E

BMT/13/10
ORIGINAL: English
DATE: November 8, 2011

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Thirteenth Session
Brasilia, November 22 to 24, 2011

COMBINING MORPHOLOGICAL AND MOLECULAR DISTANCE IN THE MANAGEMENT OF THE REFERENCE COLLECTION OF POTATO

Document prepared by experts from the Netherlands

INTRODUCTION

1. DUS testing of new potato varieties is relatively easy. Most candidate varieties are distinct from existing varieties because *Solanum tuberosum* is a tetraploid species which is extremely heterozygous by its vegetative nature of propagation. As a result of crossing, most seedlings are unique. More than 80% of the Dutch variety collection have unique profiles on the basis of only 11 lightsprout characteristics. Moreover uniformity is usually no problem as seedlings are vegetatively propagated.

2. Although DUS testing of potato varieties is easy in itself, there are a number of limitations. Worldwide, there are over 4000 varieties of common knowledge. The European Common Catalogue comprises about 1400 varieties, whereas the Dutch operational (living) reference collection consists of only 350 varieties. Maintaining a living reference collection is expensive and technically complicated. Potato tubers need to be renewed annually, free from diseases with constant high standards of quality and vitality. Quarantine regulations as well as the bulkiness of the samples are limiting factors for exchange and transport of new entries.

3. In addition to living plant material, the reference collection of potato comprises databases of variety descriptions and photographs of lightsprout and tuber characteristics. For

reasons of data consistency, these databases are produced nationally. Therefore, these collections are limited for the same reasons as mentioned before on the living reference collection. To date there is limited exchange of variety descriptions between testing stations because expressions of characteristics are influenced by environmental factors such as different origins and quality of the plant material, year- and location effects and last but not least, different interpretations of characteristics among observers.

4. Recently, a microsatellite database of potato varieties included in the European Common Catalogue has been developed in a cooperative project with the United Kingdom, the Netherlands, Germany and Poland (co-financed by the Community Plant Variety Office of the European Union (CPVO) and reported at BMT/11, Madrid, 2008). The database is operational in the United Kingdom and in the Netherlands for identification purposes. New applications of potato varieties in the Netherlands have been added since that time, resulting in validated SSR-profiles of about 1200 varieties in the database.

5. Considering the restrictions on the size, the maintenance and the quality of the different components of the reference collection there are reasons for improvement. The microsatellite database offers new opportunities for the management of the reference collection of potato.

MATERIAL AND METHODS

6. Descriptions of 183 varieties were validated in 2010 by testing these varieties for lightsprout characteristics as well as field and tuber characteristics based on the CPVO protocol TP/23/2 (which is based on the UPOV Test Guidelines, document TG/23/6, excluding characteristic 20, 23, 24, 25 and 26). Observations were done by an experienced DUS crop expert. 183 varieties provide ((183 x 182)/2)16,653 variety pairs.

7. Estimating morphological distances between variety pairs by one visual expert-score was considered inaccurate for potato due to the structure of the plant, different assessments at different stages, characteristics dominating the scores and different interpretations between experts. Instead, morphological distances between variety pairs were calculated based on individual characteristics by use of the FSIMILARITY directive of the statistical package GenStat. Similarities between varieties i and j were calculated according to the formula: $\sum_k \{ w_k(x_{ik}, x_{jk})\, s_k(x_{ik}, x_{jk}) \} / \sum_k w_k(x_{ik}, x_{jk})$, where x_{ik} and x_{jk} are the values of the characteristic k of variety i and j respectively. Different similarity coefficients, such as Cityblock, Euclidean, Minkowski, Divergence and others were tested in order to analyze the effects of the type of characteristic (QL/PQ/QN[1]), differences in contributions (the contribution function s_k can be different for each characteristic depending on the range of observations) and different weights w_k (depending on the reliability of the characteristic). For further details see Gower 1971, 1985. Similarity coefficients (between 0 and 1; 1 = similar) were transformed into distances by subtracting them from 1. Varieties are identical at distance 0.

8. SSR-profiles of the 183 varieties were available in the database. SSR-profiles were determined by Naktuinbouw based on 9 microsatellite markers selected in the project mentioned in paragraph 4. Molecular distances were based on the Jaccard similarity coefficient, calculated by using the same FSIMILARITY directive of the statistical package GenStat.

RESULTS AND DISCUSSION

9. Differences between similarity coefficients (for the same dataset) are mainly caused by the contribution function s_k which accounts for the differences in the range of the characteristics. PQ[1] (pseudo-qualitative) and QL[1] (qualitative) characteristics were excluded from the dataset due to their unbalanced contribution to morphological distances. In addition two relatively unreliable QN (quantitative)[1] characteristics were also removed from the dataset: the frequency of flowers (TG/23/6 – char 29) and the anthocyanin coloration of the skin of the tuber in reaction to light (TG/23/6 – char 42). Remaining QN characteristics (table 1) were taken into account with equal weights ($w_k = 1$), as differences for reliability of a characteristic will be reflected in the "Distinct plus" (see Figure 1) levels already. The Cityblock coefficient was considered most appropriate for comparison of morphological distances. Cityblock distances ranged from 0 to about 0.6 for 16,653 variety pairs. Five pairs, made up by 7 varieties, were morphologically (nearly) similar.

Table 1. Morphological characteristics included in the Cityblock distance coefficient.

type char.	**char. nr. TG/23/6**	**range of char.**	**description of characteristic**	**smallest reliable difference for distinctness (# notes)**
QN	01	1 – 9	Lightsprout: size	3
QN	03	1 – 9	Lightsprout: intensity of anthocyanin coloration of base	2
QN	04	1 – 3	Lightsprout: proportion of blue in anth. coloration of base	2
QN	05	1 – 9	Lightsprout: pubescence of base	2
QN	06	1 – 9	Lightsprout: size of tip in relation to base	3
QN	07	1 – 5	Lightsprout: habit of tip	2
QN	08	1 – 9	Lightsprout: anthocyanin coloration of tip	2
QN	09	1 – 9	Lightsprout: pubescence of tip	2
QN	10	1 – 9	Lightsprout: number of root tip	2
QN	11	1 – 9	Lightsprout: length of lateral shoots	3
QN	12	1 – 3	Plant: foliage structure	2
QN	13	1 – 9	Plant: growth habit	3
QN	14	1 – 9	Stem: anthocyanin coloration	3
QN	15	1 – 9	Leaf: outline size	2
QN	16	1 – 5	Leaf: openness	2
QN	17	1 – 9	Leaf: presence of secondary leaflets	3
QN	18	1 – 9	Leaf: green color	2
QN	19	1 – 9	Leaf: anthocyanin coloration of midrib of upper side	2
QN	21	1 – 9	2nd pair of lateral leaflets: width / length	2
QN	22	1 – 9	Terminal and lateral leaflets: freq. of coalescence	3
QN	27	1 – 9	Flower bud: anthocyanin coloration	3
QN	28	1 – 9	Plant: height	2

---

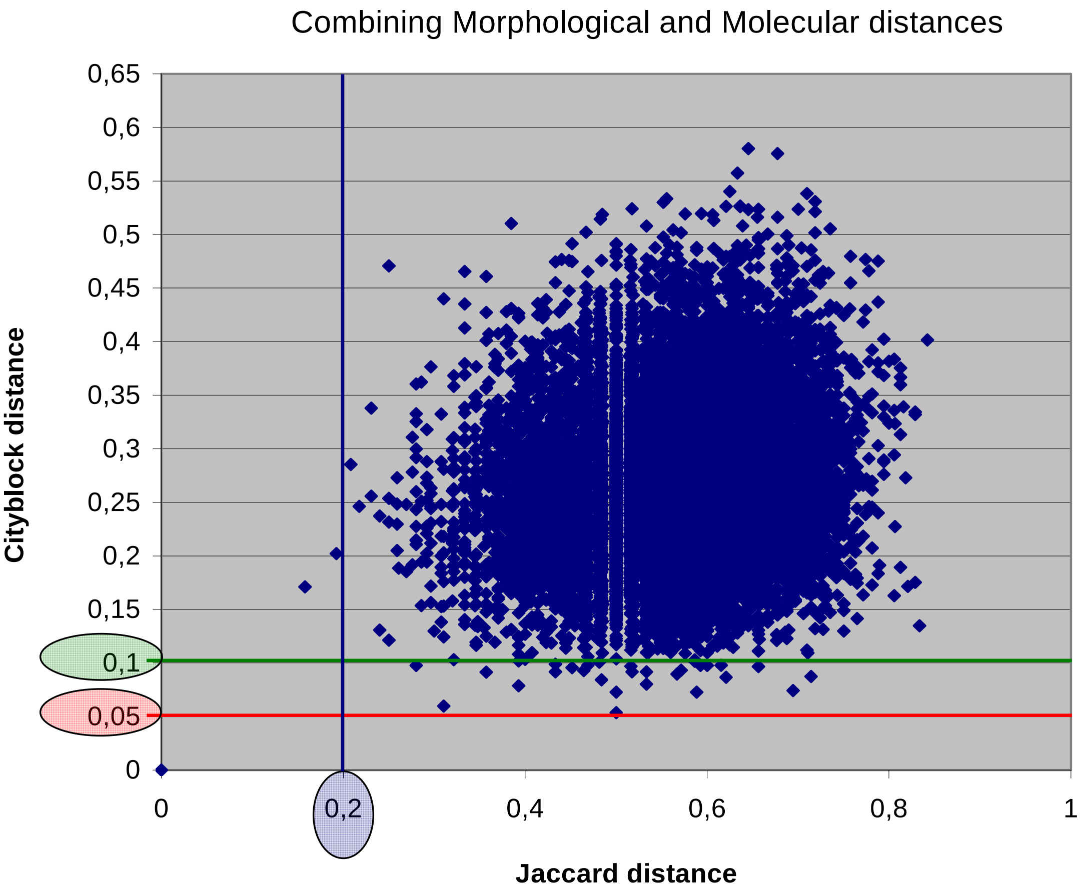
[1] see document TGP/7 Development of Test Guidelines: Annex 3: Guidance Notes (GN) for the TG Template: GN 20 (TG Template: Chapter 7: column 3) – Presentation of characteristics: States of expression according to type of expression of a characteristic

QN	30	1 – 9	Inflorescence: size	2
QN	31	1 – 9	Inflorescence: anthocyanin coloration of peduncle	3
QN	32	1 – 9	Flower corolla: size	2
QN	33	1 – 9	Flower corolla: intensity of anthocyanin coloration	3
QN	34	1 – 3	Flower corolla: proportion of blue in anth. coloration	2
QN	35	1 – 9	Flower corolla: extent of anthocyanin coloration	2
QN	36	1 – 9	Plant: time of maturity	2
QN	37	1 – 6	Tuber: shape	2
QN	38	1 – 9	Tuber: depth of eyes	2

10. Jaccard distances for the molecular data of the same 16,653 variety pairs were obtained from the microsatellite database. Jaccard distances ranged from 0 to about 0.8. The same five pairs (made up by 7 varieties), which were morphologically (nearly) similar, also had similar SSR profiles (Jaccard distance = 0).

11. Besides the 5 (nearly) similar pairs the remaining pairs (16,653 minus 5) were all distinct on the basis of a characteristic-by-characteristic approach, following the general rules on distinctness of visually observed QN-characteristics (General Introduction document TG/1/3 or document TGP/9 “Examining Distinctness”). In this approach the smallest difference for each characteristic was weighted by the DUS crop expert depending on the range and reliability of the characteristic concerned (table 1). The threshold for distinctness corresponded with a Cityblock distance of 0.05 (figure 1).

Figure 1
- Minimum Cityblock distance for distinctness (red) derived from 16,653 (minus 5) variety pairs which were distinct on the basis of a characteristic-by-characteristic approach.
- Cityblock distance for “Distinct plus” (green) = minimum distance + 1 additional state of expression.
- Most probable “Distinct plus” threshold for molecular distance (blue).
- note the 5 (nearly) similar pairs at the lowest end of both scales (0,0).

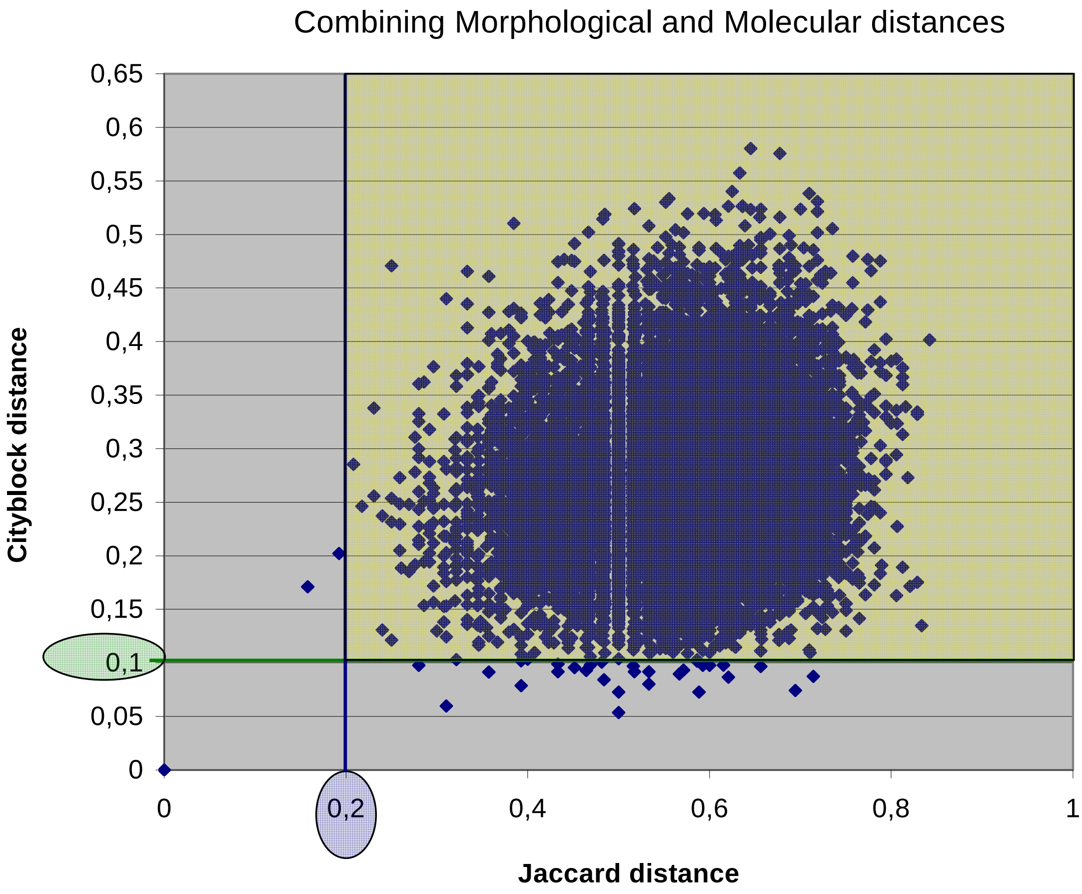


12. Varieties above the so-called “Distinct plus” level can be excluded safely from the growing trial (see document TGP/9). The “Distinct plus” threshold of each characteristic was defined by the smallest difference for distinctness + 1 additional note (i.e. one state of expression extra difference than the weighted minimum differences set by the DUS crop expert in table 1). Based on this definition – again in a characteristic-by-characteristic approach – only 20 pairs should be included in the growing trial. The “Distinct plus” threshold corresponded with a Cityblock distance of 0.10 (figure 1).

13. Considering the fact that all 16,653 pairs (minus 5) were distinct and the fact that the Jaccard threshold for distinctness was already demonstrated to be 0.15 for over 400,000 pairs in the database construction project (paragraph 4 and document BMT/11/9), the most probable “Distinct plus” threshold for molecular distance is close to Jaccard distance 0.20 (figure 1).

14. Combining both morphological and molecular “Distinct plus” thresholds leads to the following decision rules: all variety pairs which do meet the “Distinct plus” threshold for morphological distance (Cityblock > 0.1), except the varieties which do not meet the “Distinct plus” threshold for molecular distance (Jaccard < 0.2) are considered “Distinct plus” and do not need to be tested again, in the second growing cycle (highlighted upper right quadrangle in Figure 2).

Figure 2. Variety pairs declared "Distinct plus" (highlighted quadrangle) based on a combination of "Distinct plus" thresholds for morphological distance and molecular distance.



IMPLEMENTATION

15. The example illustrated for potato fits within the Model of Combining morphological and molecular distances in the management of variety collections, all assumptions being met (see document BMT/DUS Draft 5 "Possible Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)"). The example can be implemented as follows:

16. DUS testing of potato usually starts with a lightsprout test (lasting 10-12 weeks). During this period lightsprout photographs can be taken (to be included in the photo database) at optimal stage and DNA profiles can be determined. At the end of the first growing cycle decisions on distinctness can be taken, based on the traditional characteristic-by-characteristic approach for morphological differences (table 1). Positive decisions can be taken after only one year of testing for varieties which meet the "Distinct plus" threshold for morphological distance (Cityblock > 0.1), excluding the varieties which do not meet the combined "Distinct plus" threshold for morphological and molecular distance (Cityblock > 0.1 ánd Jaccard < 0.2).

17. Testing needs to be continued for varieties which do not meet the "Distinct plus" threshold for morphological distance (Cityblock distance < 0.1) or the "Distinct plus"

threshold for molecular distance (that is varieties with Jaccard distances < 0.2). Those are the varieties in the non-highlighted quadrangles of Figure 2. Mutants, synonyms, nearly similar and closely related varieties (among others) will be detected, to be compared side-by-side in the second growing cycle.

CONCLUSION

18. - Expanding the reference collection of potato with DNA profiles makes the collection much more complete – potentially up to 100% of varieties of common knowledge - which greatly enhances the quality of the testing, overcoming most of the limitations mentioned in paragraph 2 and 3.
- DNA profiles developed under strict laboratory protocols (see document UPOV/INF/17/1 "Guidelines for DNA-Profiling: Molecular Marker Selection and Database Construction ("BMT Guidlines"), chapter 5.3) are robust, thus providing checks on morphological descriptions, especially for (nearly) similar varieties.
- Molecular distances based on robust DNA profiles are a back-up on DUS decisions.
- Efficiency is increased by reducing the number of growing cycles with one year for most of the candidate varieties.
- Expanding the database with DNA profiles of original identity samples are a spin-off for identification purposes.

Henk Bonthuis
Naktuinbouw, Wageningen, The Netherlands

REFERENCES

document BMT/11/9 (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.

document BMT/11/10 (2008). Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU Common Catalogue. Part 2. the database.

document BMT/DUS Draft 5 (2011). Possible use of molecular markers in the examination of DUS, UPOV – TC/CAJ.

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27, 857-871.

Gower, J.C., 1985. Measures of similarity, dissimilarity and distance. In: Encyclopedia of Statistical Sciences, Volume 5, 397-405.

ACKNOWLEDGEMENTS

Hedwich Teunissen for providing the DNA profiles, Jan Kees Schipper for validating the morphological descriptions and Biometris for assistance in distance calculations.

[End of document]