Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular	**BMT/16/18**
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]

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 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

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

ILVO

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.

Datasets on Amplified Fragment Length Polymorphisms, Sequence Tagged Microsatellite Sites and Cleaved Amplified Polymorphic Sites markers in eight 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.

ILVO

(De Riek et al. 2007; Crop Science)

Variety Identification in Sugar Beet Varieties
Research Aims

Major issues addressed are

(i) which classification method gives the most consistent results when three marker techniques are compared,

(ii) given different classification techniques available, for which marker technique is the output generated least constrained by the way data analysis is performed.

(De Riek et al. 2007; Crop Science)

Variety Identification in Sugar Beet Varieties
Assignment 10 most similar plants

		To							
	From	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	27	26	12	22	15
	H66377	10	20	14	138	26	30	37	25
	KWS8123	21	28	14	26	169	12	22	8
	MK9907	14	27	22	51	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	16	15	37	14	46	39
	Fortis	33	37	101	5	31	32	31	30
	H66377	30	26	5	145	15	26	30	23
	KWS8123	14	24	24	11	195	1	24	7
	MK9907	26	29	34	46	12	92	25	36
	Princesse	29	48	25	13	23	14	105	33
	Sylvester	21	63	24	18	13	26	58	77
STMS	Ariana	137	88	6	10	14		33	12
	Aurelia	62	149	20	12	28	4	16	9
	Fortis	3	27	195	24	16	1	14	20
	H66377	4	11	28	194	8	15	13	27
	KWS8123	4	9	6	1	275	2	2	1
	MK9907		1	6	19	8	255	2	9
	Princesse	45	32	18	10	16	4	161	14
	Sylvester	12	16	28	29	16	14	15	170

Assignment based on the pairwise Jaccard similarity data

ILVO

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

Similarity by assignment $Sa_{x,y}$

	From	To 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

ILVO

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)

ILVO

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.

ILVO

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.

ILVO

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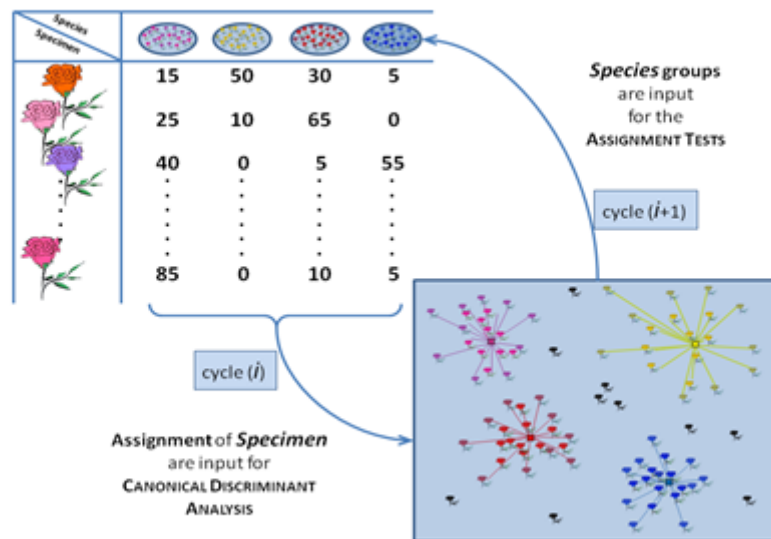
Assignment test scheme

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

Moroccan local lentil landraces and varieties

Assignment of genotypes to their origin (collection sites or variety) was computed using the assignment test combined with canonical discriminant analysis (De Riek et al. 2001, 2013) using SPSS-Statistics 22 to display genetic variation coming from the origin of the landraces.

The assignment test was carried out by first ranking all individual genotypes (single plants) to each other based on chi-square distances for the SSRs analysis and on Jaccard's similarity index for AFLP.

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 origin (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 their origins.
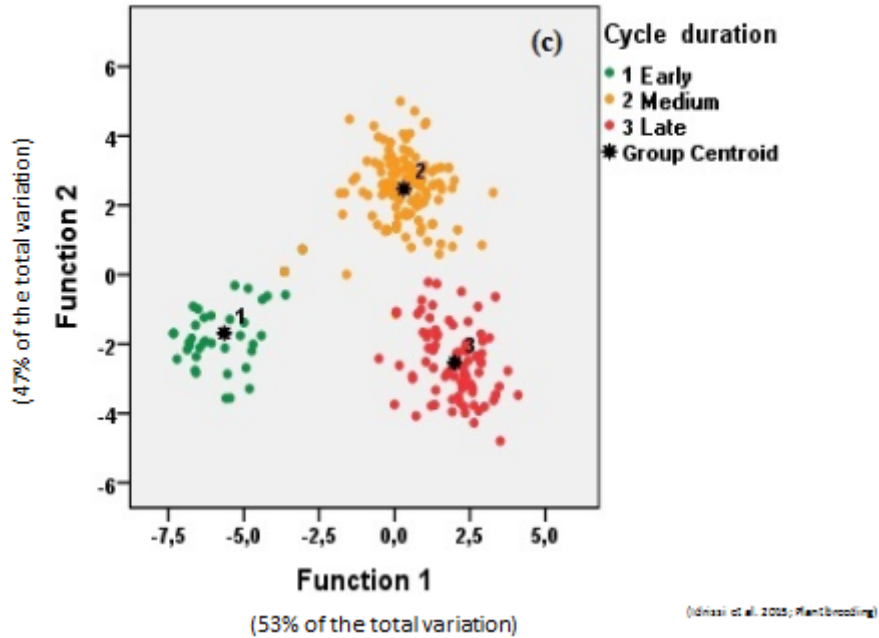
(Idrissi et al. 2015; Plant breeding)

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

(Idrissi et al. 2015; Plant Breeding)
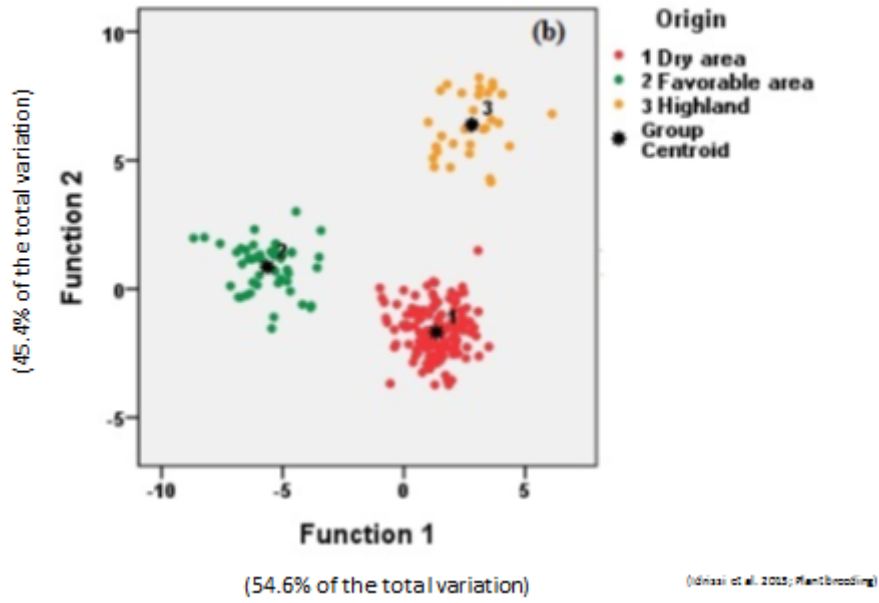
Discriminant analysis based on landraces cycle duration using combined data sets (SSRs, AFLPs and agronomical)

(Idrissi et al. 2015; Plant Breeding)

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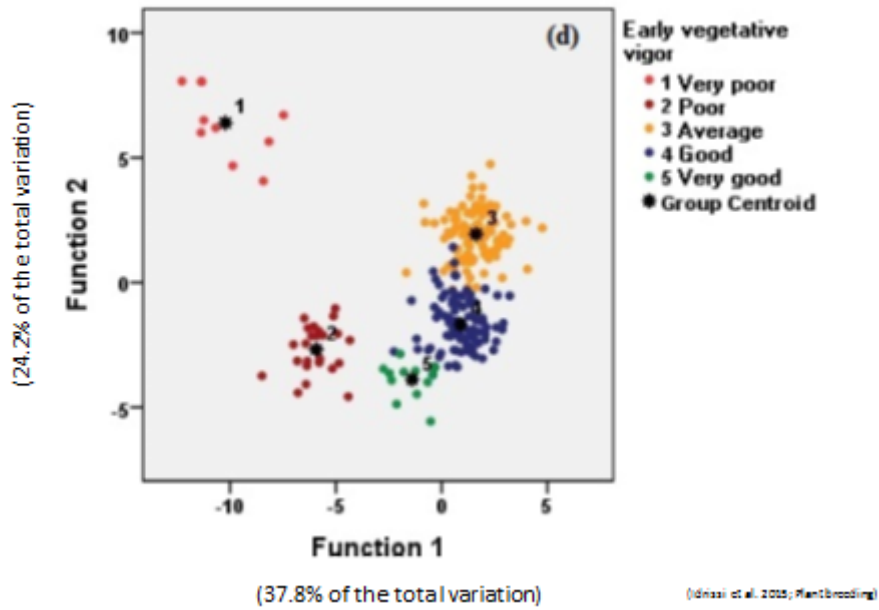
Thank you

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

[End of Annex and of document]