E

BMT/11/9 Rev.
ORIGINAL: English
DATE: August 28, 2008

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA PROFILING IN PARTICULAR

Eleventh Session
Madrid, September 16 to 18, 2008

CONSTRUCTION OF AN INTEGRATED MICROSATELLITE AND KEY MORPHOLOGICAL CHARACTERISTIC DATABASE OF POTATO VARIETIES ON THE EU COMMON CATALOGUE
PART 1: DISCUSSION OF MORPHOLOGICAL AND MOLECULAR DATA (REVISED)

Document prepared by experts from the Netherlands

CONSTRUCTION OF AN INTEGRATED MICROSATELLITE AND KEY MORPHOLOGICAL CHARACTERISTIC DATABASE OF POTATO VARIETIES ON THE EU COMMON CATALOGUE
PART 1: DISCUSSION OF MORPHOLOGICAL AND MOLECULAR DATA (REVISED)

Dr Ir Lysbeth Hof[1)] & Dr Alex Reid[2)]

[1)]Naktuinbouw, Wageningen/Roelofarendsveen, NL
[2)]SASA, Edinburgh, UK

INTRODUCTION

1. Each year, in Europe alone more than 100 applications for registration and/or Plant Variety Rights of potato varieties reach the examination offices. Apart from national applications, four DUS testing stations are responsible for the vast majority of applications for European Plant Breeder's Rights: the Bundessortenamt in Germany, Naktuinbouw in the Netherlands, COBORU in Poland and SASA in the United Kingdom.

2. For assessment of distinctness, new potato applications need to be compared with all other varieties of common knowledge. Selection of similar varieties for side-by-side comparison in the field is a crucial step in this procedure. To date, there is no exchange of variety descriptions between examination offices because expression of most of the currently used characteristics is influenced by environmental factors and may also be subject to interpretation differences between observers.

3. Maintenance of live reference collections of potato for DUS testing is becoming impractical, as it is a vegetatively propagated crop and the European Union Common Catalogue (EUCC) alone already comprises over 1,100 varieties. As a consequence, reference varieties are increasingly obtained direct from the breeders and a quick method for verification of identity would be very valuable. A possible solution to the problems mentioned above would be the establishment of an international database, comprising all potato varieties of the EUCC, characterized by molecular markers. Microsatellite markers (also known as simple sequence repeats or SSRs) are very suitable for this purpose, as they are highly discriminatory, can be reproducibly obtained in different laboratories, their analysis can be automated and the results can be stored easily.

4. The four testing stations (BSA, Naktuinbouw, SASA and COBORU) therefore joined forces, and with co-financing by the Community Plant Variety Office of the European Community (CPVO) initiated a 2-year project to construct an integrated microsatellite and key morphological characteristic database of potato varieties. In order to keep it manageable, it was decided to restrict the database to varieties of the European Common Catalogue. The project finished in April 2008.

5. This paper gives a brief overview of the results from this project with emphasis on discussion of the morphological and molecular data. Aspects concerning the construction of the actual database will be discussed in another paper.

Materials and methods

6. The project started in April 2006, using the 24th edition of the EUCC as a point of reference for the database. Of the 1,104 entries in the EUCC, 900 varieties were collected, with over 200 varieties from more than one source. Many varieties were retrieved from the respective collections of the partners, and samples sent to the laboratories in the United Kingdom and the Netherlands for DNA extraction. Morphological data of characteristics from the Technical Questionnaire (TQ) and light sprouts were retrieved from databases of the partners for as many varieties as possible, as well as light sprout pictures of good technical quality.

7. Varieties not available from own collections were requested from breeders/official maintainers. Of these newly collected varieties light sprout descriptions were made, pictures of the light sprouts taken and samples sent to both labs for DNA extraction and genotyping.

8. Genotyping of all samples was performed with 9 microsatellite markers which prior to this project were selected by the United Kingdom and the Netherlands as being suitable for variety characterization in potato. The 9 markers were analyzed in 3 multiplexes of 3 markers. DNA analyses of most samples were carried out at both labs, to ensure that the same results were obtained when using different equipment. Both labs use an ABI DNA sequencer (Applied Biosystems), though a different type. In case of doubt on interpretation of molecular profiles, characterization was repeated and/or new samples were collected. For simplicity, these duplicates as well as identical profiles from both labs were removed from the final database.

Morphological data

9. In the course of the project morphological data for 733 varieties were collected. For some varieties, descriptions from more than one country were received, adding up to a total of 856 descriptions: 622 from a single country, 99 from 2 countries, 11 from 3 countries and 1 from all 4 countries. Light sprout photographs were provided for 377 varieties (of which 28 varieties with photographs from 2 countries).

10. The collected morphological data included descriptions of all light sprout characteristics as well as the characteristics required in the UPOV Test Guidelines for potato (document TG/23/6). A list of the characteristics is shown in Table 1.

Table 1. Morphological characteristics included in variety descriptions

Char. number (TG/23/6)	Character description	Type of expression*	Notes
1	Light sprout: size	QN	1-9
2	Light sprout: shape	PQ	1-5
3	Light sprout: intensity of anthocyanin coloration of base	QN	1-9
4	Light sprout: proportion of blue in anthocyanin coloration of base	QN	1-3
5	Light sprout: pubescence of base	QN	1-9
6	Light sprout: size of tip in relation to base	QN	1-9
7	Light sprout: habit of tip	QN	1-9
8	Light sprout: anthocyanin coloration of tip	QN	1-9

9	Light sprout: pubescence of tip	QN	1-9
10	Light sprout: number of root tips	QN	1-9
11	Light sprout: length of lateral shoots	QN	1-9
29	Plant: frequency of flowers	QN	1-9
33	Flower corolla: intensity of anthocyanin coloration on inner side	QN	1-9
34	Flower corolla: proportion of blue in anthocyanin coloration on inner side	QN	1-3
36	Plant: time of maturity	QN	1-9
37	Tuber: shape	QN	1-6
39	Tuber: color of skin	PQ	1-7
40	Tuber: color of base of eye	PQ	1-4
41	Tuber: color of flesh	PQ	1-9

*QN= quantitative, PQ= pseudo-qualitative

11. For the varieties requested from breeders/maintainers, only light sprout data were available. Most of the descriptions from the United Kingdom were based on the previous version of the UPOV Test Guidelines (TG/23/5), whilst the other countries supplied data based on TG/23/6. The Test Guidelines differ in scoring for characteristics 4, 7, 34, 39, 40 and 41. When comparing descriptions from different countries, the United Kingdom descriptions were therefore excluded for those characteristics.

12. For 12 varieties, morphological descriptions of 3 or more countries were available. For each of these varieties, for all quantitative characteristics the difference between the highest and lowest score of the different descriptions was calculated. From a total of 174 comparisons, 17.2% of the scores was identical, 37.4% showed 1 note difference, 27.0% 2 notes difference, 13.2% 3 notes difference, 3.4% 4 notes difference, 1.1% 5 notes difference and 0.6% (a single incident) 6 notes difference.

13. Of the pseudo-qualitative characteristics, tuber skin and eye color was found to be stable: for all 12 varieties no differences between scores amongst countries were found. For light sprout shape 5 varieties scored the same in all countries. For tuber flesh color, only 4 varieties had the same score, but the differences in the other 8 varieties were small; all were scored having yellow flesh, but with some variation in intensity of the yellow color.

14. In almost all cases the various descriptions of the same variety would have been considered to be distinct, whilst in reality should have been similar. An exchange of variety descriptions between DUS testing stations therefore does not seem to be useful, and applications and similar varieties should be compared side-by-side in the same trial field. These results support the conclusions of a project carried out in the framework of a UPOV study to consider the publication of variety descriptions, presented at the thirty-fourth Session of the Technical Working Party for Agricultural Crops, held in 2005 (see document TWA/34/13 Add. 2).

Marker information and molecular data

15. Details of the nine markers used in the project are given in Table 2. 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'.

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

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

16. As the actual size of the alleles can slightly vary between scoring platforms, it was decided that for reference each defined allele was designated with a letter (A, B,…) rather than its product size. In the analyses a set of reference varieties was always included to ensure calibration of allele sizes.

17. In Table 3, statistics for the markers as found in this project are presented. Potato is a tetraploid species; for each variety the number of different alleles for each marker theoretically can range from 1 (all the same) to 4 (all different). The lowest average number of different alleles was found in MS 3009 (1.91), the highest in MS 5148 (3.14). MS 5148 also had the highest number of alleles (Table 4), and therefore not surprisingly showed the highest number of different profiles, and the highest number of unique profiles *viz*. 13.9%. Varieties in this category can be distinguished on the basis of this marker alone. MS SSR1 was also found 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, MS 3009 only showed 48 different profiles, with 19 (2.1%) unique profiles. MS 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).

Table 3. Marker results based on ± 900 potato varieties of the EU Common Catalogue 2006.

Marker	Number of alleles	Average number of different alleles per phenotype	Number of different profiles	Number of unique profiles	% Unique profiles	Frequency of most common allelic phenotype	PIC value*
MS 0019	10	2.14	61	16	1.8	0.17	0.92
MS 2005	6	2.56	21	4	0.4	0.37	0.80
MS 2028	9	2.31	62	20	2.2	0.23	0.90
MS 3009	14	1.91	48	19	2.1	0.34	0.81
MS 3012	7	2.25	27	2	0.2	0.19	0.87
MS 3023	4	2.26	14	1	0.1	0.32	0.79
MS 5136	11	2.76	54	25	2.8	0.14	0.92

MS 5148	20	3.14	251	126	13.9	0.05	0.98
MS SSR1	14	2.81	119	50	5.5	0.17	0.93

* PIC values based on allelic phenotypes

18. 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

19. 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).

20. Using these 9 markers, all of the varieties (excluding known mutants) can be differentiated with a few exceptions. In two cases, samples with the same variety name showed different molecular profiles. One of those samples was probably mislabelled but during the course of this project the correct sample could not be identified.

21. Eight pairs of varieties showed an identical profile. Three of these cases could be explained: one pair was already suspected of being identical, one pair was already suspected of being mislabeled, and for one pair, one variety was already suspected of being a mutant of the other and not a seedling as reported. Another 5 pairs could not yet be explained. One of these pairs concerns two very old varieties (respectively >100 and >50 years old), of which no ancestry is known. For another pair, both varieties are derived from the same cross. Morphological descriptions of all of these pairs are very similar. In an attempt to differentiate these varieties, each was analyzed with an additional 31 markers (making a total of 40 markers), but in all cases the members of the sets remained identical. Material of these varieties has been resampled from official maintainers, but the results were the same.

22. 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 a few mutants, an estimation of this upper limit is based on the presented frequencies of most common allelic phenotypes (see Table 3): 0.17 x 0.37 x … x 0.17 = 3.6×10^{-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.

23. 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.

24. One variety appeared to have a genuine polymorphism. For this variety, samples from Germany and Poland consistently differed by a single allele of marker SSR1 (the Polish sample has alleles BDFI and the German sample BDF). This is the only variety which yields such a polymorphism.

25. The ability of the database to discriminate between varieties can best be shown by calculating the pair wise comparisons of single entries of all varieties. The total number of pair-wise comparisons for 900 varieties is 404,100 ($900^2/2 - 900$). The average Jaccard similarity between varieties with these markers was around 40-45%. Eighteen pairs of varieties showed a similarity coefficient of 1 (100%). These include known mutants as well as the unexplained sets of matching varieties as mentioned above.

26. The next closest pair had a Jaccard similarity of 91%. This similarity value represents a difference of 2 alleles. The high value can partially be explained by ancestry: one variety results from a cross between the other with a third variety, which is supported by comparison of the three respective profiles. Apparently the number of marker alleles inherited from one parent is somewhat higher than the number of marker alleles inherited from the other parent. (Remember that allelic phenotypes are scored, not allelic genotypes.)

27. The following group represents five variety pairs with a similarity value between 85 and 87.5%. This is equivalent to a difference of 3 alleles. One of these pairs is related, but the other four pairs do not seem to have clear common ancestry.

28. If mutants and other identical pairs are not taken into account, only 0.0015% (6 pairs: 6/404,100 * 100%) of all possible comparisons show a similarity of 85-91%, equivalent to a difference of 2-3 alleles. This means 99.9985% of all pair wise comparisons between different varieties have a similarity lower than 85%, and at least 4 alleles difference.

29. It is extremely unlikely that, with these markers, mutants can be discriminated from their original variety. Mutants may result from a single point mutation and the chances of this mutation being involved in one of the microsatellites is remotely small. On the other hand, when two independent samples show an identical molecular profile, the results indicate that the chance of these samples being genetically identical or mutants is very high.

30. Comparison of profiles from both laboratories gave a good indication of the scorability of the markers/alleles. Initially a substantial number of varieties with one or two different allele calls was found. All differences were discussed separately. For some markers with large differences in size of the alleles, particularly the larger alleles amplified rather poorly resulting in small peaks which sometimes were just below the set detection level of 15% surface of the largest peak. During the project the interpretation of the peak profiles at both labs was fine-tuned, leading to far less differences to be discussed towards the end. Some markers or alleles continuously led to discussion: e.g. the G allele in MS 0019 (often too small to be called in the Netherlands) and the A allele in MS 3023 (difficult to distinguish between peak at A being genuine allele or stutter of B). However, after discussion of results in almost all cases consensus could be reached on the profile to be entered in the database. These small differences in scoring generally have no impact with regard to the use of the database for variety identification as differences between varieties largely exceed the differences caused by interpretation of alleles. However, when constructing a database analyses at two labs greatly enhances its robustness.

31. During the project, 21 cases of wrongly labeled samples were resolved. 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 only from verified sources when entering molecular data in the database.

32. Some confusion was also caused by reuse of variety names. It is possible in some collections to encounter varieties which have been deleted from the EUCC several years ago. National legislation in many countries allows the reuse of denominations for new varieties after a certain period of time. Also, potato varieties sometimes are very long-lived (>100 years), and date from far before European denomination regulations or even National variety registration. In the course of this project several cases of samples with identical names but different profiles were encountered. Apparently some were caused by mislabeling of the samples as one of the profiles matched another known variety. In other cases, however, neither profile matched any other variety. All but two of these cases were solved by searching for pedigree information of these varieties; it soon became clear that more than one variety with that name (had) existed. The use of the Wageningen Potato Pedigree Database (*www.plantbreeding.wur.nl/potatopedigree*) and the European Cultivated Potato Database (*www.europotato.org*) proved to be very helpful.

Blind test

33. Twenty varieties were submitted for blind testing, ten provided by Poland and ten by Germany. All varieties were blind tested in the United Kingdom 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 database. The only exceptions were two samples submitted by Poland that could not be separated. These varieties were already in the list of pairs with identical profiles that could not be explained (see above). As a measure of how well the system works the next closest match from the database yielded an 83.3% similarity between varieties.

Conclusions

34. Some 900 potato varieties of the European Common Catalogue (EUCC) have been collected and characterized by 9 microsatellite markers. Morphological descriptions of 654 varieties and light sprout pictures of 377 varieties were collected and stored in a BioNumerics (Applied Maths) database.

35. Morphological variety descriptions from different origins showed considerable variation. Exchange of these descriptions between countries is therefore considered not useful for purposes of selection of similar varieties for DUS testing and extension of digital reference collections (description database).

36. Morphological instability of varieties is unlikely to be seen in the molecular profile using these markers.

37. Almost all varieties (99.5%), except mutants, have unique molecular profiles. The unexplained matching pairs of varieties have very similar morphological descriptions. Average similarity between varieties was low (40-45%) and PIC values of most markers was high, underlining the discriminative power of the system.

38. Jaccard similarity of the closest variety pair (except mutants) was 91%, representing a difference of 2 alleles. These varieties were related as one of the varieties had the other as one of it parents. The closest pair of varieties not having a clear common ancestry showed a Jaccard similarity of 86%, representing a difference of 3 alleles. In this dataset more than 99.99% of all pair wise comparisons had a similarity lower than 85%, and at least 4 alleles difference.

39. Scoring of results in two laboratories led to identical descriptions for almost all varieties. In the case of differences, these usually referred to different interpretation of presence of one allele. MS 0019 and MS 3023 most often led to differences in scoring. Even without these markers all varieties can be differentiated.

Suggestion for use of microsatellite markers in future DUS testing of potatoes

40. Because of the highly discriminative power of the used microsatellite markers, this technique can be used as an efficient tool for variety identification.
Reference varieties in potato DUS tests need to be newly planted (and, if no live reference collection is maintained, collected) each year. Molecular identification prior to planting could prevent the use of mislabeled or mixed-up material. In the course of this project 21 cases of mislabeling/variety mix-ups have been brought to light and solved (≈ 1% of all samples).

41. Morphological descriptions have shown to be insufficiently stable for exchange between countries, and foreign descriptions therefore are unreliable for selection of most similar varieties to be included in the DUS trial. Microsatellite profiles have shown to be sufficiently stable across laboratories, and can be exchanged easily. Molecular profiling of all candidate varieties prior to or during DUS test and screening of these profiles against the database would allow identification of genetically closely related varieties, which subsequently could be included in the growing trial for establishing distinctness in addition to reference varieties selected on the basis of other information .

42. As microsatellite markers have shown to be reliable, repeatable and stable, they could serve as a basis to develop a system to combine molecular and morphological/phenotypic information for the establishment of distinctness.

43. If a molecular system is implemented in the DUS test in any way, it is recommended that DNA samples should be extracted from the identity material (submitted for DUS) and stored at two separate locations for reasons of security.

44. When molecular profiles of varieties are entered in a database, it is highly recommended that samples are analyzed in duplicate to minimize the risk of mis-scoring of alleles. In case of doubt samples should be re-analyzed. Using independent samples from different, verified sources decreases the risk of entering profiles with an incorrect variety name.

Acknowledgements

45. Funding of this project was provided by the CPVO, The Dutch Ministry of Agriculture, Nature and Food Quality (Netherlands), Naktuinbouw (Netherlands), SASA (United Kingdom), the Bundessortenamt (Germany) and COBORU (Poland). The microsatellite profiles for the Netherlands were provided by Plant Research International, Wageningen (Netherlands).

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, 7-13.

TWA/34/13, addendum 2 (2005) Project to consider the publication of variety descriptions: potato.

[End of document]