E

BMT/9/13
ORIGINAL: English
DATE: June 15, 2005

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Ninth Session
Washington, D.C., June 21 to 23, 2005

THE POTENTIAL OF SNP MARKERS IN EXPRESSED GENES FOR IDENTIFICATION OF POTATO VARIETIES AND DETERMINATION OF DISTINCTNESS

Document prepared by experts from the Netherlands

THE POTENTIAL OF SNP MARKERS IN EXPRESSED GENES FOR IDENTIFICATION OF POTATO VARIETIES AND DETERMINATION OF DISTINCTNESS

Gerard van der Linden, Martijn van Kaauwen, Roeland Voorrips, Elena Kochieva, Ben Vosman
Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands
Contact: ben.vosman@wur.nl

Introduction

1. Before a candidate variety can obtain PBR it needs to meet certain standards regarding Distinctness, Uniformity and Stability (DUS). In Europe, DUS testing is carried out by several (national) stations. The testing usually involves the scoring of a large number of morphological characteristics that determine whether a variety conforms to the DUS criteria. In Article 1 of the UPOV act 1991 (UPOV, 1991) a variety is described as a plant grouping within a single botanical taxon of the lowest known rank, which grouping, irrespective of whether the conditions for the grant of a breeder's right are fully met, can be defined by the *expression of the characteristics* resulting from a given genotype or combination of genotypes, distinguished from any other plant grouping by the expression of at least one of the said characteristics and considered as a unit with regard to its suitability for being propagated unchanged.

2. In this definition the '*expression of the characteristics'* is the key issue. It has been argued that polymorphisms observed with molecular markers cannot be used to determine distinctness as they cannot be considered as expressed characters. One of the concerns that has been raised is that differences in molecular profiles might be detected that do not have an effect on the phenotype, which it is argued ultimately might lead to erosion of the protection offered under UPOV 1991. Therefore molecular markers are for the moment not applicable for establishing DUS within the UPOV framework. However, discussions within the UPOV-BMT working group have resulted in three possible options for the use of molecular markers within the DUS framework:

3. Option 1 is the use of molecular characteristics as a predictor of traditional characteristics; option 2 describes calibration of threshold levels for molecular characteristics against the minimum distances found from use of traditional characteristics; option 3 would be the development of a new system for assessing DUS. Following a meeting of the BMT Review Group, it was agreed that the use of approaches under options 1 and 2, with certain provisions, would not erode PBR, whereas there was no consensus on the use of option 3 type approaches.

4. Recent studies have shown that DNA polymorphisms (SNPs and microsatellites) can be found in expressed sequences e.g. in genes. These polymorphisms are expressed in the RNA and might have an effect on the phenotype. It can be argued that markers detecting such polymorphisms might be acceptable to determine distinctness, perhaps resulting in a fourth option for the introduction of molecular markers for DUS testing or an addition to option 1.
In this paper we describe the discovery of SNP markers in a number of potato genes and their potential in variety identification and determination of distinctness. In addition, we evaluate the performance of these markers in relation to an option 2 approach for the introduction of molecular markers in DUS testing.

Methods:

Plant Material

5. Three hundred and five potato varieties were used in this study. Most of the varieties used for molecular analysis were obtained from CGN. Material was harvested as light sprouts from morphologically verified tubers from the CGN collection (Wageningen, the Netherlands). A few varieties were from UK. This material was obtained from morphologically verified tubers from the NIAB potato collection, or from leaves from potato plants in the SASA potato collection. In addition, leaf material of a number of varieties was sampled from test fields of the Bundessortenamt in Germany. Also a few varieties were obtained as DNA from the Laboratory of Plant Breeding, Wageningen University. DNA was extracted from leaf or tuber material using the DNeasy DNA extraction kit (Qiagen).

SNP discovery:

6. Close to 87,000 EST sequences from the NCBI nucleotide database were downloaded in FASTA format. The majority of these sequences (72,151) are from the potato variety Kennebec, and the remaining sequences (15,863)) are from the variety Bintje. Several criteria were used for the selection of putative SNPs in order to get the highest confidence that these electronic SNPs are indeed based on genetic variation in the Kennebec and/or Bintje DNA.
In total 107 SNPs were identified that passed all criteria.

SNP analysis:

7. Pyrosequencing was used to detect SNP polymorphisms. The primers for pyrosequencing, including the sequence primer, were designed using the on-line primerdesign software of Pyrosequencing AB (Biotage AB, Sweden).

8. The SNP loci including the SNPs were amplified with a Biotin-labeled primer and a non-labeled primer. Depending on which strand the pyrosequencing reaction will be performed, either the forward primer or the reverse primer was labeled with a Biotin label. The amplified loci were purified and subjected to pyrosequencing.

Analysis of the pyrograms:

9. The SNPs were scored by first calculating the ratio of the SNP peaks relative to a reference peak in the pyrogram. subsequently the relative contribution of each allele was determined. For accurate assessment of the number of alleles present in a variety, the relative peak heights for a single SNP of all varieties were taken together and used to determine bins for each of the allelic states 0:4, 1:3, 2:2, 3:1 and 4:0. SNPs for which no distinct bins could be found for each of the allelic states were discarded. SNPs were scored from 0 to 4 (0=0:4, 1=1:3, 2=2:2, 3=3:1, and 4=4:0).

Data analysis:

10. Genetic similarity analysis using the SNP markerdata was performed using the FSIMILARITY command with TEST=euclidean (GenStat® Release 8, by the GenStat Committee, 2005). Overall morphological Euclidean similarities were calculated based on DUS trait scores (UPOV guidelines for potato TG/35/5) without any transformation or normalization, also using the Genstat FSIMILARITY command with TEST=Euclidean.

Results:

11. For potato, an extensive EST database is now available at TIGR (http://www.tigr.org), which we used for SNP discovery. In addition, several SNPs with linkage to resistance loci in the potato genome were extracted from the PoMaMo database https://gabi.rzpd.de/PoMaMo.html (Meyer et al, 2005). These markers are not necessarily located in expressed genes, but are physically linked to known resistance genes. The SNP set was complemented with one SNP that was discovered by sequencing of target genes available in nucleotide databases. Table 1 lists the SNPs used in this study. Indicated in this table is the effect that the nucleotide substitution has on the protein sequence: A synonymous substitution does not affect the protein sequence and the function of the protein is not changed. A non-synonymous substitution changes the amino acid encoded by the codon that contains the SNP. Non-synonymous SNPs may result in either a conserved mutation at the protein level (an aminoacid is replaced by another aminoacid with similar properties, which may affect function of the protein, but it is not likely) or a non-conserved mutation: an aminoacid is substituted by another aminoacid with different properties (for instance a negatively charged aminoacid replaced by a positively charged amino acid) which is likely to affect the functional properties of the protein. SNPs likely to affect the function of the protein were SNP 6, 17, 19, 27 and 35 (table 1).

Table 1: SNP origin, position and homology

SNP		**kind**	**location**	**Homology**
1	**EST**	**S***		Succinate semi-aldehyde dehydrogenase (*Arabidopsis thaliana*
2	**EST**	**Ns** (cons)**		ethylene responsive factor (AP2-like) (potato 100%)
3	**EST**			shepherd gene (Hsp-like) (*Arabidopsis thaliana*):
4	**EST**			Squalene monooxygenase
5	**EST**			Squalene monooxygenase
6	**EST**	**Ns (non-cons)**		proteasome gene family (*Arabidopsis thaliana*). 96% DNA homology to tomato mRNA (accession number gi:47105920).
7	**PoMaMo**		**ChrIX**	Na+/K+/Cl-cotransport protein (Tobacco). High homology, so putative analog. A plant cation-chloride co-transporter promoting auxin-independent tobacco protoplast division. Overexpression induces auxin-independent protoplast growth.
8	**PoMaMo**		**Chr II**	RGA (annotation by Leister et al (1996). I can not confirm this annotation… Near Marker TG5. QTLs for phytophthora resistance, and E. carotovora.
9	**PoMaMo**		**Chr IV**	No annotation. St3.2 locus. Phytophthora resistance QTL (half of the chromosome…).
10	**PoMaMo**			No annotation. Does hit an EST, so appears to be a gene. No location.

SNP		kind	location	Homology
11	**PoMaMo**		**Chr X**	lysyl-tRNA synthetase (LysRS) (tomato). A root-specific iron-regulated gene involved in protein translation machinery. In vicinity of E. carotovora resistance.
12	**PoMaMo**		**Chr II**	Potato genomic fragment. Also may have an AP2 domain. Fruit ripening, stress induced? Near QTLs for phytophthora resistance, and E. carotovora.
13	**PoMaMo**		**Chr IV**	Putative Serine-threonine protein kinase. In QTL (large one) for phytophthora resistance. No EST hit potato
14	**EST**	**s**	**Chr XI**	Potato sucrose transporter gene (100%)
15	**EST**	**Ns (cons)**		ubiquitin conjugating protein [Avicennia marina], near 100% homology)
16	**Sequencing**	**s**		Potato SKT-1 gene. Potassium channel. SNP in coding region
17	**EST**	**Ns (non-cons)**		VTC2-like (Arabidopsis).
18	**PoMaMo**		**Chr III**	Potato 4-coumarate:coenzyme A ligase (4CL-2b) gene (expressed gene). key enzyme of phenylpropanoid metabolism in plants. QTL for phytophthora resistance. Well-known RFLP marker.
19	**EST**	**Ns (non-cons)**		Putative LytB analog. Function still unknown, possible role in cell wall formation. nonmevalonate pathway, DXP pathway involved for isoprenoid synthesis
20	**EST**	**Ns (cons)**		putative (Sti1-like) stress-inducible protein [Arabidopsis thaliana]
21	**EST**	**s**		putative ribosomal protein (chloroplast) [Arabidopsis thaliana]
22	**EST**	**s**		Putative B-subunit of K-channel (opotato; 100%)
23	**EST**	**s**		glucose-6-phosphate/phosphate translocator 2 (GTP2 gene), potato, 100% homology.
24	**PoMaMo**		**Chr IX**	inorganic phosphate transporter PT1 gene (potato, homology 100%). Close to phytophthora QTL.
25	**PoMaMo**		**Chr V**	Putative gene. In R-gene hot spot at the COSA locus. Qualitative resistance to amongst others phytophthora, nematodes.
26	**EST**	**s**		glucose-6-phosphate 1-dehydrogenase (G6PDH) (potato 100% homology).
27	**EST**	**Ns (non-cons)**		putative peptidylprolyl isomerase (Arabidopsis thaliana).
28	**EST**	**s**		cellulose synthase family (Arabidopsis thaliana, limited homology).

SNP		kind	location	Homology
29	**EST**	**s**		Potato 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase (potato, 100% homology).
30	**EST**	**s**		Ephethon-regulated gene (Capsicum annuum)
31	**PoMaMo**		**Chr V**	(Likely in) gene: Putative methyltransferase, which may be responsive to dehydration. Slightly above COSA, in R-gene hot spot: qualitative resistance to amongst others phytophthora, nematodes
32	**EST**	**s**		UDP-glucose glucosyltransferase (potato, 100% homology).
33	**EST**	**s**		No annotation possible
34	**EST**	**Ns (Cons)**		delta-tonoplast intrinsic protein (*Tabacum nicotianum*)
35	**EST**	**Ns (non-cons)**		UDP-glucose glucosyltransferase (potato 50% homology]); homology is limited, no clear functional annotation
36	**EST**	**Ns (cons)**		dolichyl-di-phosphooligosaccharide-protein glycotransferase (*Arabidopsis thaliana*) (oligosaccharyltransferase)
37	**EST**	**s**		S-adenosylmethionine decarboxylase (SAMDC) (*Ipomoea nil*).
38	**EST**	**s**		ribosomal protein S5 (*Spinacia oleracea*)

s*: Synonymous (neutral) substitution: The nucleotide substitution does not change the amino acid sequence

Ns**: Nonsynonymous substitution: The nucleotide substitution changes the amino acid sequence. Cons: Conserved mutation: An amino acid is replaced by an amino acid with similar properties. Non-cons: An amino acid is replaced by an amino acid with different properties (likely to change protein function)

SNP assay development:

12. The primers for locus-specific amplification and pyrosequencing of the potato SNP loci were developed using the online software available for registered pyrosequencing users. To improve the efficiency of pyrosequencing SNP markers were, when possible, analysed in duplex e.g. two loci were taken into account in design of the primers. We were able to analyse 24 SNPs in 12 duplexes, starting with the locus PCR. An additional 10 SNPs were amplified in separate PCR reactions and duplexed in the pyrosequencing reaction. Seven SNPs were done in simplex. With these SNP assays, 305 varieties were analyzed.

SNP scoring

13. In an allotetraploid species such as potato an SNP can have three heterozygous states (3:1, 2:2 and 1:3) in addition to the two homozygous states. With pyrosequencing, it is possible to reliably detect these states separately, provided that accurate quantification is possible (Rickert et al, 2002).

14. For simplex assays, allele ratios can be determined automatically by the pyrosequencing software. For duplexes however automated determination of allele ratios was not possible. Ratios were therefore calculated manually.

15. Scorings of the SNPs were evaluated per locus. For each SNP, the relative peak heights were determined in each variety. This results in a set of relative peak heights for each allele ratio: 0:4, 1:3, 2:2, 3:1 and 4:0. Thirty-eight SNPs could be reliably scored in the full set of varieties. These SNPs are the SNPs listed in Table 1, along with several properties of these SNPs and the SNP loci.

Identification of varieties using SNP markers

16. With a set of 38 SNPs, all 305 potato varieties could thus be uniquely identified. Two pairs of varieties were nearly identical: Cardinal and Diamant as well as Eersteling and Rode Eersteling, both differing at the same marker (heterozygote vs homozygote). The high genetic similarity of these pairs of varieties is explained by the fact that Diamant is a Cardinal mutant, and Rode Eersteling is a mutant of Eersteling. The most likely explanation for the observed difference in molecular profile between the mutants is that the SNP involved (SNP 15) cannot always be scored reliably. PIC values of the different SNP markers range from 0.75 to 0.40, with one SNP being nearly monomorphic (PIC value 0.02). Using the 12 most informative markers, all but 6 varieties (in 3 pairs of two) could still be distinguished. Two of these pairs involved mutants (Cardinal/Diamant, and Eersteling/Rode Eersteling). The third variety pair consisted of varieties which are unrelated by pedigree. In contrast, with the 12 most informative SNP markers based on three-state scores (two homozygote and one heterozygote state), 46 varieties in 23 pairs could not be uniquely identified. This clearly illustrates the added value of using pyrosequencing and thus being able to discriminate the five allelic states of the tetraploid potato varieties (two homozygotic states and three heterozygotic states). The increased resolution of each SNP marker enables the unique identification of potato varieties with a limited number of markers, thus providing a cost-effective alternative to other molecular marker systems for identification.

DUS characteristics vs SNP markers for variety identification

17. For over 300 potato varieties DUS characteristics 1-50 were scored according to the UPOV guideline for potato. Most of the characteristics were scored on a 1 to 9 scale. The variety set was analyzed with DUS characteristics using an Euclidian measure for genetic similarity All varieties could be uniquely identified. The mutant pair Diamant/Cardinal was also judged highly similar based on DUS characteristics; the other mutant pair (Rode Eersteling/Eersteling) was less similar. The DUS genetic similarities were compared pairwise to the SNP genetic similarities (fig.1). Each point in the figure represents a comparison of genetic similarities of a pair of varieties.

Discussion

SNP scoring in potato

18. The use of pyrosequencing as a SNP scoring platform permitted the scoring of biallelic SNPs in five different states in the tetraploid potato, and thus determining the actual genotype. It allows the discrimination of the two homoygotic states as well as the three heterozygotic states. The set of SNP markers used in this study was very well capable of uniquely identifying all varieties in the set, except pairs of mutants, which are expected to be

genetically highly similar. Scoring five different states obviously enhances the resolution of the marker set for variety identification. This is exemplified by the fact that, with the top ten most informative SNP markers scored in five states, all but one pair of the potato varieties could still be uniquely identified, whereas with the top 15 most informative SNPs scored in three states (two homozygote and a heterozygote states), more than 5 pairs (in addition to the mutant pairs) could not be distinguished. The possibility to uniquely identify potato varieties with 6 pairs of markers in 6 duplex pyrosequencing reactions may offer a highly cost-effective and reliable alternative to currently used systems for potato variety identification.

Relationship between genetic distances observed with molecular markers and distances measured with morphological characteristics

19. Option 2 formulated by the BMT-UPOV workgroup is defined as: Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics.

20. The Crop Subgroups developed this option with the aim to ensure that there would be no significant shift in the typical minimum distances as measured by traditional characteristics. However, they noted that the lack of a clear relationship between molecular marker distances and differences in traditional characteristics would lead to the need to consider how to handle potentially different decisions on distinctness. An important question that needs to be answered is whether variety pairs, which are not distinct using traditional characteristics, would be judged as distinct using molecular characteristics and whether such decisions would be acceptable for maintaining the value of protection.

21. We show here that in the potato variety set analysed, there is no obvious relationship between the similarities measured between two varieties on the basis of the chosen molecular markers and the similarities observed on the basis of DUS characteristics (fig.2). The SNP markers in this study therefore would provide no basis for an option 2 approach for potato. However, this conclusion is strongly dependent on the way the morphological data are treated and similarities calculated. It is conceivable that some correlation might be found if characteristics are weighted differently..

SNPs in expressed DNA

22. The use of molecular markers for DUS testing has been the subject of discussions within UPOV for over 10 years now. In Article 1 of the UPOV Act 1991 (UPOV, 1991) an important point is made by stating that Distinctness of a variety is determined by the *expression of characteristics* that describe the genotype or combination of genotypes. In this study we have used a collection of molecular characteristics that are unarguably expressed. Part of the SNPs used to describe a set of over 300 potato varieties were discovered in ESTs, which are Expressed Sequence Tags. The SNP polymorphisms are therefore present in genes that are transcribed in RNA and subsequently translated into functional proteins. In this respect they are analogous to protein markers currently in use for some cereals and potato. All the EST-based polymorphisms presented in this study give rise to distinct mRNAs, and a subset (about 50%) of these SNPs (the non-synonymous substitutions) will change the sequence of the protein. The SNPs that result in non-conserved substitutions of amino acids in the protein are the most likely to also affect the function of the protein, which can have consequences for the phenotype of the variety.

An Option 4?

23. As argued above, non-synonymous SNPs in genes are expressed characters, and as such they would comply with Article 1 of UPOV 1991. In our view, this gives rise to a potential Option 4 or to an addition to option 1, which could read: "Molecular markers detecting polymorphisms in expressed genes can be considered as expressed characters and are thus acceptable as characters for determining DUS under UPOV 1991." We suggest further that the BMT Review Group should investigate this option.

Acknowledgements

This research was funded under the method development scheme of CGN.

References:

Meyer S., Nagel A., and Gebhardt C. (2005) PoMaMo—a comprehensive database potato genome data. Nucleic Acids Research 33, D666–D670

Rickert, A.M., Premstaller, A., Gebhardt, C., Oefner, P.J. (2002) Genotyping of SNPs in a polyploid genome by Pyrosequencing. BioTechniques, 32, 592–603.

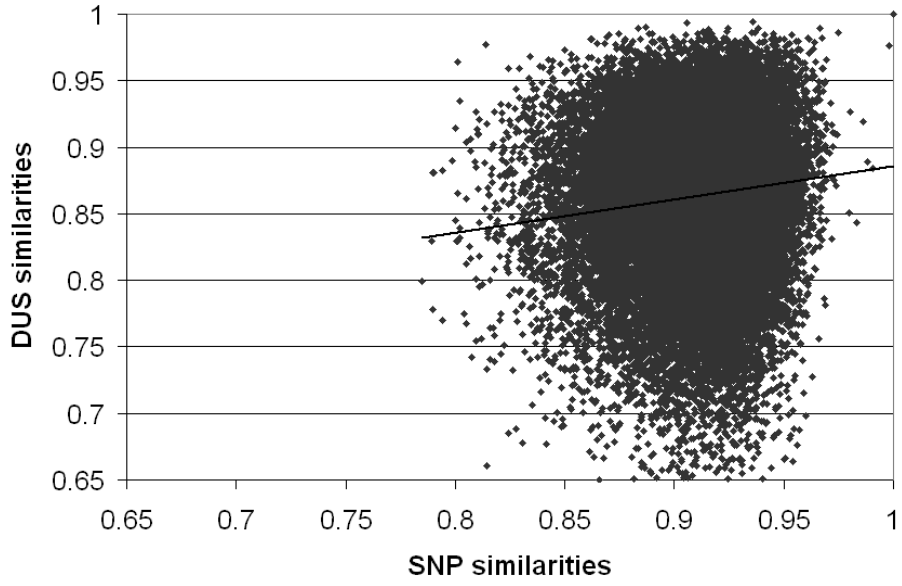

Fig. 1: Pairwise comparison of DUS genetic distances and SNP genetic similarities

[End of document]