BMT Guidelines (proj.2)
ORIGINAL: English
DATE: June 22, 2004

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

GUIDELINES FOR MOLECULAR MARKER SELECTION AND DATABASE CONSTRUCTION [~~IN VARIETY CHARACTERIZATION~~] ("BMT GUIDELINES")

DRAFT

Document prepared by the Office of the Union

to be considered by the
Ad hoc *Crop Subgroups on Molecular Techniques for: Potato; Sugarcane; and Wheat at their sessions on June 28, 2004*

Notes:

(i) Text shown in bold type is new text added to the previous version (document "BMT Guidelines (proj.1)"). Text which it proposed for deletion is shown as highlighted / strikethrough and is placed between square brackets

(ii) Where there is more than one proposal relating to a section of text, the alternatives are shown between square brackets and highlighted.

(iii) Background notes to new proposals are provided in the form of endnotes.

A. INTRODUCTION

The purpose of this document **(BMT Guidelines)** is to provide guidance for molecular marker selection and database construction. **The BMT Guidelines are intended to address the need to build databases containing molecular profiles of varieties, possibly produced in different laboratories using different technologies. This sets high demands on the quality of the markers and on the necessity of reproducing data using these markers in situations where equipment and/or reaction chemicals might change. In addition, specific precautions need to be taken regarding the quality of data entered into a database**.

[~~for the purposes of variety characterization in relation to their possible use in the examination of Distinctness, Uniformity and Stability ("DUS"), enforcement of Plant Breeders' Rights, technical verification, variety identification and the consideration of essential derivation~~ [a]].

[b]**With regard to the possible use of molecular markers in the examination of Distinctness, Uniformity and Stability ("DUS") the current situation within UPOV is explained in the Annex to this document[c] and can be summarized in terms of the following options:**

Option 1: Gene specific marker of a phenotypic characteristic

On the basis of certain assumptions (see Annex I: Option 1(a)), UPOV considers that the use of a gene specific marker of a phenotypic characteristic is acceptable within the terms of the UPOV Convention and would not undermine the effectiveness of protection offered under the UPOV system.

Option 2: Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics

On the basis of certain assumptions (see Annex I: Option 2), UPOV considers that the approach of using the calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics, where used for the management of reference collections, is acceptable within the terms of the UPOV Convention and would not undermine the effectiveness of protection offered under the UPOV system.

Option 3: Development of a New System

Regarding the development of a new system based on the use of molecular characteristics, there is no consensus on the acceptability of such a system within the terms of the UPOV Convention and no consensus on whether this would undermine the effectiveness of protection offered under the UPOV system.

B. GENERAL PRINCIPLES

1. Selection of a Molecular Marker System

1.1 The most important criteria for choosing a method are:

(a) reproducibility of data production between laboratories **and detection platforms**[d] and **repeatability** over time;
(b) discrimination power;
(c) possibilities for robust databasing; and
(d) accessibility of method for all interested parties[e].

1.2 Cost aspects are an important consideration. However, cost factors are different for each laboratory and **technology and**[f] it is not possible to make a general recommendation at this time.

1.3 As improvements in technology and new equipment become available, it is important for the continued sustainability of databases that the data produced are independent of the equipment used to produce them. This is, for example, the case with DNA sequencing data. Initially, radioactively labeled primers and sequencing gels were used to produce such data, whereas **this** can now be done using fluorescent dyes followed by separation on high throughput, largely automated, capillary gel electrophoresis systems. Despite these differences, the data produced with the various techniques are consistent with each other and independent of the techniques used to produce them. This can also apply to DNA microsatellite (simple sequence repeat, SSR) data, if attention is paid to certain specific factors (see below). This **repeatability and** reproducibility is important in the construction, operation and longevity of databases. Only in this way can a centrally maintained database, populated with verified data from a range of sources, be constructed in a cost-effective way such that the significant investment required in its establishment is only made once. On the basis of these constraints, the use of currently available multi-locus techniques is ruled out and this document confines its recommendations to well defined and researched SSRs (microsatellites) and, for the future, to sequencing information (i.e. single nucleotide polymorphisms, SNPs).

2. Selection of Molecular Markers

2.1 General Criteria

Some criteria for choosing a specific marker or set of markers apply to all marker systems. These include, in particular:

(a) useful level of polymorphism **(indicated, for example, by PIC value obtained on a set of representative varieties**[g]**)**;
(b) for scoring: repeatability within, and[h] **reproducibility between, laboratories**;
(c) distribution of the markers throughout the genome (i.e. map position) **[– useful information, but not necessarily essential for all variety characterization applications**[i]**];** and
(d) as far as possible[h] avoidance of markers with "null" [**definition**[j] **…**] alleles **- not necessarily essential for all variety characterization applications**[f]**;**

2.2 System-specific criteria

The experts from France propose that this section should contain more detailed descriptions with reference to established systems and more examples.

2.2.1 Microsatellite Markers (SSRs)

[**new section: brief explanation of Microsatellite Markers**[k]]

2.2.1.1 For effective microsatellite analysis, selecting high quality markers is essential. This includes consideration of:

(a) the degree of ["stuttering"] and other artifacts (e.g. ["echo"] bands in some systems)[l];
(b) the size of the amplification product;
(c) effective separation between the various alleles in all detection systems;
[(d)][m] reliable and repeatable scoring of the alleles/patterns **in different detection systems**[f]; and
[(e)][m] the level of polymorphism (number of alleles detected) between varieties, which requires analysis of a substantial number of varieties.

[2.2.1.2[n]] [For scoring microsatellites in different laboratories and using different detection platforms, it is crucial that reference alleles (varieties) are defined and included in all analyses. These reference alleles are necessary because molecular weight standards behave differently in the various detection **platforms** and, therefore, are not appropriate for allele identification.] / [Recommended detection platforms should be established. Reference alleles should be provided in the form of a reference mix of lines to produce an "allele-ladder".[d]]

[2.2.1.3[n]] Primers **used in a particular laboratory**[f] should, where possible, be synthesized by [a single] / [an assured] supplier to avoid the possibility of different DNA profiles being produced from primers synthesized through different sources.

2.2.2 Single nucleotide polymorphism (SNP)

[Single nucleotide polymorphisms (SNPs) can be detected via DNA sequencing, a routine technique which generally shows very high levels of repeatability **over time** and reproducibility between laboratories. However, detection of specific SNPs is currently carried out with a range of techniques, many of which are not yet routine. By their nature, SNPs have effectively two allelic states, which makes scoring relatively straightforward and reliable and should reduce or remove many of the problems noted above. However, it means that a large number of markers **may**[f] need to be analyzed, either singly or in multiplexes, to allow the unique identification of a particular genotype.] / [Specific guidance on the use of single nucleotide polymorphisms (SNPs) has not yet been developed.]

2.3 Access to the technology

Some methods and materials are publicly available. However, a large investment is necessary to obtain, for example, high quality markers and consequently markers and other methods and materials may be covered by intellectual property rights. UPOV has developed guidance for the use of characteristics which are examined by patented methods (see document TGP/7/1

Annex 3, GN 14) and this guidance should be followed for the purposes of these guidelines. It is recommended that matters concerning intellectual property rights should be addressed at the start of any work.

3. Material to be Analyzed

The source and type of the material and how many samples need to be analyzed are the main issues with regard to the material to be analyzed

3.1 Source of plant material

It is essential that as far as possible, the plant material is obtained from a reliable source. In particular, it is recommended that, wherever possible, the material be obtained from the sample of the variety submitted for examination for the purposes of Plant Breeders' Rights or for official registration. Where appropriate, the individual plants from which the samples are taken should be traceable in case some of the plants subsequently prove not to be representative of the variety.

3.2 Type of plant material

The type of plant material to be sampled and the procedure for sampling the material for DNA extraction will, to a large extent, depend on the crop or plant species concerned. For example, in seed-propagated varieties, seed may be used as the source of DNA, whereas, in vegetatively propagated varieties, the DNA may be extracted from leaf material. Whatever the source of material, the method for sampling and DNA extraction should be standardized and documented. Furthermore, it should be verified that the sampling and extraction methods produce consistent **results by**[g] DNA analysis.

3.3 [Sample size[g]] / [Number of samples]

3.3.1 Vegetatively propagated varieties

In principle, a single sample could be analyzed for vegetatively propagated varieties, as all individuals should be identical, except for mutations are which are not usually detected using the current molecular tools. However, it is advisable to analyze at least duplicate [accessions[o]] / [samples].

3.3.2 Self-pollinated and mainly self-pollinated varieties

It is not **always** appropriate to assume that self-pollinated and mainly self-pollinated varieties are homozygous at all **loci, including** SSR **or SNP**[f] loci. Heterozygosity can, for example, arise from residual heterozygosity, cross-pollination or physical admixture. [Although there are some advantages in the use of bulked samples (e.g. cost savings) and there are statistical approaches that can be useful, it is recommended to analyze individual seeds.] / [The use of single plant samples may lead to problems due to the effects of maintenance and/or shifts in allele frequency and the use of bulk samples is recommended to buffer or normalize this variation.[p]]

3.3.3 Cross-pollinated varieties

[For cross-pollinated varieties, it is recommended to analyze individual seeds because differences between varieties **may be** the result of differences in allele (or band) frequencies, **as well as** the presence or absence of **particular**[f] alleles/bands. However, in future, methods based on SNP analysis might allow the estimation of allele frequencies in bulked seed samples.] / [The use of single plant samples may lead to problems due to the effects of maintenance and/or shifts in allele frequency and the use of bulk samples is recommended to buffer or normalize this variation.[p]]

3.4 DNA reference sample

It is recommended that a DNA reference sample collection should be created from the plant material sampled according to sections 3.1 to 3.3. This has the benefit that the DNA reference sample, rather than live plant material, can be stored and supplied to others. Furthermore, it would eliminate variations due to different extraction conditions.

4. Standardization of Analytical Protocols

ISF proposal: avoid detailed guidelines for internal laboratory procedures and develop a performance-based approach, similar to that developed by ISTA for GMO testing.

4.1 Introduction

[~~In principle, it is advantageous to standardize as few things as possible.~~[q]] Methods used for genotyping and the construction of databases should be technically simple to perform and robust, allowing easy and indisputable scoring of marker profiles **in different laboratories**. [~~However, some level of~~] standardization is **desirable**, [~~such as~~ [q]] in the selection of markers, reference alleles and allele calling**/scoring**[f].

4.2 Quality criteria

4.2.1 Regarding analytical protocols, it is not essential to standardize these, but rather to agree on certain quality criteria concerning, for example:

(a) quality of DNA [to be assessed by…[r]];
(b) polymerase to be used [list of assured products to be developed[d]];
(c) [amount / concentration of each PCR component] / [other components] (e.g. PCR buffer, $MgCl_2$, dNTP, Primer, [~~Taq polymerase, DNA template~~[h]]);
(d) [~~final reaction volume~~[s]];
(e) PCR cycling conditions (including: length and temperature of initial denaturation; number of cycles; length and temperature of denaturation, annealing - for each primer pair - and extension; and length and temperature of final extension)

4.2.2 The agreed quality criteria should be set out in a protocol.

[~~4.2.3It is recommended that the protocol should include a specification that PCR products should be loaded onto the detection system in the same sequence in all laboratories, in order to ease interpretation and facilitate comparisons between laboratories.~~]

4.3 Evaluation Phase

[**new section: recommendation on number of primers**[t]]

4.3.1 Multiple laboratories

In order to select suitable markers and laboratory protocols for a given species, a preliminary evaluation phase involving more than one laboratory **("ring test")** is **[strongly**[f] recommended] / [necessary].

4.3.2 Variety choice

An appropriate number of varieties, [based on the genetic variability within a crop[d]] / [normally between 20 and 30], should be selected as the basis for the evaluation phase. The choice of varieties should reflect an appropriate range of diversity and should also include some closely related and some morphologically similar varieties to assess the level of discrimination in such cases.

4.3.3 Interpretation of results

The evaluation stage should include a blind ring test to assess the method in an objective way. Any marker which causes difficulties in any of the laboratories involved in the evaluation should be rejected for subsequent use. Ideally, as most errors in the analysis of large variety collections seem to arise from scoring errors, construction of databases should be based on duplicate samples [explanation[u] …], analyzed in different laboratories. This approach is also very effective in spotting sampling errors, or those due to heterogeneity within the samples, and eliminates possible laboratory artifacts (systematic errors).

4.4 Scoring of molecular data

4.4.1 A protocol for scoring should be developed in conjunction with the evaluation phase. The protocol should address how to score the following:

(a) rare alleles [explanation[j] …];
(b) null alleles [explanation[j] …];
(c) "faint" bands [explanation[j] …];
(d) missing data [explanation[j] …]; and
(e) monomorphic bands [explanation[j] …];

[~~4.4.2For each primer pair, a minimum and maximum size range for scoring markers **(smallest and largest marker allele[d])** should be established.~~[h]]

[~~4.4.3In cases where a gel-based system is used for revelation of markers, a suitable size ladder should be used to simplify interpretation of results between laboratories.~~[v]]

5 Constructing the Database

The experts from France comment that this is an essential aspect and the section should be developed further. In particular, it is proposed that, at least, the structure of the database (database model) and the data dictionary should be defined.

5.1 Type of database

There are many ways in which molecular data can be stored, therefore, it is important that the database structure is developed in a way which is compatible with all intended uses of the data.

5.2 Transfer of data to the database

To reduce the number of errors in data transfer and transcription, it is advisable to automate transfer of data to databases as much as possible.

5.3 Data access / ownership

It is recommended that all matters concerning ownership of data and access to data in the database should be addressed at the outset of any work.

5.4 Data analysis

The purpose for which the data will be analyzed will determine the method of analysis, therefore, no specific recommendations are made within these guidelines.

6. Validating the database

When the first phase of the database is complete it is recommended to conduct a ‘blind test’, i.e. distribute a number of samples to different laboratories and ask them to use the agreed protocol in conjunction with the database to identify them.

7. Summary

The following is a summary of the approach **for each crop**[h]**,** to the selection and use of SSRs to construct central databases of molecular profiles of varieties that are sustainable (i.e. can be populated in the future with data from a range of sources, independent of the technology used). The principles of the approach would also be suitable for use with markers such as SNPs.

(a) consider the approach on a crop-by-crop basis[d]**;**
(b) agree on a marker source[d]**;**
(c) agree on acceptable detection platforms[d]**;**
(d) agree on laboratories to be included in the test[d]**;**
(e) agree on quality issues (see section 4.2);
(f) verify source of the plant material used (see section 3);
(g) agree which SSRs are to be used in a preliminary collaborative phase, involving more than one laboratory and detection platform (see section 2);
(h) conduct evaluation (see section 4.3);
(i) develop protocol for scoring molecular data (see section 4.4);
(j) agree on the plant material reference set[d]**;**
(k) analyze **the** larger number of varieties in different laboratories/different platforms – use duplicate samples, and exchange samples/DNA extracts if problems occur. (This exercise will also provide information on sample heterogeneity);
(l) use reference varieties/DNA sample/alleles in all analyses;
(m) verify all stages (including data entry) – automate as much as possible;
(n) conduct a 'blind test' in different laboratories using the database;
(o) adopt the same procedures for adding new data.

C. CROP SPECIFIC BMT GUIDELINES

It is recommended that, as appropriate, the general principles set out in section B of this document be used to develop crop specific BMT guidelines for approval within UPOV in the following way:

(a) The Technical Committee (TC), on the basis of a recommendation by the relevant *Ad Hoc* Crop Subgroup on Molecular Techniques (Crop Subgroup), to commission the development of crop specific BMT guidelines and appoint a leading expert to coordinate, with all interested experts, the development of crop specific BMT guidelines in accordance with the principles set out in section B of this document.

(b) Draft crop specific BMT guidelines to be reviewed by the relevant Crop Subgroup and the Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular (BMT) and, subject to agreement, to be submitted for approval by the TC.

(c) Crop specific BMT guidelines approved by the TC to be published

[a] proposal to remove reference to the purposes of variety characterization to avoid some concerns e.g. concern of ISF that it implied that PBR offices had a role in decisions on essentially derived varieties
[b] redrafted to clarify UPOV position on the possible use of molecular markers in the examination of DUS
[c] this will be document TC/40/9 Add.
[d] ISF proposal
[e] proposal made by experts from Australia. Office notes that this could be elaborated on the basis of document TGP/7: GN 14 "Characteristics examined by patented methods".
[f] proposal made by experts from the United Kingdom
[g] proposal made by experts from France
[h] proposal made by experts from the Netherlands
[i] text proposed by experts from the UK. Experts from France propose to delete section in square brackets. Experts from Australia request further explanation as to why the application of variety characterization should be taken into account.
[j] Experts from Australia suggest definition be provided.
[k] proposal made by experts from Australia
[l] Experts from Australia suggest explanation of difference between "stuttering" and "echo"
[m] Experts from Australia suggest that these are general criteria which apply to any marker system – not just micro-satellite markers.
[n] Experts from Australia propose that this section be moved to Section 4.2. or 4.3
[o] Experts from Australia request clarification of what is meant by "accession" – e.g. same variety from a different location, plot etc. Also, would it be different for other types of varieties (self-pollinated, cross-pollinated etc.)?
[p] ISF notes that using individual plant samples will lead to substantially greater costs than bulk samples and would also produce extra costs for maintaining DNA reference samples.
[q] Experts from France propose to delete section in square brackets.
[r] ISF requests further information on how to determine the quality of DNA
[s] ISF questions the value of this recommendation
[t] Experts from Australia propose to include the number of primers to be tested in the evaluation phase.
[u] Experts from Australia propose the definition of "duplicate sample" (e.g. duplicate DNA or leaf samples, same or different plants etc.)
[v] ISF proposes that gel-based systems should be avoided

[End of document]