CPV.6608 Final Report

Construction of an integrated microsatellite and key morphological characteristic database of potato varieties in the EU Common Catalogue

Introduction

As outlined in the original project proposal (see Annex 1) the objective was to construct a database for potato varieties in the EU Common Catalogue containing SSR marker data; some key morphological characteristics used in DUS testing and lightsprout photographs. The project started in April 2006 using the 24th edition of the Common Catalogue as a point of reference for the database which contained 1,104 varieties. The aims were to genotype as many of these varieties as possible with 9 microsatellite, also known as simple sequence repeat (SSR), markers in the UK and Netherlands and to collect the key morphological descriptions. A large number of these varieties were already held in at least one of the partners' collections and the remaining varieties were to be obtained from the official maintainers by Poland and Germany. These newly collected samples were to be put up for lightsprout characterization and photography and samples sent to the UK and the Netherlands for DNA extraction and genotyping.

Project Aims

1) To collect samples of varieties on the Common Catalogue not present in the collections of the four partners.

2) Lightsprout characterization of newly collected varieties.

3A) Molecular characterization of varieties with 9 SSR markers.

3B) Comparison of genetic fingerprints of varieties held in more than one of the partner collections as a check on the identity and molecular stability of varieties maintained in different locations.

4) Construction of a database containing SSR data for the varieties on the Common Catalogue and morphological data for as many of these as possible. Lightsprout photographs (where available) to also be included.

5) Validation of the SSR data in the database by a blind test.

6) Exchange of data between the partners.

7) Suggestions for implementation of this technology in future DUS testing.

The official start date of the project was 6th April 2006. This later than anticipated start of the project meant that the timing of some of the milestones was slightly revised. This mainly affects the lightsprout ID and characterization as this can only be undertaken at specific times of the year. The revised milestones table is below.

Milestones/Stage:	**Project month**																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Collection of new varieties on EU list and not present in the four collections	X	X					X	X	X										X	X	X			
2. Lightsprout ID and photographs[1]	X									X	X	X	X									X	X	X
3. Characterisation of collected varieties with SSR markers			X	X	X	X	X	X	X	X					X	X	X	X	X	X	X	X		
4. Database set up	X	X	X	X	X	X																		
5. Construction and validation of the database by blind testing				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
6. Harmonisation & exchange of data between testing stations					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
7. Implementation of the database in PBR protocols																					X	X		
8. Output																						X	X	X
Meetings	X										X											X		

[1)]Lightsprout description only for varieties not present in reference collections. Description is divided over two years because of limited capacity in lightsprout cabinet. Timing of identification within the project depends on starting date, as the window for growing lightsprouts is limited to winter and spring.

The minutes of the three project meetings are annexes 2-4.

1st Meeting, 25th April 2006, Edinburgh (Minutes see Annex 2)
2nd Meeting, 20th February 2007, Wageningen (Minutes see Annex 3)
3rd Meeting 6th February 2008, Hannover (Minutes see Annex 4)

Material

Germany
The identity of samples sent by Germany for all varieties listed in DE has been checked morphologically on all characters of the CPVO guideline in the framework of the DUS testing (DB field "origin of the sample" = Office). All varieties not in the DE references collection have been requested from the maintainer of the variety according to CC. Samples have been sent to SASA after assessing lightsprout characteristics without a comparison with an official standard sample (DB field "origin of sample" = Breeder).

Poland
For all varieties registered in PL the identity of samples dispatched from Poland has been verified in the framework of DUS tests carried out in accordance with the CPVO TP.
Samples of the varieties listed on CC, requested and obtained from the maintainers for the purpose of the project were forwarded to SASA and simultaneously the lightsprout test for these varieties was conducted (but samples not checked against the official standard sample).

The Netherlands
Samples were collected from the official central clone field of the Dutch Inspection Service (NAK). Some samples were obtained from lightsprout tests performed at Naktuinbouw.

United Kingdom
UK samples were obtained from the culture collection held at SASA with a few exceptions which were obtained from breeders.

Outputs

1) Collection of new varieties on EU list and not present in the four collections

In total 362 varieties were requested over the lifetime of the project from official maintainers with 136 returns (37% success rate). Many varieties were requested on more than one occasion. The greatest reason for the varieties that were not obtained was no response from the official maintainer.

2) Morphological characterization.

The morphological characteristics included in the database are shown in Table 1.

CPVO code	Character	Variation of notes in DB
1	Lightsprout: size	1-9
2	Lightsprout: shape	1-5
3	Lightsprout:intensity of anthocyanin colouration of base	1-9
4	Lightsprout: proportion of blue in anthocyanin colouration at base	1-3
5	Lightsprout: pubescence of base	1-9
6	Lightsprout: size of tip in relation to base	1-9
7	Lightsprout: habit of tip	1-5
8	Lightsprout: anthocyanin colouration of tip	1-9
9	Lightsprout: pubescence of tip	1-9
10	Lightsprout: number of root tips	1-9
11	Lightsprout: length or lateral shoots	1-9
24	Plant: frequency of flowers	1-9
28	Flower corolla: intensity of anthocyanin colouration on inner side	1-9
29	Flower corrolla: proportion of blue in anthocyanin colouration on inner side	1-3
31	Plant: time of maturity	1-9
32	Tuber: shape	1-6
34	Tuber: colour of skin	1-7
35	Tuber: colour of base of eye	1-4
36	Tuber: colour of flesh	1-9

Table 1. The morphological characters recorded in the database with corresponding CPVO codes and character states (CPVO TP/23/2).

Morphological data for 733 varieties were collected (see Table 2). These data comprised 622 descriptions from a single country, 99 descriptions from 2 countries, 11 from 3 and 1 from all four (856 entries in total). Lightsprout photographs are available for 392 varieties (26 varieties with photographs from 2 countries).

	Total number of entries	Total number of varieties
Lightsprout data only	80	79
Full morphogical data	776	654
Photographs	418	392

Table 2. Numbers of entries and varieties with different types of data in the database.

Not all of the descriptions were complete, some descriptions were only based on light sprout characteristics (for the newly collected varieties from CC), and for others scores of some characteristics were missing. The year of the description differed with most of the UK descriptions based on the previous UPOV guideline (UPOV TG/23/5), whilst the other countries supplied data based on CPVO TP/23/2 which for the used characteristics is identical to UPOV TG/23/6.

For the used characteristics, both protocols differ in scoring for characteristics 4 (colour of light sprout base), 7 (habit of light sprout tip) and 29 (colour of flower corolla). For UK descriptions, characteristics 4, 7 and 29 were therefore excluded from the comparison. Comparison of the results was based predominantly on the 12 varieties supplied by 3 or more countries, as these give the most informative results. For the selected characteristics the differences between highest and lowest score between descriptions of each variety were calculated, and frequencies of observed differences are presented in Figure 1. For example for variety Adora, of the 19 available characteristics 3 scores were identical for all three countries, 9 characteristics had 1 note difference between scores, 4 characteristics showed 2 notes difference, 1 characteristic showed 3 notes difference and 2 characteristics showed a range of 5(!) notes. For *cv*. Bonanza a six note difference was found between the NL and UK description for characteristic 3: ‘intensity of colour of base of the lightsprout’. This almost suggests that descriptions were made on different varieties, but this could not be supported by the other characteristics. Historically there was a candidate variety with the proposed denomination Bonanza in UK which was never released. The variety in the CC was listed and protected in Germany in 1993. The morphological description of the UK sample was carried out in 1989 and it is therefore probable that this description is of the UK candidate variety. However, the microsatellite profiles of samples of Bonanza from both the UK and Netherlands collections are identical suggesting that now both collections contain the same variety. This is a good example of the confusion that can arise when names are reused and the material in the variety collection is not correctly linked to the relevant administrative data.

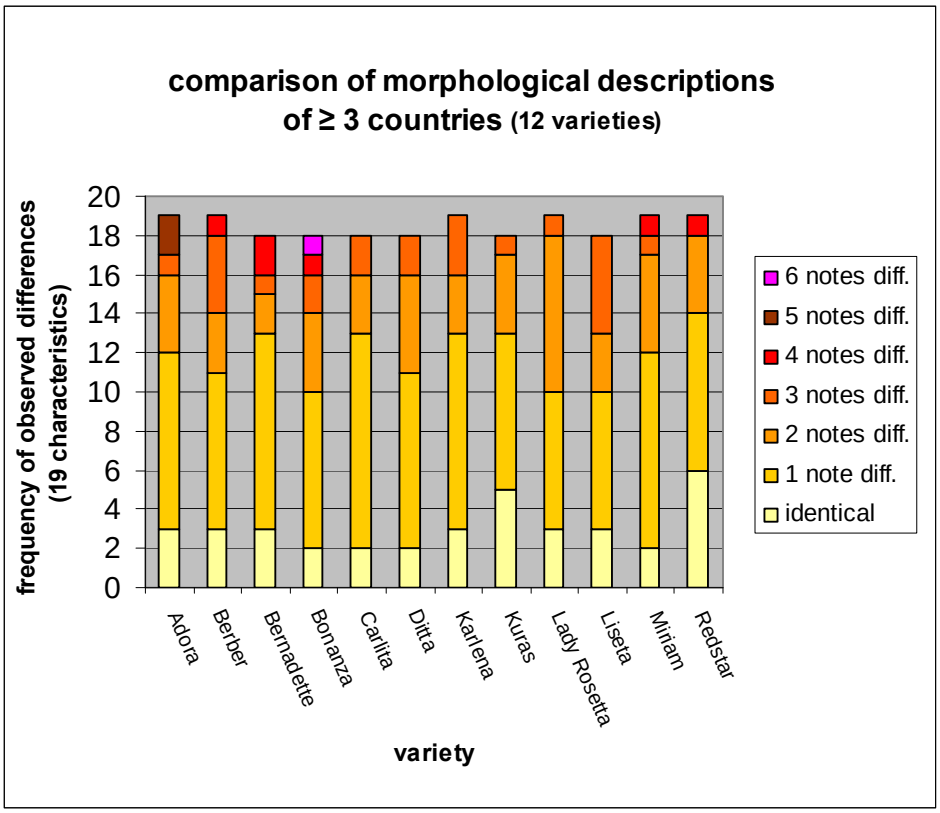

Figure 1. Comparison of morphological descriptions from 3 or more countries. For the selected characteristics the differences between highest and lowest score among descriptions were calculated, and frequencies of observed differences are presented per variety.

Some of the variation is due to the different environmental factors as light/temperature and humidity for light sprout characteristics and weather and soil conditions during field growth for the other characteristics. Discrepancies between descriptions may also partly be due to differences in interpretation of expression of characteristics between observers. For instance, score 1 for skin colour of tuber (a qualitative characteristic) was remarkably often given by the UK, but very rarely or not in the other countries. For characteristic 4 'proportion of blue in anthocyanin colouration of base of lightsprout, score 2 ("intermediate") was given relatively often by Germany, but not so often by Poland and hardly ever by Netherlands.
Based on these data, in almost all cases the 3 descriptions of the 12 varieties would have been declared distinct, whilst in reality should have been similar. From these data it can be concluded that variety descriptions cannot be exchanged between DUS testing stations, and morphological comparison of similar varieties must be carried out side-by-side. The same conclusions were reached in a study of Henk Bonthuis, in the framework of a UPOV study to consider the publication of variety descriptions, presented at TWA/34 in 2005 (addendum 2 to doc. TWA/34/13).

3A) Molecular characterization of varieties with 9 SSR markers.

Details for the nine markers used in the project are given in Table 3.

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

Table 3. Marker information showing the repeat motif of the microsatellite, linkage group, numbers of alleles found during the project and original reference.

In total 1161 samples (including the 20 blind test samples) are included in the database using the 9 SSR markers in Table 3. It should be noted that far more samples than this number were actually analyzed. The UK routinely runs two separate samples for each variety to ensure correct scoring of profiles within the lab. The DNA of most collected varieties was analyzed at two labs, NL and UK, to ensure that the same results were obtained when using different electrophoresis equipment. The dual running of the samples in both laboratories has greatly enhanced the robustness of the database. The cases where the presence or absence of alleles was scored differently between labs were relatively few, and very often referred to the same alleles (e.g. the STMS 0019 G allele, which in NL often was so small it was below the detection level). As a result, this particular allele was decided not to contribute to distinctness between varieties if STMS 0019 G was the only observable difference between samples.
In the case of the small number of problem varieties encountered during the project numerous samples were run where possible. For simplicity these duplicate samples from a single source were removed from the final database, as well as identical profiles from both labs. Altogether 892 varieties from the CC were subjected to SSR analysis. Of these, 669 are represented by samples collected from a single source, 197 from two sources and 26 from 3 sources.

The number of different alleles per marker ranged from 4 to 20. The frequency in which these alleles occur ranged from 0.1% (rare) to 98.0% (common) with an average of 23.6%. The alleles could only be scored qualitatively (present or absent), not quantitatively (number of copies per allele.) These frequencies therefore cannot be interpreted as genuine allele frequencies. The profiles that were scored per marker therefore are called ‘allelic phenotypes’, as they do not represent the ‘allelic genotypes’.

marker	avg, # of diff. alleles per phenotype	# different profiles	# unique profiles	% unique profiles	frequency of most common allelic phenotype	# diff. alleles in most common allelic phenotype	PIC value*
STMS 0019	2.14	61	16	1.8	0.17	2	0.92
STMS 2005	2.56	21	4	0.4	0.37	3	0.80
STMS 2028	2.31	62	20	2.2	0.23	2	0.90
STMS 3009	1.91	48	19	2.1	0.34	2	0.81
STMS 3012	2.25	27	2	0.2	0.19	2	0.87
STMS 3023	2.26	14	1	0.1	0.32	2	0.79
STMS 5136	2.76	54	25	2.8	0.14	3	0.92
STMS 5148	3.14	251	126	13.9	0.05	3	0.98
STMS SSR1	2.81	119	50	5.5	0.17	3	0.93

Table 4. Marker information for potato varieties subjected to SSR analysis during this project on the 2006 EU Common Catalogue. * PIC values based on allelic phenotypes

In Table 4 statistics for the markers are presented. The number of different alleles per marker can range from 1 (all the same) to 4 (all different). The lowest average number of different alleles was found in STMS 3009 (1.91), the highest in STMS 5148 (3.14). STMS 5148 also had the highest number of alleles (Table 4), and therefore not surprisingly showed the highest number of different profiles (allelic phenotypes), and the highest number of unique profiles: 13.9%. These varieties can be distinguished on the basis of this marker alone. STMS SSR1 also showed to be a powerful marker for discrimination between varieties. With 13 alleles it showed 119 different profiles, of which 50 (5.5%) were unique. With a comparable number of 14 alleles, STMS 3009 only showed 48 different profiles, with 19 (2.1%) unique profiles. STMS 3023 was the least discriminating marker of this set, with 4 different alleles, only 0.1% unique profiles and the highest frequency of the most common allelic phenotype (0.32).
The combined effect of number of alleles and different allelic phenotypes is best represented in the PIC (polymorphism information content) value of the markers. In diploid species this is calculated from allele frequencies. As these are not available from our data, the presented PIC values were calculated on the basis of allelic phenotypes:

$PIC_{mark} = 1 - \sum (p_i)^2$, with mark=marker, and p_i = frequency of allelic phenotype per marker

PIC values range from 0 to 1; the closer to 1, the more discriminative the marker is. The PIC values calculated from these data are very similar to values presented for the same markers in earlier studies (Reid and Kerr, 2007).
In Figure 1 the results for each marker are presented as pie-charts: each slice representing an allelic phenotype. The unique profiles are all combined in the last slice (black), as the resolution of the images did not allow for individual representation.

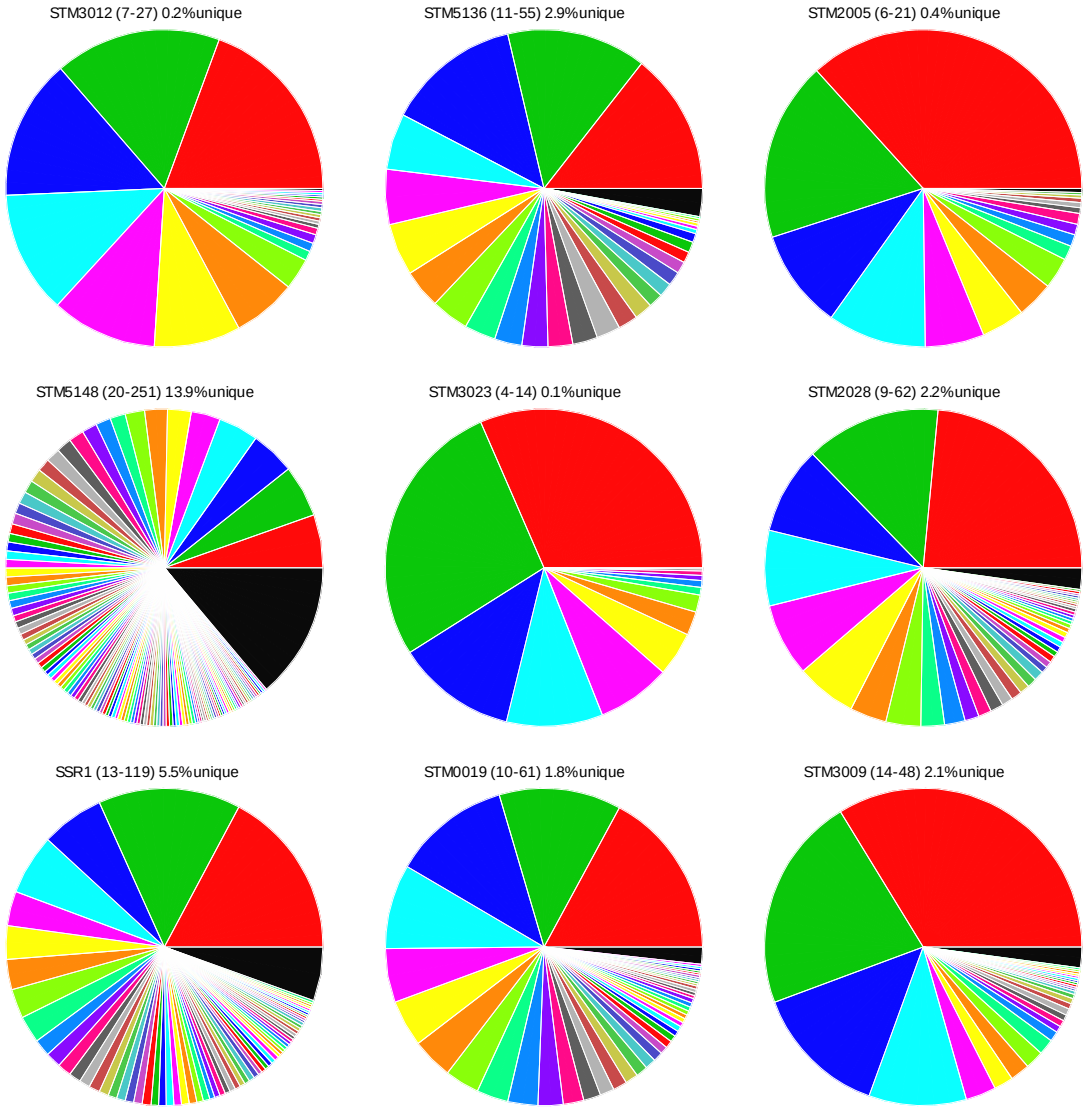

Figure 2. Graphic presentation of the discriminative power of the 9 used markers (STMS 3012, STMS 5136, STMS 2005, STMS 5148, STMS 3023, STMS 2028, STMS SSR1, STMS 0019 and STMS 3009) for the potato varieties subjected to SSR analysis during this project from the EU Common Catalogue. In brackets the number of alleles followed by the number of different profiles. The black slice (top south-east quarter) represents the combined unique profiles.

Using these 9 markers all of the varieties (excluding known mutants) can be differentiated with a few exceptions (indeed it is possible to exclude STMS 0019, STMS 3009 and STMS 3023 and still differentiate all the varieties). The problems fall into one of two categories, firstly, varieties which match and are not known mutants, and secondly, samples with the same variety name that do not match (see section 3B Testing of genetic fingerprints of varieties held in more than one of the partners collections). The problem varieties are listed in Table 5.

Variety(ies)	Problem	Origin of samples	Possible cause
Allerfrüheste Gelbe	NL & UK different profiles	NL & UK	3 NL samples Allerfrüheste Gelbe match Arran Comrade (not in EUCC) from UK collection. Possible mislabel of Allerfrüheste Gelbe and Arran Comrade in UK collection?
Sava	NL & UK different profiles	NL & UK	Neither match any other variety so can't determine which is correct.
Albas Astarte	Identical	NL	Albas likely mutant of Astarte, not seedling as described (varieties morphologically very similar except flower colour).
Naglerner Kipfler[1] Ratte (syn. Asparges)	All identical	Asparges (NL & UK) Naglerner Kipfler (DE & NL) Ratte (NL & UK)	Naglerner Kipfler is not a known mutant.
Bernadette Dali	Identical	Bernadette (NL & UK) Dali (NL & PL)	Not known mutants.
Denar Lord	Identical	Denar (NL & PL) Lord (NL & PL)	Both varieties are described as progeny from the same cross.
Elfe Marabel	Identical	Elfe (DE & NL) Marabel (NL & UK)	Not known mutants.
King David Royal Kidney	Identical	Both UK	Already suspected to be the same variety.
Manna Mistral	Identical	Both UK	Previously suspected variety mix up Mistral probably correct.
Lady Florina Timate	Identical	Lady Florina NL, PL & UK Timate NL	Not known mutants.
Satina	Slight difference	DE & PL	STMS SSR1 allele I appears to be genuine polymorphism.

Table 5. Problem varieties discovered during the project.
[1] listed in Austria. Material from maintainer. No morphological identification check by DE.

Varieties which match and are not known mutants.
Of the list of problem varieties there are only 5 sets that would not have been expected to yield identical profiles. These are Bernadette/Dali, Denar/Lord, Elfe/Marabel, Lady Florina/Timate and Asparges/Ratte/Naglerner Kipfler. In an attempt to differentiate these varieties each was analysed with an additional 31 SSR markers (making a total of 40

markers). In each case the members of the sets remained identical. Material of these five sets has been re-sampled from breeders or official maintainers, put up for light sprouts and planted in the field for side-by-side comparison by the Netherlands in 2008. Prior to planting, molecular identity was checked by means of the 9 markers, leading to identical results. Three other sets of varieties also yielded identical profiles. Albas and Astarte only differ morphologically by flower colour and it is suspected that Albas is a mutant variety of Astarte and not a seedling as previously reported, Royal Kidney and King David were already suspected as being the same variety and Manna and Mistral which were a previously suspected variety mix up.

As only allelic phenotypes are scored, no data on allele frequency can be calculated. The chance of two unrelated profiles showing an identical profile can therefore not be derived from allele frequencies. However, an upper limit can be calculated based on the frequency of the most common allelic phenotype of each marker (mutants and duplicates excluded).
As this dataset does not include duplicates and only few mutants, an estimation of this upper limit is based on the presented frequencies of most common allelic phenotypes (Table 4):
$0.19 \times 0.14 \times \ldots \times 0.34 = 3.6 \times 10^{-7}$. In other words, this chance is at best 1 in 2.8 million ($1/3.6 \times 10^{-7}$). The chance of 2 varieties yielding identical profiles for 40 markers is infinitesimally small.
The 9 markers can be regarded as independent as they are positioned on different linkage groups (chromosomes). However, the assumption of unrelatedness is not strictly true for many varieties, as they may have common ancestors at some point in their pedigree. When related, the chance of identical allelic phenotypes obviously increases. In addition, the selection towards agronomical important characteristics of the superior ancestor may have unknown influences on some allele frequencies.
The theoretical probabilities of the 5 sets of identical varieties yielding identical profiles for just the 9 markers used during the project are calculated at: Bernadette/Dali 1 in 1.56×10^{9}; for Denar/Lord 1 in 3.03×10^{8} (although this is undoubtedly much higher); for Elfe/Marabel 1 in 1.02×10^{12}; for Lady Florina/Timate 1 in 7.71×10^{9} and for Naglerner Kipfler/Asparges 1 in 4.24×10^{13}. As common ancestry is not taken into account, the actual probabilities will be higher, but clearly still extremely small!

The ability of the database to discriminate between varieties can best be shown by calculating the pairwise comparisons of single entries of all varieties. The total number of pairwise comparisons for 900 varieties is:

$$\frac{900^2}{2} - 900 = 404{,}100$$

In Figure 3, a frequency distribution of all pairwise comparisons is presented (based on Jaccard similarity coefficients). The average similarity between varieties with these markers is around 40-45%.

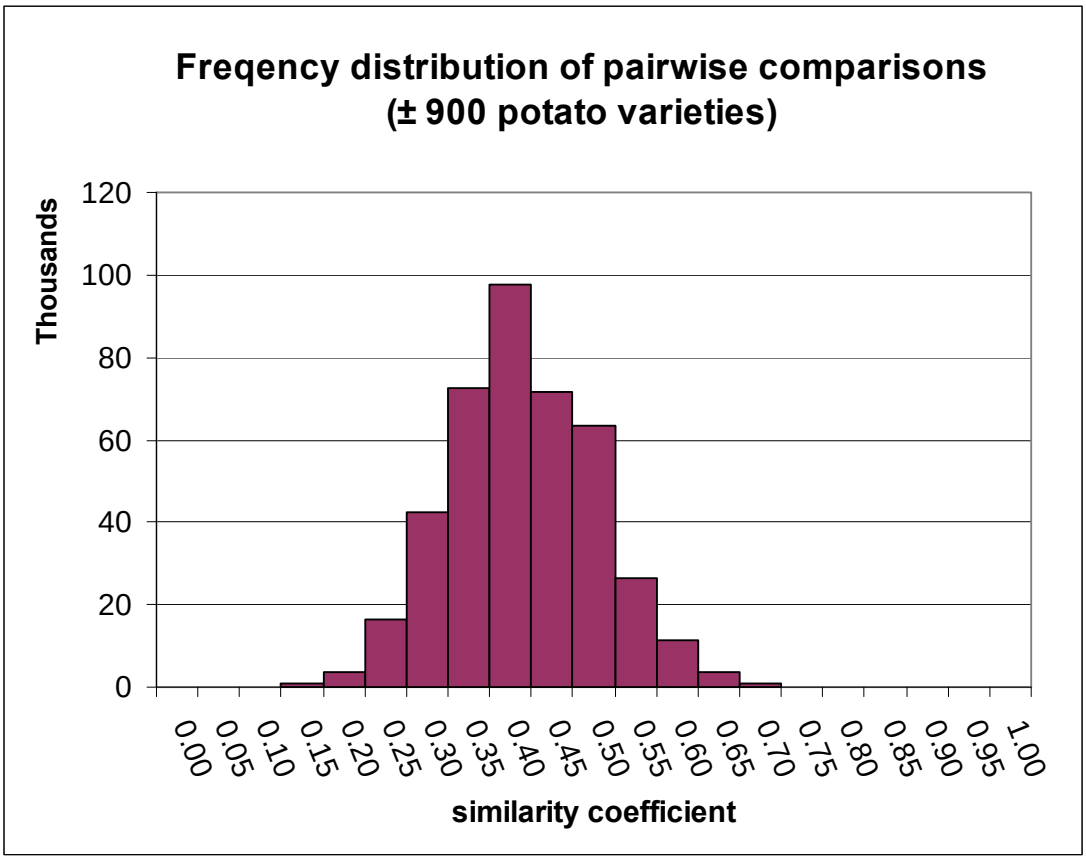

Figure 3. Frequency distribution of pairwise comparisons of ± 900 potato varieties (Jaccard coefficient).

The most interesting part is the upper tail end of this graph, which cannot clearly be seen because of the scale of the y-axis (thousands). Therefore, a close-up of the upper tail end is presented in Figure 4.

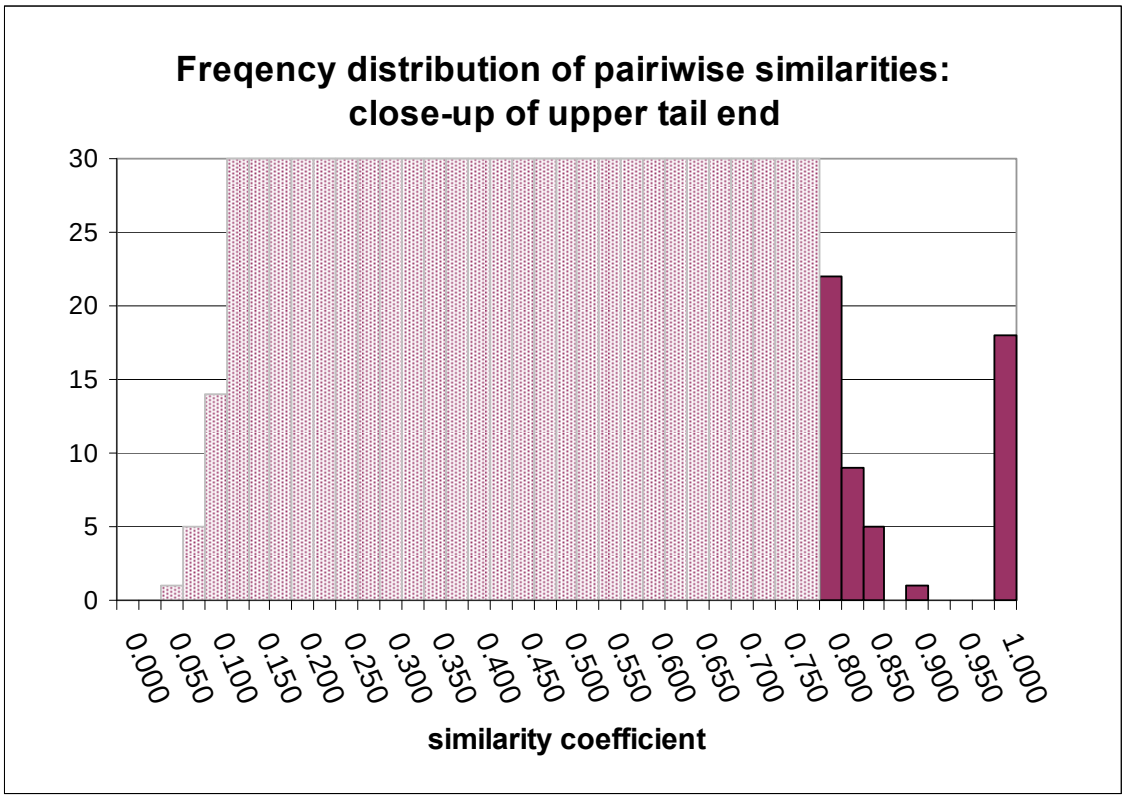

Figure 4. Close-up of upper tail end of frequency distribution of pairwise comparisons of ± 900 potato varieties (Jaccard coefficient).

The bar at similarity coefficient 1.0 represents the group of molecular identical varieties. These include the known mutants (e.g. Cara, Red Cara, Druid and Avondale; Duke of York and Red duke of York) as well as the five sets of varieties mentioned above. The next closest pair (Jaccard similarity 91%) is Nikita/Janine. This similarity represents a difference of 2 alleles. The high similarity can be explained by ancestry: Janine results from a cross between Nikita and Obelix, which is supported by comparison of the three profiles. Apparently for Janine the number of marker alleles (allelic phenotype) inherited from Nikita is somewhat higher than the number of marker alleles inherited from Obelix.
The next bar represents a group of 5 variety pairs with a similarity between 85 and 87.5%. This is equivalent to a difference of 3 alleles. One of these pairs (Pentland Ivory/Pentland Dell) is related (Pentland Ivory has Pentland Dell as one of its parents), but the other four pairs do not seem to have clear common ancestry.
If mutants and and other identical pairs are not taken into account, only 0.0015% (6 variety pairs) of all possible comparisons show a similarity of 85-91%, equivalent to a difference of 2-3 alleles. This means 99.9985% of all pairwise comparisons between different varieties have a similarity lower than 85%, and at least 4 alleles difference.

3B) Testing of genetic fingerprints of varieties held in more than one of the partners collections.

The database contains 223 varieties submitted for DNA analysis from more than one partner (197 from 2 sources and 26 from 3). Of these the majority of samples match each other exactly with all nine markers. The only exceptions are Allerfrüheste Gelbe and Sava, where

the UK and Netherlands samples are so different they are obviously different varieties one of which must have been mislabelled. In the case of Allerfrüheste Gelbe the samples from the Netherlands match *cv*. Arran Comrade (not in the CC) from the UK collection and it is therefore believed that the Netherlands samples are the actual Allerfrüheste Gelbe and this problem has arisen in the UK collection due to a mix up of Arran Comrade and Allerfrüheste Gelbe although this is not confirmed and both samples remain in the database as a result. Neither samples of Sava match any other variety held in other databases in the UK and it remains unclear which is the 'correct' Sava.
A few other varieties display slight differences between profiles obtained in the Netherlands and the UK. These are mostly due to the G allele in STMS 0019, a known problem with this marker as the peak for this allele can often be small if other alleles are present (due to competition during PCR). In all cases where the only difference between two samples is STMS 0019 G, the samples are deemed to belong to the same variety.
One variety that would appear to have a genuine polymorphism is *cv.* Satina where samples from Germany and Poland consistently differ by a single allele with marker SSR1 (the Polish sample has alleles BDFI and the German sample BDF). This is the only variety which yields such a polymorphism.
During the project, we resolved 21 cases of wrongly labeled samples which showed clear differences between one partners collection and another. At this point it is not known where these errors arose and, within the scope of this project, this does not really matter. The important message is that using this technology it is possible to highlight these errors. For potato, it therefore proves to be advisable to collect samples from only verified sources when entering molecular data in the database.

4) Database set up

A database has been constructed in MS Access containing 11 tables and has been designed as such so that it can serve the BioNumerics software package which is used to perform identifications. In the Access database there are tables for each of the nine markers, a table for the morphological data and a table containing general information for each sample in the database. As BioNumerics requires a unique identifier for every entry and as some varieties had multiple samples the DNA sample code has been used for the SSR data (in the UK we tried to analyse every variety as duplicate samples from separate tubers). Entries for the morphological data have only one entry for each country and here the unique identifier is the variety name and the country code where the description was carried out. For example 3 samples of *cv.* Ditta were submitted for SSR analysis from the Netherlands, Poland and the UK and the unique identifier for these samples are NL-089, PL-021 and UK-0598 respectively. All 4 countries submitted a morphological description for Ditta and the unique identifiers are Ditta_DE, Ditta_NL, Ditta_PL and Ditta_UK.

The information table contains details for each sample in the database (fields and explanation are given in Table 6.

Field	Explanation
Key	The unique identifier for use in BioNumerics
Variety Denomination	The name of the variety
Origin of sample e.g. Breeder (B), office (O) or other (T)	Where the sample was obtained from either the breeder or from one of the partners collections or from another source
Submitting office (S SASA, B BSA, C COBORU, N Naktuinbouw)	The office which submitted the sample for DNA analysis or submitted the morphological description
Harvest year	The year the sample analyzed was harvested
DNA extraction laboratory (S SASA, N Naktuinbouw)	The laboratory the DNA sample was extracted (either UK or Netherlands)
Extraction year	The year the DNA extraction was made
Place of storage of DNA sample (S SASA, N Naktuinbouw)	The place where the DNA sample is kept in long term storage
SSR analysis performed at (S SASA, N Naktuinbouw)	The laboratory where the SSR analysis was performed
SSR analysis year	The year the SSR analysis was performed
Technical protocol used for morphological description	The technical protocol used for the morphological description (if known)
Description year (either before 1995 or actual year)	The year that the latest description was carried out, N.B. not necessarily from official description
Place description carried out	The office the description was carried out at
Photograph availability (and link to photograph)	The file name for the photograph (if available) and hyperlink
Place photograph taken (office)	The office the photograph was taken at
Photograph year (if known)	The year the photograph was taken
National Reference	The national reference number of the sample (if there is one)
Comments	Any other comments (e.g. is the variety a mutant of another variety)

Table 6. Explanations of database fields

The data held in the Access database is linked via an ODBC (Open DataBase Connectivity) to BioNumerics allowing the BioNumerics database to be updated directly from Access. Two libraries have been constructed for all of the varieties obtained during the project on the 24th edition of the EU Common Catalogue. One library contains the data from the morphological descriptions and has (for the varieties with more than one description) multiple entries for each variety. For example all 4 descriptions of Ditta are included in the morphological library thus theoretically allowing a positive identification of an unknown sample from any country. It should be noted that BioNumerics does not impose a penalty for missing data when performing an identification against a library unit therefore if only the lightsprout characters are available for an unknown sample then any data from the plant or tuber characteristics present in the library example are ignored during the identification process. The second library contains the SSR data and in this case each library unit has only a single example except for the problem varieties in Table 5 where all of the entries are included in the library. This will allow an identification to be made of an unknown sample of one of these varieties if

it only matches one of the examples in the library. For example if a subsequent sample of Satina is screened against the library it will be identified as Satina whether it has the SSR1 I allele or not.

5) Validation of the SSR data in the database by a blind test.

Twenty varieties were submitted for blind testing, ten from Poland and ten from Germany. The results are shown In Table 7.

Germany			
Sample No. (DNA code No.)	BioNumerics Identification	Variety submitted	Next closest match (% similarity)
1 (D-207)	Kuras	Kuras	Amado (64.3)
2 (D-208)	Frühgold	Frühgold	Juwel (83.3)
3 (D-209)	Eldena	Eldena	Marena (70.4)
4 (D-210)	Delikat	Delikat	Acapella (75.0)
5 (D-211)	Charlotte	Charlotte	Manuela (73.1)
6 (D-212)	Russet Burbank	Russet Burbank	Early Puritan (74.1)
7 (D-213)	Terrana	Terrana	Turdus (65.4)
8 (D-214)	Solara	Solara	Exquisa (70.4)
9 (D-215)	Priamos	Priamos	Orla (61.1)
10 (D-216)	Pirol	Pirol	Umiak (59.3)
Poland			
1 (PL-205)	Denar/Lord	Denar	Czapla (80.0)
2(PL-206)	Denar/Lord	Lord	Czapla (80.0)
3 (PL-207)	Bard	Bard	Zebra (73.9)
4 (PL-208)	Ruta	Ruta	Gorbea (66.7)
5 (PL-209)	Rumpel	Rumpel	Pasat (64.5)
6 (PL-210)	Kolia	Kolia	Andante (63.0)
7 (PL-211)	Dorota	Dorota	Clarissa (69.2)
8 (PL-212)	Korona	Korona	Orlik (77.8)
9 (PL-213)	Bila	Bila	Lady Olympia (65.5)
10 (PL-214)	Orlik	Orlik	Acapella (79.3)

Table 7. The results of the blind test samples showing the identification obtained from interrogation of the BioNumerics SSR data library, the actual variety and the variety which was the next closest match and the % similarity (Jaccard coefficient).

All varieties were blind tested in the UK as well as the Netherlands, with identical results. Of the 20 varieties submitted for blind testing 18 were unequivocally identified (a 100% match) by interrogation of the BioNumerics library. The only exceptions were samples of Denar and Lord submitted by Poland which are in the list of problem varieties (Table 5). As a measure of how well the system works the most similar 2nd place match (excluding Denar and Lord) was Juwel which yielded an 83.3% similarity to Frühgold.

6) Exchange of data between the partners.

At the final project meeting held in Germany in February 2008 it was agreed by the partners that at this stage there was no need for the BioNumerics database to be accessible via the

internet. Instead the database is to be distributed amongst the partners on CD. It will include the full version of the Access database. The UK and the Netherlands will also exchange the BioNumerics files (including the Libraries).

7) Suggestions for implementation of this technology in future DUS testing.

The molecular description of nearly all varieties listed and protected in Europe has shown that the vast majority of SSR profiles are unique and specific for one variety. Only mutants and a few pairs of varieties yielded the same SSR profile. These pairs of varieties all have very similar morphological descriptions. The analyses of duplicated samples of the same varieties from different sources did not expose any major problem with stability of the molecular profile. The SSR system was harmonized successfully in a way that both laboratories carrying out the molecular analysis came to identical descriptions for all samples analyzed in both places. The technique can be transferred to other laboratories if needed. Results from different laboratories can be combined in the database if the same SSR system is applied.

The results of this project open the possibility to use the analyzed SSR markers for two main applications in relation to DUS testing.

(a) Variety identification – Quality management for the DUS system

Almost all varieties (99.5%, excluding mutants) have unique molecular profiles. Thus, this technique can be used as an efficient additional tool for variety identification. Any mix up of varieties can be identified immediately. The identity of the molecular profile of a new sample gives a high probability that the material represents the variety, notwithstanding that the definite identity of a sample can only be proven by a morphological identity check because morphological instability is unlikely to be seen in the molecular profile using these markers.

The molecular profile can give an important contribution in the quality management of the DUS testing system:

- Reference varieties in DUS test need to be newly planted (and, if no live reference collection is maintained, collected) each year. In any case it is necessary to validate the variety identity of a new sample of a candidate or reference variety which is used for DUS purposes. It has to be proven that the variety was maintained and submitted correctly: there was no change in the expression of characteristics and there was no mix up of varieties. If all newly submitted samples were to undergo molecular profiling nearly all cases of mislabelling and variety mix up could be identified before the material will be included in the trial or in the living variety collection.

- During the course of this project several problems have been identified in currently maintained living variety collections by molecular profiling. 21 samples were found to have been mislabelled. Through the use of molecular markers these were brought to light and solved.
  Other mistakes were linked to the reuse of variety denominations. The collection contained samples of varieties which have been deleted several years ago, including candidate varieties which have never been released. In some cases this caused confusion when the same variety denomination was reused for a new variety. The problem cases have been identified by molecular markers (as well as morphological

descriptions in the case of Bonanza). In future such cases can be prevented by linking relevant administrative data to the living plant material in the variety collection.

(b) Management of the reference collection

The identification of existing varieties which are similar to a candidate variety is a decisive challenge for the DUS expert. The Common Catalogue alone comprises more than 1000 potato varieties which cannot be grown completely in the DUS trials. Because of the high polymorphism of SSR markers in potato it can be expected that varieties with identical molecular profiles are also similar in the morphological characteristics. Through molecular profiling of all candidate varieties prior/during the DUS test and screening of the candidates profile against the database such variety pairs can be identified and it can be ensured that a side-by-side comparison will be carried out in the same growing trial to establish distinctness or prove identity.

While, at present, distinctness cannot be based on molecular information alone in the course of this project it has been shown that the use of the SSR marker system provides reliable, stable and repeatable molecular variety descriptions. This is an important precondition for the development of a system to combine molecular and phenotypic information for the establishment of distinctness.
The morphological descriptions and lightsprout photographs contained in the database are not yet appropriate for the identification of similar varieties, in particular if descriptions have been produced in different locations and different years. If possibilities of harmonized variety descriptions are to be explored it is important to realize that stability of variety descriptions over years and locations needs to be analyzed further, and an appropriate morphological dataset needs to be developed

If a molecular system will be implemented in the DUS test in any of the described ways, it is recommend that DNA samples should be extracted from the identity material (submitted for DUS) and stored at two separate locations.

Other Outputs

The setup of the BioNumerics database for SSR data has been reported at the BMT meeting in November 2006 (see document BMT/10/5 and BMT/10/9) and the BMT-TWA Subgroup meeting for potato in April 2007 (see document BMT-TWA/Potato/2/2). Final results of this project will be presented at the BMT/11 meeting, September 2008.
A peer reviewed methods chapter is to be published towards the end of 2008 (or early in 2009) detailing the method developed for this project (Reid *et al.*, 2008).

References

BMT/10/5 (2006) Identification of potato cultivars on the European Union Common Catalogue using Simple Sequence Repeat (SSR) markers.

BMT/10/9 (2006) The Use of a BioNumerics database for the rapid identification of potato cultivars.

BMT-TWA/Potato/2/2 (2007) Identification of potato cultivars on the European Union Common Catalogue using Simple Sequence Repeat (SSR) markers.

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.

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.

Milbourne, D., Meyer, R.C., Collins, A.J., Ramsey, L.D., Gebhardt, C. and Waugh, R. (1998) Isolation, characterisation and mapping of simple sequence repeat loci in potato. *Molecular and General Genetics*, **259**, 233-245.

Reid, A., Hof, L., Esselink, D. and Vosman, B. (2008) Potato cultivar genome analysis. In: Plant Pathology Techniques and Protocols (Ed. Burns, R.) Humana Press.

Reid, A. and E.M. Kerr (2007) A rapid simple sequence repeat (SSR)-based identification method for potato cultivars. *Plant Genetic Resources: Characterization and Utilization* **5**, 7-1