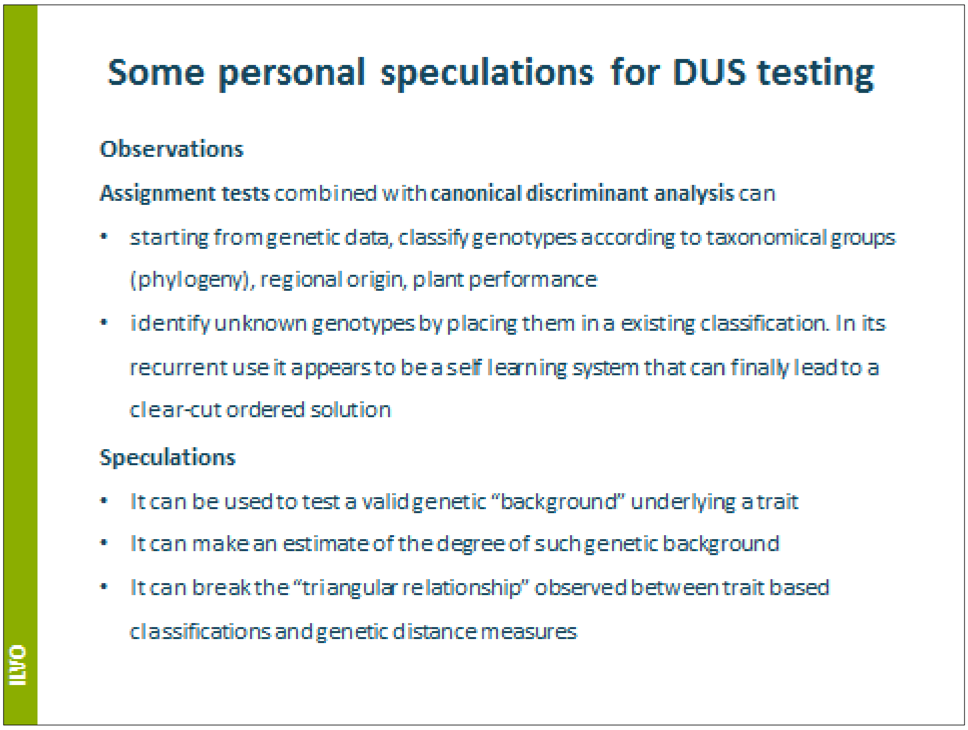
Observations

Assignment based measures gave more reliable classifications in heterogeneous (polyploidy) sugar beet hybrid varieties

- More balanced as the genetic distances are averaged because of the allocation pattern over the "best friends"
- Highly independent of marker technique used
- Supervised classification outcompetes straightforward clustering etc.

Speculations

- A database related measure as $Sa_{x,y}$ is better to work with a set of reference varieties in stead of e.g. the Jaccard or Nei genetic measures directly
- It can deal in a more consistent way with the shift of standards over time as the allocation pattern among the tested varieties can even be taken as a running yard stick



Some personal speculations for DUS testing

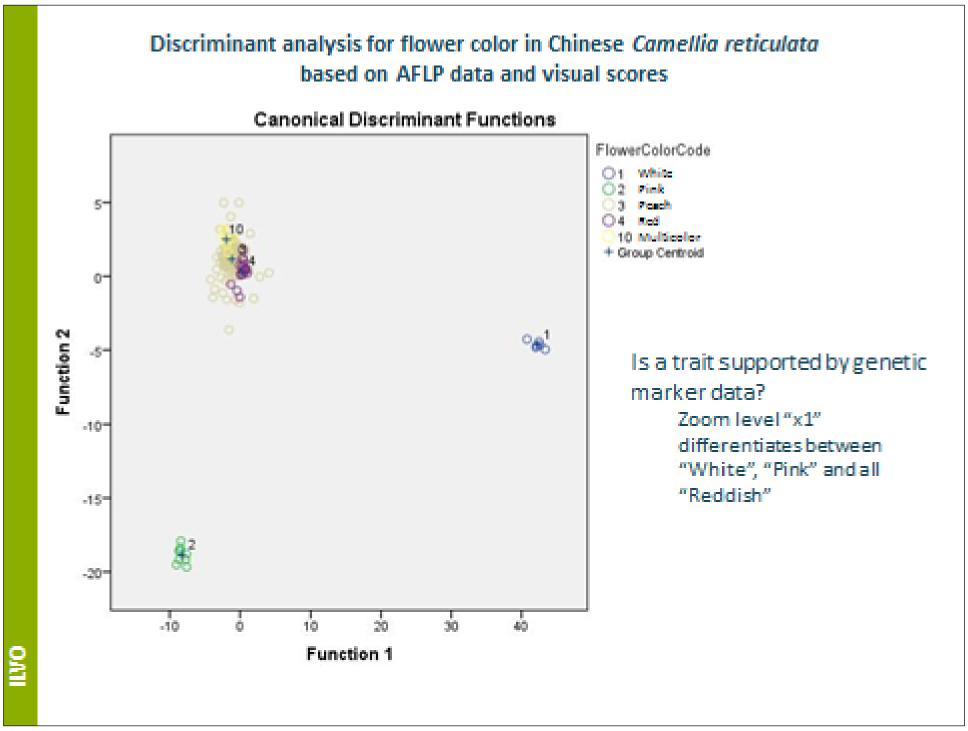
Observations

Assignment tests combined with **canonical discriminant analysis** can

- starting from genetic data, classify genotypes according to taxonomical groups (phylogeny), regional origin, plant performance
- identify unknown genotypes by placing them in a existing classification. In its recurrent use it appears to be a self learning system that can finally lead to a clear-cut ordered solution

Speculations

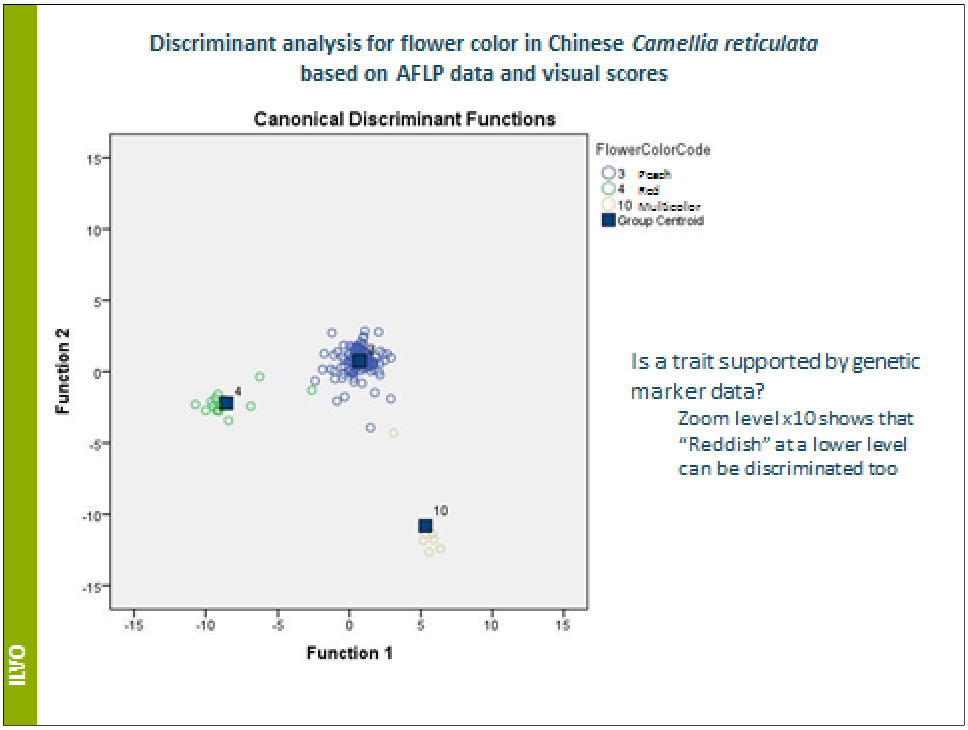
- It can be used to test a valid genetic "background" underlying a trait
- It can make an estimate of the degree of such genetic background
- It can break the "triangular relationship" observed between trait based classifications and genetic distance measures



Discriminant analysis for flower color in Chinese *Camellia reticulata* based on AFLP data and visual scores

Is a trait supported by genetic marker data?

Zoom level "x1" differentiates between "White", "Pink" and all "Reddish"



Discriminant analysis for flower color in Chinese *Camellia reticulata* based on AFLP data and visual scores

Is a trait supported by genetic marker data?

Zoom level x10 shows that "Reddish" at a lower level can be discriminated too

Thank you

Flanders Research Institute for
Agriculture, Fisheries and Food
Plant Genetics and Breeding
Caritasstraat 39
9090 Melle – Belgium
T +32 (0)9 272 29 00
F +32 (0)9 272 29 01

Jan.deriek@ilvo.vlaanderen.be
www.ilvo.vlaanderen.be

ILVO

Flanders
is agriculture and fisheries

ILVO

[End of Annex and of document]