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/17/12 Add.

Seventeenth Session
Montevideo, Uruguay, September 10 to 13, 2018

Original: English
Date: October 8, 2018

ADDENDUM TO CONSTRUCTION OF A EUROPEAN POTATO DATABASE WITH VARIETIES OF COMMON KNOWLEDGE AND ITS IMPLEMENTATION IN THE POTATO DUS TESTING SYSTEM PART II: GENERATION OF MOLECULAR DATA

Document prepared by experts from the United Kingdom and the Netherlands

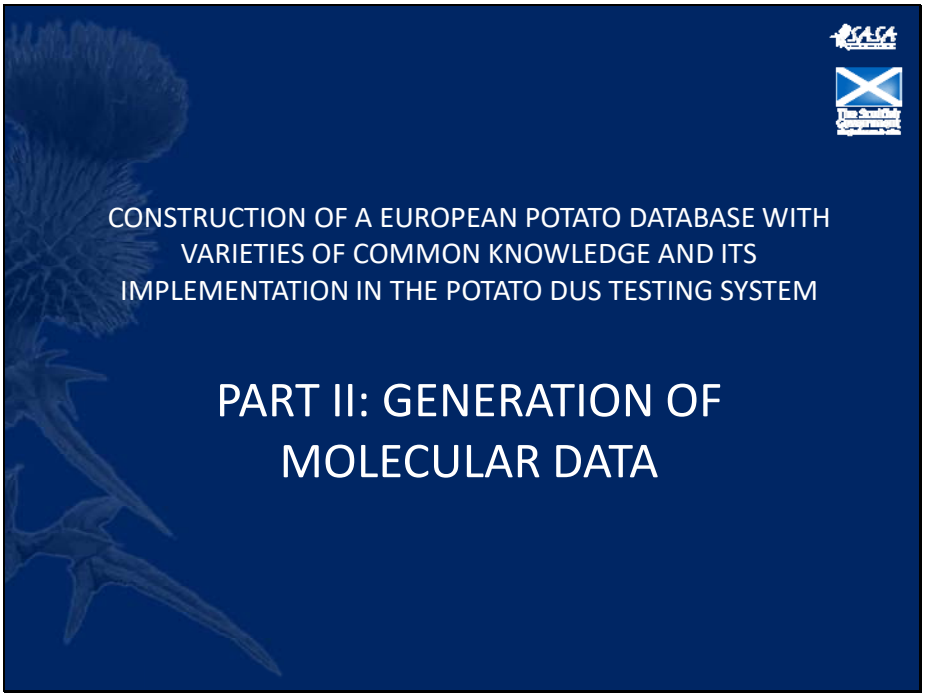
The Annex to this document contains a copy of a presentation on “Construction of a European potato database with varieties of common knowledge and its implementation in the potato DUS testing system Part II: Generation of molecular data”, prepared by experts from the United Kingdom and the Netherlands, which was made at the seventeenth session of the Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular (BMT).

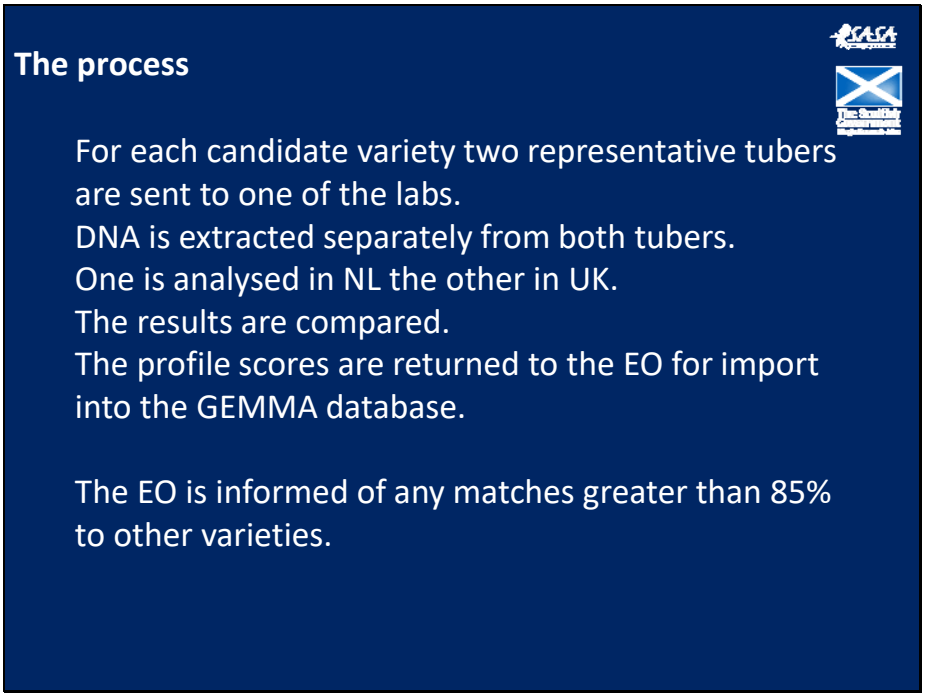
[Annex follows]

ANNEX

CONSTRUCTION OF A EUROPEAN POTATO DATABASE WITH VARIETIES OF COMMON KNOWLEDGE AND ITS IMPLEMENTATION IN THE POTATO DUS TESTING SYSTEM
PART II: GENERATION OF MOLECULAR DATA

Presentation prepared by experts from the United Kingdom and the Netherlands





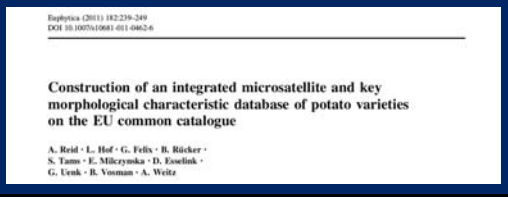
Evolution of molecular analysis

2004 Centre for Genetic Resources/Plant Research International (CGN/PRI) and SASA collaborated to produce, from the public domain, a harmonized set of 9 SSR markers for potato variety differentiation which yield results that are both robust and easy to interpret.

Name	Repeat motif	Linkage group	Number of alleles	PIC value	Reference
STMS 0019	$(AT)_7$ $(GT)_{10}$ $(AT)_4$ $(GT)_5$ $(GC)_4$ $(GT)_4$	VI	10	0.92	Milbourne *et al.*, 1998
STMS 2005	$(CTGTTG)_3$	XI	6	0.8	Milbourne *et al.*, 1998
STMS 2028	$(TAC)_5.(TA)_3.(CAT)_3$	XII	9	0.9	Milbourne *et al.*, 1998
STMS 3009	$(TC)_{13}$	VII	14	0.8	Milbourne *et al.*, 1998
STMS 3012	$(CT)_4.(CT)_8$	IX	7	0.87	Milbourne *et al.*, 1998
STMS 3023	$(GA)_9.(GA)_8.(GA)_4$	IV	4	0.79	Milbourne *et al.*, 1998
STMS 5136	$(AGA)_5$	I	11	0.92	Ghislain *et al.*, 2004
STMS 5148	$(GAA)_{17}$	V	20	0.98	Ghislain *et al.*, 2004
STMS SSR1	$(TCAC)_n$	VIII	14	0.93	Kawchuk *et al.*, 1996

Evolution of molecular analysis

These 9 markers have been used for the European Potato Database since 2006 (see BMT-TWA/POTATO/1/4, BMT/10/5, BMT/11/9 and BMT/11/10 for background).
During the initial stages both laboratories carrying out the molecular work used a similar capillary based platform making harmonization 'relatively' easy.
The first European potato project ran from 2006 to 2008 and typed around 900 varieties in the EU Common Catalogue.



Euphytica (2011) 182:239–249
DOI 10.1007/s10681-011-0462-6

Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU common catalogue

A. Reid · L. Hof · G. Felix · B. Rücker · S. Tams · E. Milczynska · D. Esselink · G. Uenk · B. Vosman · A. Weltz

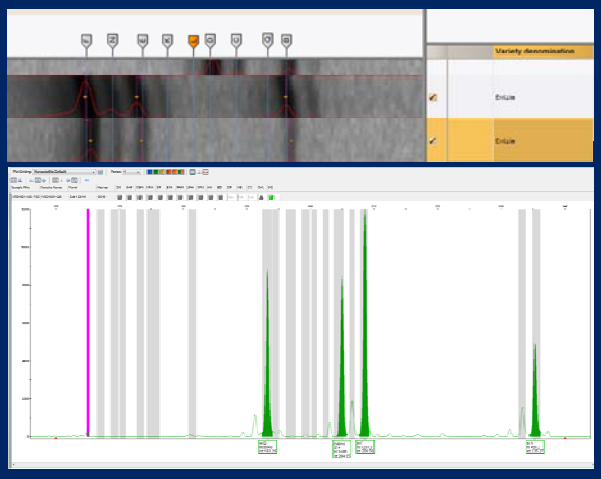
Evolution of molecular analysis

There have now been two subsequent projects culminating this year in the end of potato project III.

After the end of the initial project the SSR analysis of potato varieties in the Netherlands moved labs to Naktuinbouw and was set up using a different gel based platform.
NL – Li Cor gel based system
UK – Thermo Fisher capillary based system

This required some additional harmonization.

Because the two labs use different platforms



Not surprisingly the data looks quite different and when we compare results there can be discrepancies.

What are these discrepancies?

Category 1

There is a new allele (for example the allele-bins for 2028:F and 5148B actually contained 2 separate alleles now called 2028:F & K and 5148:B & W).

Normally resolved fairly easily by the labs agreeing on the presence of a new allele. We also find alleles in totally new positions some years.

Category 2

An allele was miscalled by one lab (can either be missed completely or assigned an incorrect score).

Normally resolved fairly easily by both labs checking their data and agreeing on the correct call.

What are these discrepancies?

Category 3

Genuine difference in the profiles obtained by the two labs. Fortunately a rare occurrence, usually the result of a sample with poor DNA quality that does not amplify well (a rare occurrence but between the two labs and two tubers we have always managed to obtain a complete profile for the candidate).

Alternatively, this can be caused by a mixture of varieties, this is very easy to spot as the profiles are very different.

What are these discrepancies?

Category 4

An allele is called as questionable by one lab (the lab cannot decide whether to call an allele or not as it is on the threshold) and IS called by the other lab.

Category 5

An allele is called as questionable by one lab (the lab cannot decide whether to call an allele or not as it is on the threshold) and is NOT called by the other lab.

Effect of discrepancies

Discrepancies have an effect on downstream statistical analysis.
Mistakes in the raw data (allele scores or DNA profile) can lead to incorrect similarity values.
In BioNumerics alleles can be entered as either
not present, a score of 0
present, a score of 1
or unknown (discrepancy types 4 & 5) which are scored as a ?
In this case the data point is ignored in the analysis.
On this basis, a different allele score of type 1, 2 and 3 would result in two samples not matching when they should because the similarity value of the pairwise comparison is affected.
Differences in allele scores of type 4 and 5 would not have a direct effect on the similarity value and are therefore less critical.
However, too many missing data points will affect similarity values.

The option of ? as an allele score is not possible in GEMMA at the moment.

Discrepancies after initial screen – year one

Year	# samples	# discrepancies	discrepancy %
2013	121	74	61.2

Doesn't look good!

4 year gap between the end of the 1st project and the beginning of the 2nd.

Analysis in the Netherlands changed to a gel based system.

Several new alleles were discovered.

So most of the differences were very easy to rectify.

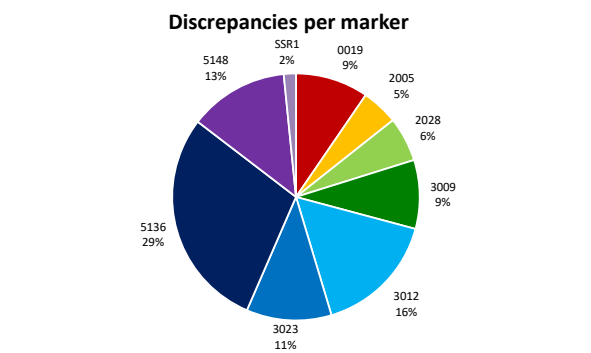
Discrepancies after initial screen – later years

Year	# samples	# discrepancies	discrepancy %
2013	121	74	61.2
2014	208	82	39.4
2015	156	32	20.5
2016	147	36	24.5
2017	121	27	22.3
2018	116	16	13.8
Total	869	267	30.7

DNA profiling is a routine activity in both labs and we both work according to an agreed on set of decision rules and there has been a steady improvement.

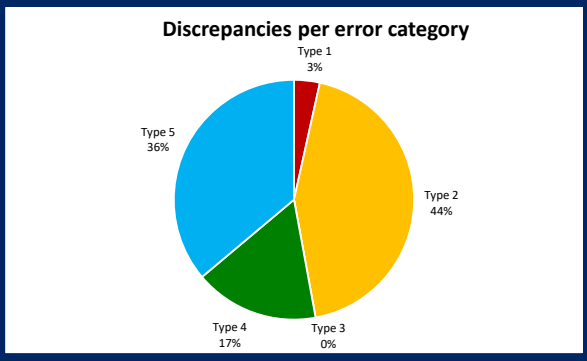
The big improvement in 2018 could partly be to new personnel performing the testing.

Discrepancies per marker



0019 generally due to new alleles (1) and a few difficult to call alleles (4&5)
2005 generally due to miss-called alleles (2)
2028 generally due to new alleles (1) and a few miss-called (2)
3009 generally due to alleles difficult to score around cut off thresholds (4&5)
3012 generally due to alleles difficult to score around cut off thresholds (4&5)
3023 generally due to one allele around the threshold value for scoring (4&5)
5136 generally due to one allele in particular (4&5)
5148 generally due to new alleles (new alleles 1bp different from existing allele)
SSR1 generally due to new alleles (1) and a few miss-called alleles (2)

Discrepancies per error category



Discrepancies of type 1, 2 and 3 are critical (have effect on similarity values). Type 4 and 5 are less critical. Both labs consider a type 4 discrepancy as not being a problem as the lab calling the allele as questionable thinks that there might be an allele present but it falls slightly below a predetermined threshold in the analysis software and the other lab definitely calls the allele as present. However, type 5 is more worrying as one lab thinks there might be an allele present but the other lab does not.
In this figure there are no Type 3 discrepancies. This is due the fact that the admixes were left out of this analysis. So far, all type 3 errors were traceable admixes.

Admix detection

Year	# Samples	# Admixes	Admix %
2013	121	0	0.00
2014	208	1	0.48
2015	156	1	0.64
2016	147	0	0.00
2017	121	0	0.00
2018	116	0	0.00
Total	**869**	**2**	**0.23**

The rate of admixtures submitted for testing has been very low.
In addition to the two detected a further admix was detected at the light sprout stage and checked by SSR analysis to confirm this observation.

The conclusion

Both labs have developed decision rules which evolve as necessary.
They are used as guidance in the interpretation and scoring of the alleles.
There is not a better or more preferred platform.
Indeed the use of the two platforms strengthens the system and makes the method more transferable.
Reciprocal results checking eliminates errors.

Up to date marker information

A number of new alleles have been detected since the original publication (including 2 new ones this year).

Name	Number of alleles 2006	Number of alleles 2018
STMS 0019	10	18
STMS 2005	6	7
STMS 2028	9	11
STMS 3009	14	18
STMS 3012	7	8
STMS 3023	4	6
STMS 5136	11	12
STMS 5148	20	23
STMS SSR1	14	15

This gives a grand total of 118 alleles.
There are now over 2000 entries that have been genotyped.

[End of Annex and of document]