



BMT/14/9

ORIGINAL: English

DATE: October 13, 2014

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**

Geneva

**WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES  
AND DNA-PROFILING IN PARTICULAR****Fourteenth Session  
Seoul, Republic of Korea, November 10 to 13, 2014****DETERMINING A THRESHOLD FOR GENETIC CONFORMITY IN POTATO SEEDLINGS***Document prepared by an expert from the United Kingdom**Disclaimer: this document does not represent UPOV policies or guidance***INTRODUCTION**

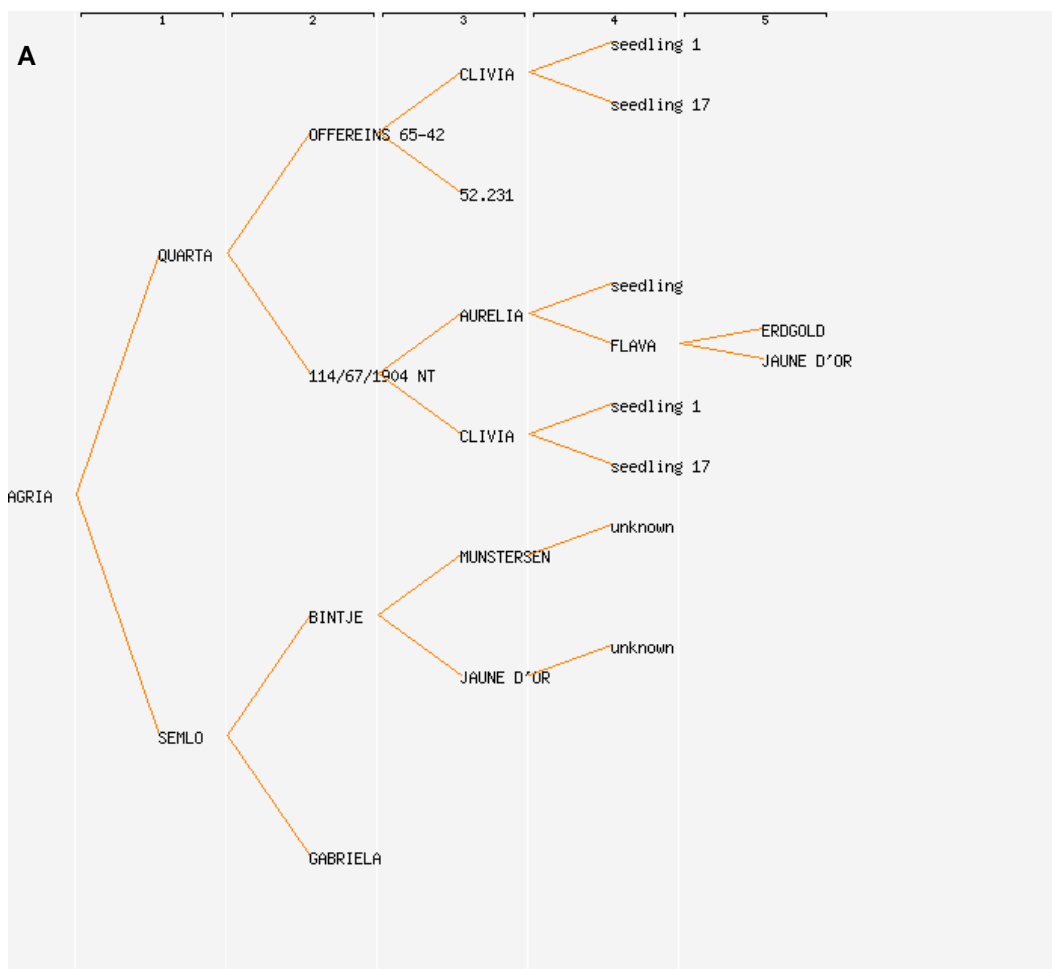
1. Numerous DNA based 'fingerprinting' studies, mainly using microsatellite markers or SSRs, have focused on commercial potato variety identification. Between 2006 and 2008, Science and Advice for Scottish Agriculture (SASA) took part in a Community Plant Variety Office of the European Union (CPVO) co-funded project to screen the varieties on the European Union Common Catalogue (EUCC) using 9 microsatellite markers (Reid *et al.*, 2011). In addition to these varieties SASA has built a library of over 1500 potato varieties typed with these 9 markers as well as an additional 3 markers. The results of these studies showed that commercial varieties can easily be differentiated using this technology. The exceptions to this rule are somaclonal variants such as King Edward and Red King Edward which yielded identical results and a small number of pairs of varieties which intuitively should be different but proved not to be. Most of these aberrant pairs can, however, be explained. Firstly, some varieties have been assigned different names in different countries (e.g. Asparagus and Ratte). The second explanation is that at some point a mix up occurred between varieties at one location resulting in one variety subsequently been distributed with the wrong name (we know that such mix ups have occurred in the past as we also found a small number of varieties from different collections that had the same name and were proved to be totally different varieties). A third explanation is that it is possible for two *bona fide* varieties to yield identical fingerprints although the chance of this was calculated at 1 in 2.8 million during the CPVO project (based on only 9 markers). It is, however, likely that for varieties with similar lineages this number would be much lower.

2. The purpose of this investigation was to determine the levels of genetic variability within seedlings (approximately 1000) produced by 5 crosses of 3 potato varieties using 12 microsatellite markers and to postulate a genetic threshold for the reversal of the burden of proof for essential derivation. Numerous studies to determine the EDV threshold for other crop species have been documented, for example Maize (Heckenberger *et al.*, 2005a, 2005b and 2006), barley and lettuce (van Eeuwijk and Law, 2004) and durum wheat (Noli *et al.*, 2012). Commonly the "tail principle" (van Eeuwijk and Law, 2004) is used to determine the threshold as this method does not require any *a priori* knowledge of the parents of the EDV. The tail principle involves analysis of the distribution of pairwise similarity values and the setting of a similarity value above which the disputed variety is deemed to be essentially derived from the initial variety. The International Seed Federation (ISF) have guidelines for determining the threshold for essential derivation and the reversal of the burden of proof (see <http://www.worldseed.org/isf/edv.html>). Below the set threshold there is no presumption of essential derivation while above it there is presumption of essential derivation and the burden of proof of non predominant derivation would fall on the breeder of the putative EDV. The threshold varies from species to species depending on the existing genetic variability within the species and the established breeding procedures. For example the threshold for lettuce is set at 0.96 Jaccard similarity using AFLPs, for oilseed

rape the trigger point is a Dice value of 0.85 for both winter and spring varieties, for rye grass it is a Jaccard of 0.6, tomato is 0.78 and cotton 0.875. Maize has a two tier system at 0.82 conformity the burden of proof shifts to breeder of putative EDV while 0.9 conformity is viewed as a strong indication of predominant derivation. During the CPVO project around 900 potato varieties were assessed from the 2006 EU Common Catalogue and the mean Jaccard coefficient was calculated at around 45% with 5 pairs between 85-90% similarity and only 1 pair between 90-95% (excluding somaclonal variants and varieties suspected to have been mislabeled). However, as most of the varieties tested were either unrelated or their parentage was not known it was not possible to set a threshold from these data hence this study.

## MATERIALS AND METHODS

3. The three varieties used in this study were Agria, Fontane and Ramos. Examination of the lineages of the 3 parent varieties (van Berloo *et al.*, 2007) show that both Fontane and Ramos have Agria as one parent, and, that the other side of the crosses also have a common background (Figure 1).



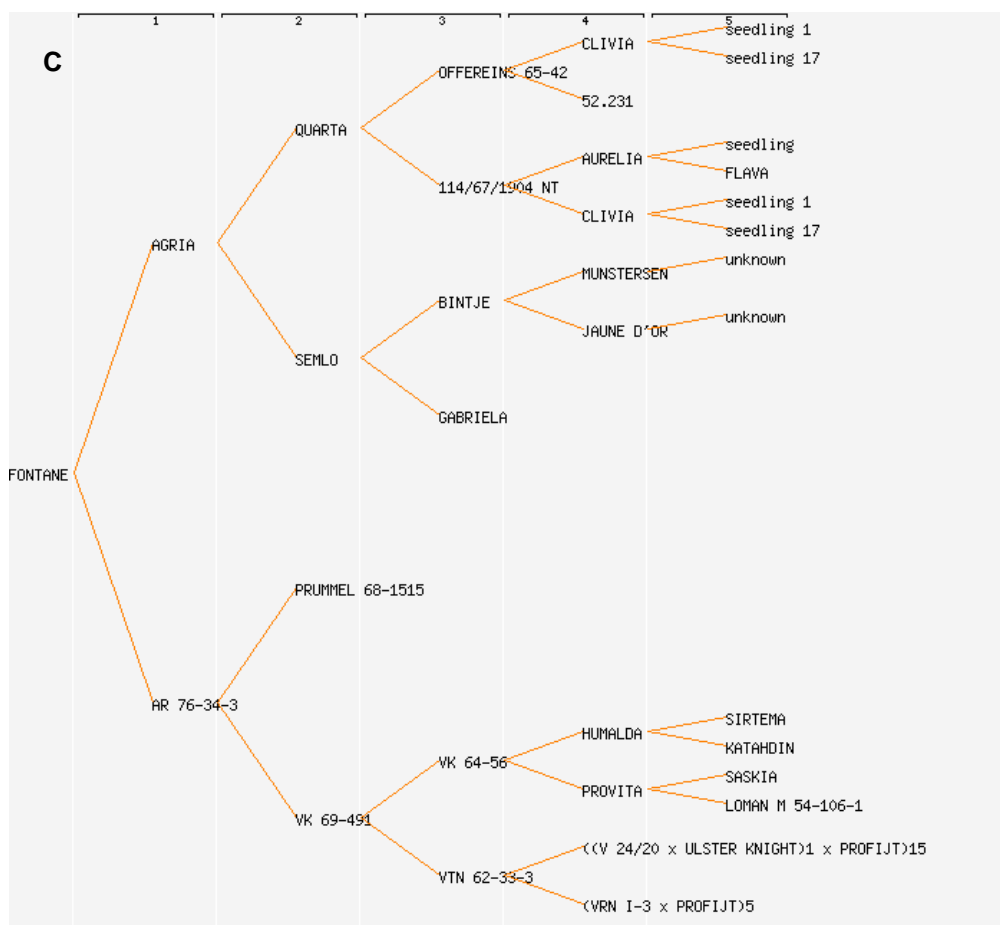
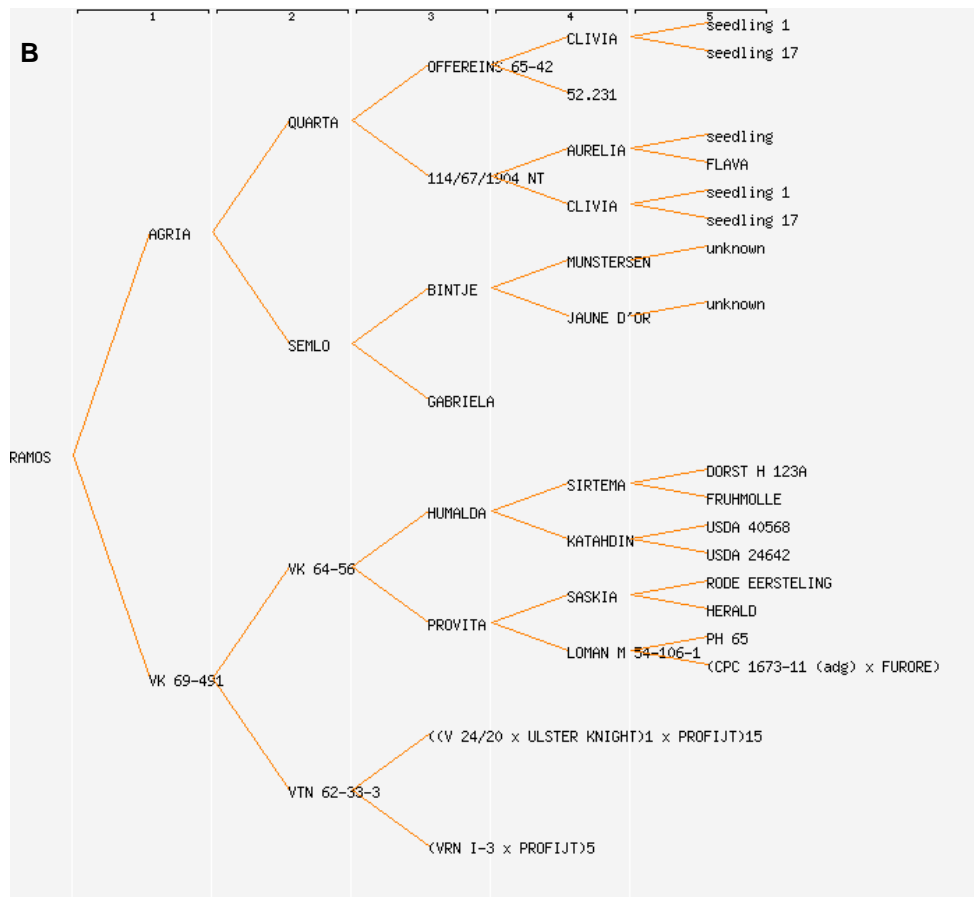


Figure 1. Pedigrees (last 5 generations) for (A) Agria, (B) Ramos and (C) Fontane. Data from Potato Pedigree Database ([www.plantbreeding.wur.nl/potatopedigree](http://www.plantbreeding.wur.nl/potatopedigree)).

4. Details of the crosses are shown in Table 1 and properties of the microsatellite markers used to analyze them in Table 2. The 12 microsatellites were amplified as per Reid *et al.* (2011) with the addition of multiplex set 4 containing the additional 3 markers. These data were stored and analyzed in BioNumerics v6.1 (Applied Maths). Similarity values were calculated using the Jaccard coefficient and the results plotted using Principle Components Analysis (PCA) and Multi-Dimensional Scaling (MDS) using the default settings for analysis within BioNumerics.

Table 1. Combinations of crosses examined in this study. The color code shows the colors used to depict the various sets of samples in subsequent figures.









Variety	Type of cross	DNA Plate numbers	Colour code
Ramos	Selfing	EDV#2-1 & EDV#2-2	
Fontane	Selfing	EDV#4-1 & EDV#4-2	
Agria x Ramos	Cross	EDV#6-1 & EDV#6-2	
Agria x Fontane	Cross	EDV#8-1 & EDV#8-2	
Ramos x Fontane	Cross	EDV#11-1 & EDV#11-2	
Agria	Parent from SASA db		
Fontane	Parent from SASA db		
Ramos	Parent from SASA db		

Table 2. Marker information showing the repeat motif of the microsatellite, linkage group, numbers of alleles and original reference.

Marker name	Repeat motif	Linkage group	Number of alleles	Reference
STMS 0019	(AT) <sub>7</sub> (GT) <sub>10</sub> (AT) <sub>4</sub> (GT) <sub>5</sub> (GC) <sub>4</sub> (GT) <sub>4</sub>	VI	10	Milbourne <i>et al.</i> , 1998
STMS 1016	(TCT) <sub>9</sub>	VII	16	Milbourne <i>et al.</i> , 1998
STMS 1024	(TTG) <sub>6</sub>	VIII	8	Milbourne <i>et al.</i> , 1998
STMS 2005	(CTGTTG) <sub>3</sub>	XI	6	Milbourne <i>et al.</i> , 1998
STMS 2022	(CAA) <sub>3</sub> ..(CAA) <sub>3</sub>	II	7	Milbourne <i>et al.</i> , 1998
STMS 2028	(TAC) <sub>5</sub> .(TA) <sub>3</sub> .(CAT) <sub>3</sub>	XII	9	Milbourne <i>et al.</i> , 1998
STMS 3009	(TC) <sub>13</sub>	VII	14	Milbourne <i>et al.</i> , 1998
STMS 3012	(CT) <sub>4</sub> .(CT) <sub>8</sub>	IX	7	Milbourne <i>et al.</i> , 1998
STMS 3023	(GA) <sub>9</sub> .(GA) <sub>8</sub> .(GA) <sub>4</sub>	IV	4	Milbourne <i>et al.</i> , 1998
STMS 5136	(AGA) <sub>5</sub>	I	11	Ghislain <i>et al.</i> , 2004
STMS 5148	(GAA) <sub>17</sub>	V	20	Ghislain <i>et al.</i> , 2004
SSR1	(TCAC) <sub>n</sub>	VIII	14	Kawchuk <i>et al.</i> , 1996

5. Due to the shared pedigrees the microsatellite profiles of the 3 parent varieties could be expected to be fairly similar (Table 3). The Potato Pedigree Database also showed that an additional 8 varieties, 4 of which were already in the SASA database (Arcade, Farmer, Florida and Recolta) resulted from crosses between Agria x VK 69-491 whereas Fontane is the only listed progeny of the Agria x AR 76 34-3 cross. The allelic profiles for the additional 4 varieties in the SASA database are also given in Table 3.

Table 3. Allelic phenotypes of 12 microsatellite markers for Agria, Fontane and Ramos and 4 additional varieties resulting from a cross between Agria and VK 69-491.

Variety	Marker											
	0019	1016	1024	2005	2022	2028	3009	3012	3023	5136	5148	SSR1
Agria	BF	DGLM	DEG	BDF	E	ABC	G	BCDF	AB	CDF	AIP	ADFI
Fontane	BF	DHLM	BCDE	DF	E	ABC	FG	BCDF	AB	DEFH	AJMP	ABDF
Ramos	F	DGHL	BCDG	DF	BE	ABC	G	BCF	AB	CEFH	IMOP	BDFI
Arcade	BF	DHLM	BCDG	ABD	BE	ABC	G	CD	AB	CDF	AMOP	ABDI
Farmer	BF	DGH	CEG	ABDF	BE	AB	G	BDF	AB	CDEH	IMOP	BFI
Florida	BF	DGM	CDE	ABDF	BE	BC	G	BCD	ABD	CDEH	LOP	AFI
Recolta	F	DGM	CDG	AD	E	ABC	G	CDF	ABD	EFH	IMOP	AFI

## RESULTS

6. All of the samples tested could be differentiated on the basis of the 12 markers with the exception of 2 progeny from the Fontane selfing (samples 319 & 320, wells G4 & H4 in the DNA plate) and 2 from the Ramos selfing (samples 99 & 102, wells C1 & F1). These pairs were examined using an additional 24 markers and could still not be differentiated. Taking into account the positioning of these pairs of plants in the DNA plates it is possible that these two pairs of DNA samples originated from the same plants.

7. No crosses yielded progeny identical to the parent varieties although the Ramos selfing had 3 plants that only differed by a single allele from Ramos (97.1% similarity) and 2 plants which differed by 2 alleles (94.1% similarity). The Fontane selfing yielded a single plant that differed by two alleles from Fontane (94.5% similarity) all other plants differed by 3 or more alleles from the parent varieties.

8. All crosses yielded pairs of plants with greater than or equal to 90% similarity however the numbers are very different for the two classes of crosses. The Fontane selfing yielded 45 pairs while the Ramos selfing 111 pairs. In contrast the three out crosses yielded 1, 2 and 4 pairs of plants greater than or equal to 90% similarity. Conversely there were pairs of progeny that were very different from each other with similarities of less than 40%.

9. A small number of crosses gave unexpected results. A few plants had more than 5 microsatellite alleles for some of the markers. Intuitively, for a tetraploid this should not be possible although this has been observed in a small number of commercial varieties. There are several possible explanations for this ranging from cross contamination of the sample (i.e. DNA was extracted from more than one plant), through to duplication of the microsatellite region on one or more chromosomes, and the possibility of unequal segregation of chromosomes at gamete formation.

10. Two of the progeny of the Fontane selfing yielded several alleles not previously seen in commercial *Solanum tuberosum* varieties. These plants are sufficiently different to appear to be hybrids with another *Solanum* species (other *Solanum* species were present at the time the crosses were performed). Microsatellite mutation could also be the cause of new alleles but on this scale would appear less likely given the presence of other species.

11. Fourteen other samples yielded alleles not seen in either parent for several of the markers. In particular one from the Fontane selfing yielded alleles not found in Fontane for 8 of the 12 markers. Again this could have been the result of mutation but due to the number of markers exhibiting 'incorrect' alleles they are more likely to be the result of pollination by another variety.

12. Due to the large amount of data it is not feasible to present a dendrogram to show the differences between the seedlings so the data are presented as 2D PCA analysis and 3D MDS analysis. Figure 2 shows the results of the PCA analysis while Figures 3 and 4 show the MDS analysis viewed from different perspectives.

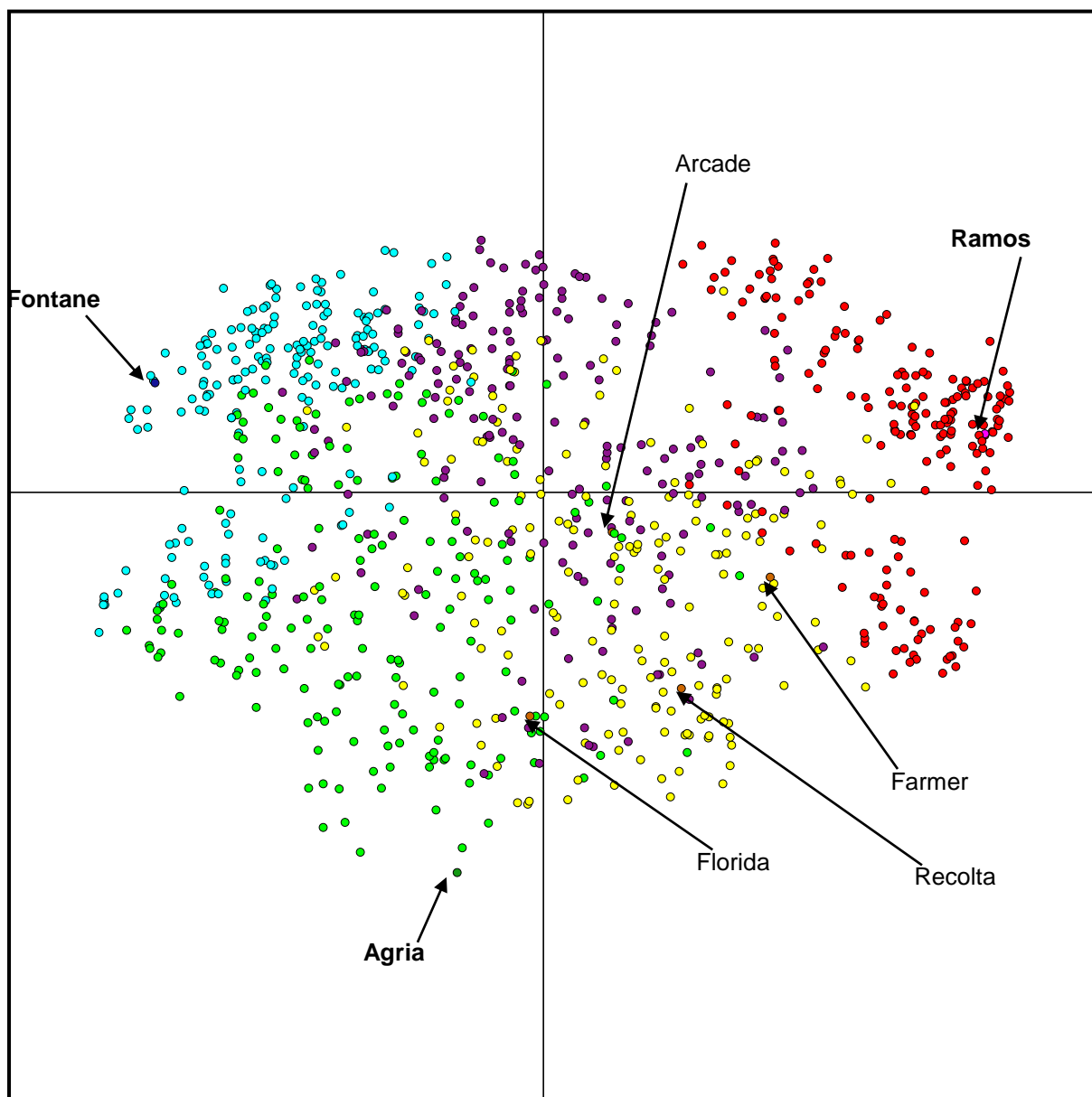


Figure 2. PCA analysis of the results from the microsatellite fingerprinting. The positions of the 3 parent varieties are arrowed along with the additional 4 varieties from the Agria x VK 69-491 cross. Note that these varieties sit between Agria and Ramos Others samples are Ramos selfing (●), Fontane selfing (●), Agria x Ramos (●), Agria x Fontane (●), Ramos x Fontane (●).

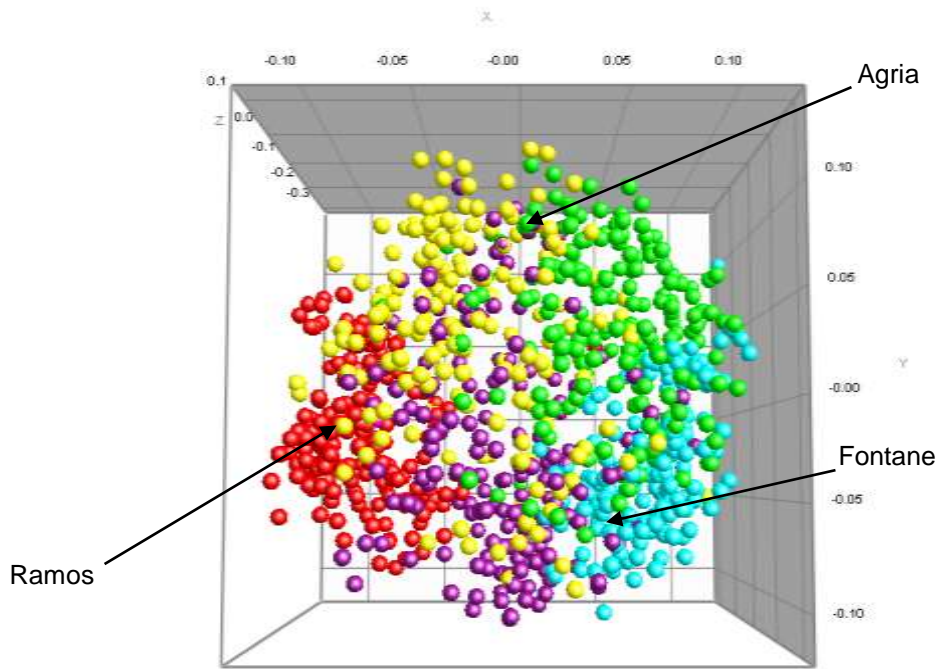


Figure 3. MDS of the microsatellite fingerprinting data viewed from front. Samples are Ramos selfing (●), Fontane selfing (●), Agria x Ramos (●), Agria x Fontane (●), Ramos x Fontane (●). The positions of the 3 parent varieties are arrowed.

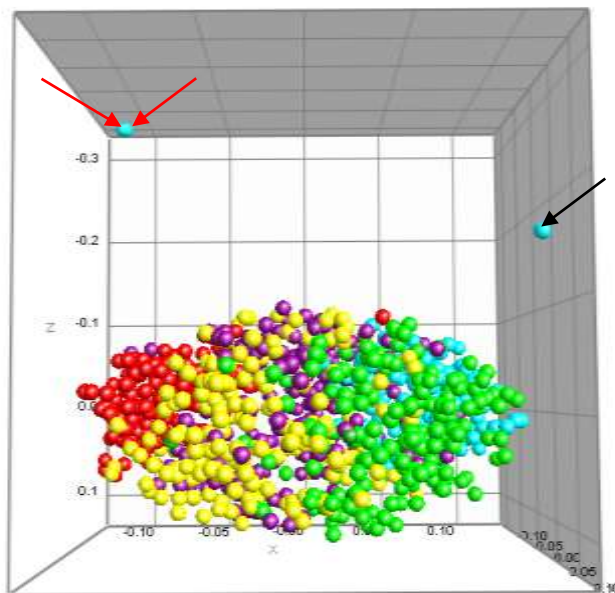


Figure 4. MDS of the microsatellite fingerprinting data viewed from above. Samples are Ramos selfing (●), Fontane selfing (●), Agria x Ramos (●), Agria x Fontane (●), Ramos x Fontane (●). The samples yielding alleles not observed in *S. tuberosum* are indicated by red arrows. A further sample from the Fontane selfing exhibiting many alleles not found in Fontane (8/12 markers) is indicated by the black arrow (the alleles in this plant are found in other commercial potato varieties).

13. The distribution of the pair wise comparisons of the Jaccard values calculated for approximately 900 varieties analyzed during the CPVO project showed a normal distribution with a mean similarity value of around 45% based on 9 markers (Reid *et al.*, 2011). Most of the samples analyzed during the CPVO project were also screened with the additional 3 markers used in this study and all data presented for the CPVO samples in this paper are based on all 12 markers. Similar comparisons of the distributions of the Jaccard coefficients for the samples from the 5 crosses (excluding samples yielding alleles not found in either parent) show similarly normal distributions but with very different means (Figure 5 and Table 4).

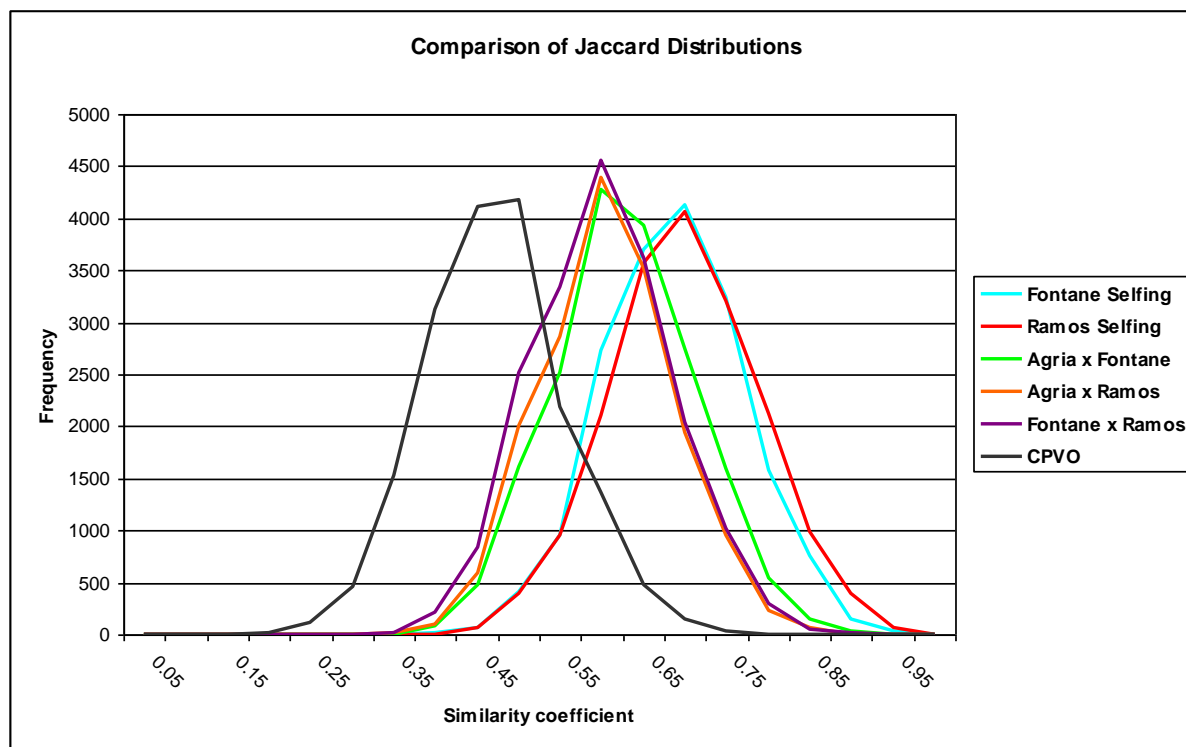


Figure 5. Frequency distributions of Jaccard coefficient pairwise comparisons for the various datasets. N.B. the CPVO data has been normalized to fit the scale of the data from this study.

Table 4. The mean % similarity (Jaccard coefficient), numbers of pairs 85-90%, 90-92%, 92-95% and above 95% similarity for the various data sets. The % of pairs for each category is shown in brackets). N.B. the pairs above 95% similarity for the CPVO database (data for all 12 markers) include somaclonal variants and varieties expected to be mix ups otherwise this value would be 0.

Dataset	Total # of pairs	Mean % similarity	85-90% similarity (% total)	90-92% similarity (% total)	92-95% similarity (% total)	Above 95% similarity (% total)
CPVO database	399,171	44.9	5 (0.001)	0 (0)	1 (0.0002)	23 (0.006)
Agria x Fontane	17,955	60.5	25 (0.139)	3 (0.017)	0 (0)	0 (0)
Agria x Ramos	16,653	58.5	7 (0.042)	2 (0.012)	0 (0)	0 (0)
Fontane x Ramos	18,528	57.9	10 (0.054)	0 (0)	1 (0.005)	0 (0)
Fontane selfing	17,766	66.4	151 (0.850)	27 (0.152)	13 (0.073)	2 (0.011)
Ramos selfing	17,995	67.7	392 (2.178)	36 (0.200)	37 (0.206)	8 (0.045)



## DISCUSSION

14. As could have been expected from the lineages of the three parent varieties the progeny from the crosses show a high degree of similarity. The three 'out crosses' yielded a higher mean similarity (57.9, 58.5 and 60.5%) than the bulk of commercial varieties from the EUCC (44.9%), and the means for the two self crosses were even higher (66.4 and 67.7%). Adopting the tail principle for these data with a cut off set at 85% would result in 6 pairs (excluding somaclonal variants) for the EUCC varieties, 28 pairs for the Agria x Fontane cross, 9 pairs for the Aria x Ramos cross and 11 pairs for the Fontane x Ramos cross. However, adopting the same cut off for the selfings would result in 193 pairs for the Fontane selfing and 473 pairs for the Ramos selfing. Increasing the cut off to above 90% would give 3, 2 and 1 pair for the out crosses and 42 and 81 pairs respectively, setting it at above 92% 1 for the out crosses, 15 and 45 for the selfings and above 95% 2 and 8 pairs for the self crosses and none for the out crosses. It would appear on the basis of these figures that the setting of the threshold for EDV is dependent on the type of cross that has been performed.

15. Future directions would include examining the progeny from crosses of varieties not as closely related as the 3 in this study and to look at the F2 generation progeny of backcrosses. It would also be very interesting to see if there is any correlation between where the individual progeny plants sit within the normal distribution and their morphological and physiological characteristics.

Alex Reid, SASA, Roddinglaw Road, Edinburgh EH12 9FJ, United Kingdom

## REFERENCES

Ghislain, M., Spooner, D. M., Rodríguez, F., Villamón, F., Núñez, J., Vásquez, C., Waugh, R. and Bonierbale, M. (2004) Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theoretical and Applied Genetics*, **108**, 881-890.

Heckenberger, M., Bohn, M & Melchinger, A.E. (2005a) Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: I Simple sequence repeat data from maize inbreds. *Crop Science*, **45**, 1120-1131.

Heckenberger, M., Bohn, M, Klein, D. & Melchinger, A.E. (2005b) Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: II Morphological distance and heterosis in comparison with simple sequence repeat and amplified fragment length polymorphism data in Maize. *Crop Science*, **45**, 1132-1140.

Heckenberger, M., Muminović, J., Rouppe van der Voort, J., Peleman, J., Bohn, M & Melchinger, A.E. (2006) Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: III AFLP data from maize inbreds and comparison with SSR data. *Molecular Breeding*, **17**, 111-125.

Kawchuk, L.M., Lynch, D.R., Thomas, J., Penner, B., Sillito, D. and Kulcsar, F. (1996) characterization of *Solanum tuberosum* simple sequence repeats and application to potato cultivar identification. *American Potato Journal*, **73**, 325-335.

[End of document]