E

TWF/47/3
ORIGINAL: English
DATE: October 19, 2016

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

Geneva

TECHNICAL WORKING PARTY FOR FRUIT CROPS

Forty-Seventh Session
Angers, France, November 14 to 18, 2016

TGP DOCUMENTS

Document prepared by the Office of the Union

Disclaimer: this document does not represent UPOV policies or guidance

EXECUTIVE SUMMARY

1. The purpose of this document is to provide an overview and to present proposals concerning revisions of TGP documents.

2. The TWF is invited to note:

(a) the revisions to documents TGP/7, TGP/8 and TGP/0 to be put forward for adoption by the Council at its fiftieth session, as set out in paragraphs 6 to 13;

(b) that the proposals for future revisions of TGP documents to be discussed by the TWPs at their sessions in 2016 will be dealt with under separate documents;

(c) the new proposals for revision of TGP documents to be discussed by the TWF at its session in 2016; and

(d) the program for the development of TGP documents, as set out in Annex III to this document.

3. The following abbreviations are used in this document:

CAJ:	Administrative and Legal Committee
TC:	Technical Committee
TC-EDC:	Enlarged Editorial Committee
TWA:	Technical Working Party for Agricultural Crops
TWC:	Technical Working Party on Automation and Computer Programs
TWF:	Technical Working Party for Fruit Crops
TWO:	Technical Working Party for Ornamental Plants and Forest Trees
TWV:	Technical Working Party for Vegetables
TWPs:	Technical Working Parties

4. The structure of this document is as follows:

I. BACKGROUND

5. The approved TGP documents are published on the UPOV website at http://www.upov.int/upov_collection/en/.

II. MATTERS PROPOSED FOR ADOPTION BY THE COUNCIL IN 2016

6. The TC, at its fifty-second session, held in Geneva from March 14 to 16, 2016, agreed to invite the Council to adopt the following revisions of TGP documents at its fiftieth ordinary session, to be held in Geneva on October 27, 2016, subject to approval by the CAJ, at its session seventy-third session, to be held in Geneva on October 25 and 26, 2016, as follows (see document TC/52/29 Rev. "Revised Report", paragraphs 86 to 95):

TGP/7: Development of Test Guidelines

(i) Coverage of the Test Guidelines

7. To add:

New standard wording: TG template, Chapter 4.2:

"These Test Guidelines have been developed for the examination of [*type or types of propagation*] varieties. For varieties with other types of propagation the recommendations in the General Introduction and document TGP/13 'Guidance for new types and species', Section 4.5: 'Testing Uniformity' should be followed."

ASW 8 (c)

"(c) Uniformity assessment by off-types (all characteristics observed on the same sample size)

"For the assessment of uniformity of [self-pollinated] [vegetatively propagated] [seed-propagated] varieties, a population standard of { x } % and an acceptance probability of at least { y } % should be applied. In the case of a sample size of { a } plants, [{ b } off-types are] / [1 off-type is] allowed."

(ii) Use of Proprietary Text, Photographs and Illustrations in Test Guidelines

8. To add:

"In the case of text, photographs, illustrations or other material that is subject to third party rights, it is the responsibility of the author of the document, including Test Guidelines, to obtain the necessary permission of the third party. Material must not be included in documents where such permission is required but has not been obtained.

"Where any text, photographs, illustrations or other material that are subject to third party rights are used in Test Guidelines it should be indicated that the third party has waived their rights for the purposes of DUS testing and development of variety descriptions (e.g. indicating 'Courtesy of [name of copyright owner]' alongside the image protected by copyright)."

9. The TC agreed to include an acknowledgement in the web-based TG template in relation to text, photographs, illustrations or other material that could be subject to third party rights.

(iii) Regional Sets of Example Varieties

10. For the purposes of developing regional sets of example varieties for Test Guidelines, to add an explanation that:

(a) a "region" should be comprised of more than one country;

(b) the TWP responsible for the Test Guidelines should decide on the need and determine the basis on which the region would be established for a regional set of example varieties;

(c) the procedure for the development of sets of example varieties for a region would be determined by the TWP concerned and could, for example, be coordinated by a leading expert for the region concerned; and

(d) example varieties would need to be agreed by all UPOV members in the region concerned.

TGP/8: Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability

(iv) New Section: Examining Characteristics Using Image Analysis

11. To add a new section on "Examining characteristics using image analysis" already agreed by the TC, as set out in Annex I to this document, for inclusion in the revision of document TGP/8 Part II: Selected Techniques Used in DUS Examination.

(v) New Section: Minimizing the Variation due to Different Observers of the Same Trial

12. To include guidance on "Minimizing the variation due to different observers of the same trial", as presented in Annex II to this document, for inclusion in the revision of document TGP/8 Part I: DUS Trial Design and Data Analysis.

TGP/0: List of TGP Documents and Latest Issue Dates

13. The Council will be invited to adopt document TGP/0/9, in order to reflect the revisions of TGP documents.

14. The TWF is invited to note the revisions to documents TGP/7, TGP/8 and TGP/0 to be put forward for adoption by the Council at its fiftieth session, as set out in paragraphs 6 to 13.

III. POSSIBLE FUTURE REVISIONS OF TGP DOCUMENTS

15. The TC, at its fifty-second session, agreed that the following matters for possible future revision of TGP documents should be considered by the TWPs at their sessions in 2016 (see document TC/52/29 Rev. "Revised Report", paragraphs 96 to 121):

TGP/7: Development of Test Guidelines

(i) Drafter's Kit for Test Guidelines
See document TWF/47/9

TGP/8: Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability

(ii) The Combined-Over-Years Uniformity Criterion (COYU)
See document TWF/47/10

(iii) New Section: Examining DUS in Bulk Samples
See document TWF/47/11

(iv) New Section: Data Processing for the Assessment of Distinctness and for Producing Variety Descriptions
See document TWF/47/12

TGP/10: Examining uniformity

(v) New section: Assessing Uniformity by Off-Types on Basis of More than One Growing Cycle or on the Basis of Sub-Samples
See document TWF/47/13

16. The TWF is invited to note that the proposals for future revisions of TGP documents to be discussed by the TWPs at their sessions in 2016 will be dealt with under separate documents.

IV. NEW PROPOSALS FOR FUTURE REVISIONS OF TGP DOCUMENTS

TGP/7: Development of Test Guidelines

Duration of DUS tests in the fruit sector

17. The TWF, at its forty-sixth session, held in Mpumalanga, South Africa, from August 24 to 28, 2015 considered the information provided in document TWF/46/25 Rev. "Revised Duration of DUS Tests in the Fruit Sector" (see document TWF/46/29 Rev. "Revised Report", paragraphs 86 to 89).

18. The TWF noted that the total duration of DUS testing for fruit crops for some authorities would include the period required for establishment of the plants. The TWF agreed that over the establishment period it should be possible to conclude the DUS testing when the examining authority was certain of a negative outcome. The TWF also agreed that the DUS examination and the variety description could be completed after the first growing cycle.

19. The TWF considered the following proposal to amend document TGP/7:

> "ASW 2 (TG Template: Chapter 3.1) – Number of growing cycles
>
> "The duration of tests should be (a single/two) independent growing cycle(s) for the purpose of observation of characteristics following an adequate number of growing cycles for establishment of plants; at the end of each growing cycle(s) for the purpose of observation of characteristics the competent authority will determine whether or not the following growing cycle(s) is required. As soon as it can be established with certainty that the outcome of the DUS test will be negative, it can be stopped independently from the number of growing cycles carried out so far."

20. The TWF agreed to invite the European Union to continue drafting a proposal for reduction of duration of DUS tests in the fruit sector taking into consideration the comments received and agreed to continue discussions at its next session.

21. The TC, at its fifty-second session, agreed to consider whether to seek to amend the guidance in document TGP/7 on the duration of DUS testing for fruit crops after further discussions by the TWF, at its session in 2016. In that regard, it requested the TWF to review whether the existing guidance in TGP documents precluded the conclusion of a DUS examination after one growing cycle (see document TC/52/29 Rev. "Revised Report", paragraph 122).

TGP/14: Glossary of Terms Used in UPOV Documents

Definition of "recurved"

22. The TWF, at its forty-sixth session, considered document TWF/46/28 "Definition of 'recurved'" (see document TWF/46/29 Rev. "Revised Report", paragraphs 105 and 106).

23. The TWF noted the current extent of use of the term "recurved" in UPOV documents and agreed that further clarification and botanical references would be needed for possibly replacing the term "recurved". The TWF agreed to request the drafter from Israel to continue drafting the document to be presented for the TWF at its next session.

24. The TC, at its fifty-second session, noted the plans of the TWF to consider whether to propose to revise the definition of "recurved" in document TGP/14.

25. The TWF is invited to note the new proposals for revision of TGP documents to be discussed by the TWF at its session in 2016.

V. PROGRAM FOR THE DEVELOPMENT OF TGP DOCUMENTS

26. Annex III to this document presents the program for the development of TGP documents as agreed by the TC, as its fifty-first session, and the CAJ, at its seventy-first session and proposals made by the TWPs, at their sessions in 2015 (see document TC/51/39 "Report", paragraph 171, and document CAJ/71/10 "Report on the Conclusions", paragraph 78, respectively).

27. The TWF is invited to note the program for the development of TGP documents, as set out in Annex III to this document.

[Annexes follow]

ANNEX I

DOCUMENT TGP/8: TRIAL DESIGN AND TECHNIQUES USED IN THE EXAMINATION OF DISTINCTNESS, UNIFORMITY AND STABILITY

NEW SECTION: EXAMINING CHARACTERISTICS USING IMAGE ANALYSIS

EXAMINING CHARACTERISTICS USING IMAGE ANALYSIS

INTRODUCTION

1. Characteristics which may be examined by image analysis should also be able to be examined by visual observation and/or manual measurement, as appropriate. Explanations for observing such characteristics, including where appropriate explanations in Test Guidelines, should ensure that the characteristic is explained in terms which would enable the characteristic to be understood and examined by all DUS experts.

COMBINED CHARACTERISTICS

2. The General Introduction (document TG/1/3, Chapter 4, Section 4) states that:

> *"4.6.3 Combined Characteristics*
>
> *"4.6.3.1 A combined characteristic is a simple combination of a small number of characteristics. Provided the combination is biologically meaningful, characteristics that are assessed separately may subsequently be combined, for example the ratio of length to width, to produce such a combined characteristic. Combined characteristics must be examined for distinctness, uniformity and stability to the same extent as other characteristics. In some cases, these combined characteristics are examined by means of techniques, such as Image Analysis. In these cases, the methods for appropriate examination of DUS are specified in document TGP/12, 'Special Characteristics'."*

3. Thus, the General Introduction clarifies that the use of image analysis is one possible method for examining characteristics which fulfill the basic requirements for use in DUS testing (see document TG/1/3, Chapter 4.2), which includes the need for the uniformity and stability of such characteristics to be examined. With regard to combined characteristics, the General Introduction also explains that such characteristics should be biologically meaningful.

4. Image analysis is the extraction of information (e.g. plant measurements) from (digital) images by means of a computer. Image analysis is used in plant variety testing to help in the assessment of plant characteristics. It can be regarded as an intelligent measurement device (advanced ruler). This document aims to give guidance when using image analysis in plant variety testing.

5. Image analysis can be used in a fully automated or semi-automated way. When fully automated, the expert just records images of plant parts with a camera or scanner and the computer automatically calculates relevant characteristics without human interference. In a semi-automated way, the computer shows the images on a screen and a user can interact with the software to measure specific plant parts, e.g. by clicking with a mouse.

IMAGE RECORDING: CALIBRATION AND STANDARDIZATION

6. An important aspect to consider when recording and analyzing digital images is standardization and calibration in cases where image analysis is automated. Standardization is done by using as much as possible the same setup (illumination, camera, camera-settings, lens, perspective, and object-camera distance) for every recording. It is important to check that the recordings are done according to a prescribed protocol, as the software may depend on it. For example, pods may have to be orientated horizontally in the images, with the beaks pointing to the left. Calibration of the system is needed to make the recording as much as possible independent of any varying conditions by correcting for the variations, e.g. in size or color.

7. Size calibration is necessary. Since the measure unit in pictures is the pixel, a relation needs to be established between the pixels on the image and millimeters. A standard way to perform this calibration is to include a ruler in every recorded image, at the same distance from the camera as the plant part being recorded. In that case the user can relate the size of the ruler to the number of pixels, and make the

calibration manually. A preferred way is to use an object of standard dimensions, e.g. a coin, which can automatically be analyzed with the software and then used for an implicit size calibration. A coin also allows checking if pixels are square (i.e. if the aspect ratio of every pixel is 1:1). In all cases, the object should be sufficiently close to the calibration object and sufficiently far from the camera, to minimize the effect of varying magnification with distance. Alternatively a telecentric lens could be used to minimize this effect.

8. Illumination calibration is also necessary: an object has to be segmented from the background in the image. An often used and very simple way to do this, is to use thresholding: a pixel with a (grey) value above a certain threshold is considered an object pixel and below the threshold a background pixel (or vice versa). If the illumination is not constant, it may occur that the segmentation is not optimal for every image and that part of the pixels are assigned to the wrong class (object/background), even if the threshold value is determined automatically. This may result in erroneous measurements. It is therefore advisable to check the segmentation results by having a quick look at the segmented binary images.

9. In many situations only a silhouette/contour of the plant material is necessary, e.g. for size and shape. In these cases it is often advisable to use a background illumination, e.g. a light box. This will increase the contrast between the background and the object, and make the segmentation result much less dependent on the threshold value.

10. It should be ensured that the lighting is homogenously distributed over the image. Darker parts in the image may result in a wrong segmentation and hence lead to incorrect and incomparable measures, especially when multiple objects are recorded in the same image.

11. For colors and (variegation or blush) patterns on the plant part, it is essential that the illumination is done correctly and checked regularly, preferably for every image. In that case illumination calibration can be done by recording (part of) a standard color chart in the image. Special algorithms are available to correct for color changes due to differing illumination conditions, but in many situations this correction causes some loss of precision.

12. The light source is of large influence on the observed color in the image. Especially for color, the type of light source is important. In many cases, lamp color and intensity change during warming up of the lamps which should consequently sufficiently be warmed up before starting the recordings. If fluorescent tubes are used, it should regularly be verified that they have more or less the same intensity/color, as they may change rather rapidly with age. Calibration charts can be used to this purpose.

13. Especially when recording shiny objects like apples or certain flowers, specular reflection needs to be taken into account. Objects with specular spots cannot be measured reliably. In such cases, attention should be paid to uniform and indirect illumination, using special light tents.

14. Both (color) cameras and scanners can be used for image recording. The choice is dependent on the application and the preference of the user. Other more advanced systems, such as 3D cameras or hyperspectral cameras are not yet used in standard plant variety testing.

15. In general image analysis is used to automate the measurement of characteristics described in the guidelines of UPOV. In that case the aim is to replace a hand measurement by a computer measurement. This requires an additional calibration in addition to the image recording calibration. The measurements can then be checked with manual measurements for consistency, e.g. by a scatterplot of hand versus computer measurement with a regression line and the line y=x.

16. In some cases, image analysis requires a more precise and mathematical definition of the characteristic than is required for human experts. E.g. the length of the pod can be redefined as the length of the medial axis of the pod, excluding the stem. In such cases, there is a special need to check for differences in behavior for different genotypes (bias). The measurement for some genotypes may be exactly the same, whereas for others a systematic difference may be present. A nice example is for determining the bulb height in onions (van der Heijden, Vossepoel and Polder, 1996), where the top of the bulb was defined as the bending point of the shoulder. As long as such a change or refinement of the definition of a characteristic is known and accounted for, this is not a problem. In general, it is advisable to consult the crop experts for redefining a characteristic and check if a minor modification of the guideline might be necessary.

17. In some cases the object consists of different parts which have to be measured separately, e.g. the pod, beak and stem of a pod of French bean. This requires a special algorithm to separate the different parts (distinguish stem and beak from the pod) and this has to be tested extensively on a large number of genotypes in the reference collection, to be sure that the implementation is robust over the entire range of expressions.

18. Shape characteristics can also be measured with image analysis, but in general it will be restricted to characteristics already in the guideline, e.g. by defining the shape as the ratio between length and width.

19. Although color is a standard UPOV characteristic, and could be measured by image analysis, it is not used often. In most cases, crop experts still rely on visual observation with RHS Colour Charts.

CONCLUSIONS

20. Image analysis is used for measurements and to automate, at least partially, the assessment of characteristics. It requires a good and precise definition of the characteristic, computerization using existing or in-house software, a good preparation of samples, checking with existing procedures, careful calibration and standardization. It often necessitates therefore an investment which can only be profitable versus hand assessment of characteristics if it concerns a significant number of measurements or measurements which are difficult and time consuming to assess by the examiner. In case of organs of a small size, seed size for example, image analysis will be more precise and more reliable.

21. Image analysis offers the possibility to store information: images can be recorded and analyzed at a later stage in order to avoid peaks of work and they can be retrieved at a later stage to compare varieties for example in case of doubt.

22. Today it is mainly used for size and shape features but with the development of techniques, it will be possible to use it for a wider range of standard UPOV characteristics in future.

REFERENCES

van der Heijden, G., A. M. Vossepoel & G. Polder (1996) Measuring onion cultivars with image analysis using inflection points. *Euphytica,* 87**,** 19-31.

[Annex II follows]

ANNEX II

TGP/8/1: PART I: NEW SECTION: MINIMIZING THE VARIATION DUE TO DIFFERENT OBSERVERS OF THE SAME TRIAL

1. Introduction

This document considers variation between observers of the same trial at the authority level. It has been prepared with QN/MG, QN/MS, QN/VG and QN/VS characteristics in mind. It does not explicitly deal with PQ characteristics like color and shape. The described Kappa method in itself is largely applicable for these characteristics, e.g. the standard Kappa characteristic is developed for nominal data. However, the method has not been developed for PQ characteristics and may also require extra information on calibration. As an example, for color calibration, you also have to take into account the RHS Colour chart, the lighting conditions and so on. Differences between observers on PQ characteristics could be tested using non-parametric methods, such as frequency of deviations.These aspects are not covered in this document.

1.1 Variation in measurements or observations can be caused by many different factors, like the type of crop, type of characteristic, year, location, trial design and management, method and observer. Especially for visually assessed characteristics (QN/VG or QN/VS) differences between observers can be the reason for large variation and potential bias in the observations. An observer might be less well trained, or have a different interpretation of the characteristic. So, if observer A assesses variety 1 and observer B variety 2, the difference observed might be caused by differences between observers A and B instead of differences between varieties 1 and 2. Clearly, our main interest lies with the differences between varieties and not with the differences between the observers. It is important to realize that the variation caused by different observers cannot be eliminated, but there are ways to control it.

1.2 It is recommended that, wherever possible, one observer should be used per trial to minimize variation in observations due to different observers.

2. Training and importance of clear explanations of characteristics and method of observation

2.1 Training of new observers is essential for consistency and continuity of plant variety observations. Calibration manuals, supervision and guidance by experienced observers as well as the use of example varieties illustrating the range of expressions are useful ways to achieve this.

2.2 UPOV test guidelines try to harmonize the variety description process and describe as clearly as possible the characteristics of a crop and the states of expression. This is the first step in controlling variation and bias. However, the way that a characteristic is observed or measured may vary per year, location or testing authority. Calibration manuals made by the local testing authority and example varieties are very useful for the local implementation of the UPOV test guideline. Where needed these crop-specific manuals explain the characteristics to be observed in more detail, and specify when and how they should be observed. Furthermore they may contain pictures and drawings for each characteristic, often for every state of expression of a characteristic.

2.3 The Glossary of Terms Used in UPOV Documents (document TGP/14) provides useful guidance for clarifying many characteristics, in particular PQ characteristics.

2.4 Once an observer is trained it is important to ensure frequent refresher training and recalibration.

3. Testing the calibration

3.1 After training an observer, the next step could be to test the performance of the observers in a calibration experiment. This is especially useful for inexperienced observers who have to make visual observations (QN/VG and QN/VS characteristics). If making visual observations, they should preferably pass a calibration test prior to making observations in the trial. But also for experienced observers, it is useful to test themselves on a regular basis to verify if they still fulfill the calibration criteria.

3.2 A calibration experiment can be set up and analyzed in different ways. Generally it involves multiple observers, measuring the same set of material and assessing differences between the observers.

4. Testing the calibration for QN/MG or QN/MS characteristics

4.1 For observations made by measurement tools, like rulers (often QN/MS characteristics), the measurement is often made on an interval or ratio scale. In this case, the approach of Bland and Altman

(1986) can be used. This approach starts with a plot of the scores for a pair of observers in a scatter plot, and compare it with the line of equality (where y=x). This helps the eye gauging the degree of agreement between measurements of the same object. In a next step, the difference per object is taken and a plot is constructed with on the y-axis the difference between the observers and on the x-axis either the index of the object, or the mean value of the object. By further drawing the horizontal lines y=0, y=mean(difference) and the two lines y = mean(difference) ± 2 x standard deviation, the bias between the observers and any outliers can easily be spotted. Similarly we can also study the difference between the measurement of each observer and the average measurement over all observers. Test methods like the paired t-test can be applied to test for a significant deviation of the observer from another observer or from the mean of the other observers.

4.2 By taking two measurements by each observer of every object, we can look at the differences between these two measurements. If these differences are large in comparison to those for other observers, this observer might have a low repeatability. By counting for each observer the number of moderate and large outliers (e.g. larger than 2 times and 3 times the standard deviation respectively) we can construct a table of observer versus number of outliers, which can be used to decide if the observer fulfills quality assurance limits.

4.3 Other quality checks can be based on the repeatability and reproducibility tests for laboratories as described in ISO 5725-2. Free software is available on the ISTA website to obtain values and graphs according to this ISO standard.

4.4 In many cases of QN/MG or QN/MS, a good and clear instruction usually suffices and variation or bias in measurements between observers is often negligible. If there is reason for doubt, a calibration experiment as described above can help in providing insight in the situation.

4.5 In the case of QN/MG observations consideration and allowance may need to be given to the possible random within plot variation.

5. Testing the calibration for QN/VS or QN/VG characteristics

5.1 For the analysis of ordinal data (QN/VS or QN/VG characteristics), the construction of contingency tables between each pair of observers for the different scores is instructive. A test for a structural difference (bias) between two observers can be obtained by using the Wilcoxon Matched-Pairs test (often called Wilcoxon Signed-Ranks test).

5.2 To measure the degree of agreement the Cohen's Kappa (κ) statistic (Cohen, 1960) is often used. The statistic tries to accounts for random agreement:$\kappa = (P(agreement) - P(e)) / (1-P(e))$, where P(agreement) is the fraction of objects which are in the same class for both observers (the main diagonal in the contingency table), and P(e) is the probability of random agreement, given the marginals (like in a Chi-square test). If the observers are in complete agreement the Kappa value $\kappa = 1$. If there is no agreement among the observers, other than what would be expected by chance (P(e)), then $\kappa = 0$.

5.3 The standard Cohen's Kappa statistic only considers perfect agreement versus non-agreement. If one wants to take the degree of disagreement into account (for example with ordinal characteristics), one can apply a linear or quadratic weighted Kappa (Cohen, 1968). If we want to have a single statistic for all observers simultaneously, a generalized Kappa coefficient can be calculated. Most statistical packages, including SPSS, Genstat and R (package Concord), provide tools to calculate the Kappa statistic.

5.4 As noted, a low κ-value indicates poor agreement and values close to 1 indicate excellent agreement. Often scores between 0.6-0.8 are considered to indicate substantial agreement, and above 0.8 to indicate almost perfect agreement. If needed, z-scores for kappa (assuming an approximately normal distribution) are available. The criteria for experienced DUS experts could be more stringent than for inexperienced staff.

6. Trial design

6.1 If we have multiple observers in a trial, the best approach is to have one person observe one or more complete replications. In that case, the correction for block effects also accounts for the bias between observers. If more than one observer per replication is needed, extra attention should be given to calibration and agreement. In some cases, the use of incomplete block designs (like alpha designs) might be helpful, and an observer can be assigned to the sub blocks. In this way we can correct for systematic differences between observers.

7. Example of Cohen's Kappa

7.1 In this example, there are three observers and 30 objects (plots or varieties).The characteristic is observed on a scale of 1 to 6. The raw data and their tabulated scores are given in the following tables:

Variety	Observer 1	Observer 2	Observer 3
V1	1	1	1
V2	2	1	2
V3	2	2	2
V4	2	1	2
V5	2	1	2
V6	2	1	2
V7	2	2	2
V8	2	1	2
V9	2	1	2
V10	3	1	3
V11	3	1	3
V12	3	2	2
V13	4	5	4
V14	2	1	1
V15	2	1	2
V16	2	2	3
V17	5	4	5
V18	2	2	3
V19	1	1	1
V20	2	2	2
V21	2	1	2
V22	1	1	1
V23	6	3	6
V24	5	6	6
V25	2	1	2
V26	6	6	6
V27	2	6	2
V28	5	6	5
V29	6	6	5
V30	4	4	4

Scores for variety	1	2	3	4	5	6
V1	3	0	0	0	0	0
V2	1	2	0	0	0	0
V3	0	3	0	0	0	0
V4	1	2	0	0	0	0
V5	1	2	0	0	0	0
V6	1	2	0	0	0	0
V7	0	3	0	0	0	0
V8	1	2	0	0	0	0
V9	1	2	0	0	0	0
V10	1	0	2	0	0	0
V11	1	0	2	0	0	0
V12	0	2	1	0	0	0
V13	0	0	0	2	1	0
V14	2	1	0	0	0	0
V15	1	2	0	0	0	0
V16	0	2	1	0	0	0
V17	0	0	0	1	2	0
V18	0	2	1	0	0	0
V19	3	0	0	0	0	0
V20	0	3	0	0	0	0
V21	1	2	0	0	0	0
V22	3	0	0	0	0	0
V23	0	0	1	0	0	2
V24	0	0	0	0	1	2
V25	1	2	0	0	0	0
V26	0	0	0	0	0	3
V27	0	2	0	0	0	1
V28	0	0	0	0	2	1
V29	0	0	0	0	1	2
V30	0	0	0	3	0	0

The contingency table for observer 1 and 2 is:

O1\O2	1	2	3	4	5	6	Total
1	3	0	0	0	0	0	3
2	10	5	0	1	0	1	17
3	2	1	0	0	0	0	3
4	0	0	0	1	0	0	1
5	0	0	0	1	0	2	3
6	0	0	1	0	0	2	3
Total	15	6	1	3	0	5	30

The Kappa coefficient between observer 1 and 2, κ(O1,O2) is calculated as follows:

- κ(O1,O2) = (P(agreement between O1 and O2) – P(e)) / (1 – P(e)) where:
- P(agreement) = (3+5+0+1+0+2)/30 = 11/30 ≈ 0.3667 (diagonal elements)
- P(e) = (3/30).(15/30) + (17/30).(6/30) + (3/30).(1/30) + (1/30).(3/30) + (3/30).(0/30) + (3/30).(5/30) ≈ 0.1867. (pair-wise margins)
- So κ(O1,O2) ≈ (0.3667-0.1867) / (1-0.1867) ≈ 0.22

This is a low value, indicating very poor agreement between these two observers. There is reason for concern and action should be taken to improve the agreement. Similarly the values for the other pairs can be calculated: κ(O1,O3) ≈ 0.72, κ(O2,O3) ≈ 0.22.Observer 1 and 3 are in good agreement. Observer 2 is clearly different from 1 and 3 and the reasons for the deviation requires further investigation (e.g. consider need for additional training).

8. References

Cohen, J. (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46.

Cohen, J. (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4): 213-220.

Bland, J. M. Altman D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement, Lancet: 307–310.

http://www.seedtest.org/en/stats-tool-box-_content---1--1143.html (ISO 5725-2 based software)

[Annex III follows]

ANNEX III

PROGRAM FOR TGP DOCUMENTS

			2016					2017						2018					
	Title of document	**Current approved documents**	**TC-EDC**	**TC/52**	**TWPs**	**CAJ/73**	**C/50**	**TC-EDC**	**TC/53**	**CAJ/74**	**TWPs**	**CAJ/75**	**C/51**	**TC-EDC**	**TC/54**	**CAJ/76**	**TWPs**	**CAJ/77**	**C/52**
TGP/0	**List of TGP Documents and Latest Issue Dates**	**TGP/0/8 ADOPTED**					**TGP/0/9 Adopt**												**TGP/0/10 Adopt**
TGP/1	**General Introduction with Explanations**	-	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
TGP/2	**List of Test Guidelines Adopted by UPOV**	**TGP/2/2 ADOPTED**																	
TGP/3	**Varieties of Common Knowledge**	C(Extr.)/19/2 Rev.																	
TGP/4	**Constitution and Maintenance of Variety Collections**	**TGP/4/1 ADOPTED**																	
TGP/5	**Experience and Cooperation in DUS Testing**	**ADOPTED**																	
TGP/6	**Arrangements for DUS Testing**	**ADOPTED**																	
TGP/7	**Development of Test Guidelines**	**TGP/7/4 ADOPTED**																	
	Drafter's Kit for Test Guidelines (Drafter: Office of the Union)			TC/52/28	TWP/[xx]/9			x	x		x			x	x	x			**TGP/7/6 Adopt**
	Coverage of the Test Guidelines (Drafter: Office of the Union)		TC-EDC/Jan16/2	TC/52/15		CAJ/73/[xx]	**TGP/7/5 Adopt**												
	Use of Proprietary Text, Photographs and Illustrations in Test Guidelines (Drafter: Office of the Union)		TC-EDC/Jan16/3	TC/52/14		CAJ/73/[xx]	**TGP/7/5 Adopt**												
	Regional Sets of Example Varieties (Drafter: Office of the Union)		TC-EDC/Jan16/4	TC/52/15		CAJ/73/[xx]	**TGP/7/5 Adopt**												
TGP/8	**Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability**	**TGP/8/2 ADOPTED**																	
	Examining characteristics using image analysis (document TC-EDC/Jan16/2 Annex I)		TC-EDC/Jan16/2	TC/52/5		CAJ/73/[xx]	**TGP/8/3 Adopt**												
	Minimizing variation due to different observers (Drafter: Nik Hulse (Australia))		TC-EDC/Jan16/5	TC/52/16		CAJ/73/[xx]	**TGP/8/3 Adopt**												
	Method of Calculation of COYU (Drafter: Adrian Roberts (United Kingdom))			TC/52/17	TWP/[xx]/10				x		x			x	x	x			**TGP/8/4 Adopt**
	Examining DUS in Bulk Samples (Drafter: Office of the Union)			TC/52/18	TWP/[xx]/11				x		x			x	x	x			**TGP/8/4 Adopt**
	Data Processing for the Assessment of Distinctness and for Producing Variety Descriptions (Drafter: Office of the Union)			TC/52/19	TWP/[xx]/12				x		x			x	x	x			**TGP/8/4 Adopt**

PROGRAM FOR TGP DOCUMENTS

			2016					2017						2018					
	Title of document	**Current approved documents**	**TC-EDC**	**TC/52**	**TWPs**	**CAJ/73**	**C/50**	**TC-EDC**	**TC/53**	**CAJ/74**	**TWPs**	**CAJ/75**	**C/51**	**TC-EDC**	**TC/54**	**CAJ/76**	**TWPs**	**CAJ/77**	**C/52**
TGP/9	**Examining Distinctness**	**TGP/9/2 ADOPTED**																	
TGP/10	**Examining Uniformity**	**TGP/10/1 ADOPTED**																	
	Assessing Uniformity by Off-types on the Basis of More than One Growing Cycle or on the Basis of Sub-Samples (Drafter: Office of the Union)			TC/52/20	TWP/[xx]/13				x		x			x	x	x			**TGP/10/2 Adopt**
TGP/11	**Examining Stability**	**TGP/11/1 ADOPTED**																	
TGP/12	**Guidance on Certain Physiological Characteristics**	**TGP/12/2 ADOPTED**																	
TGP/13	**Guidance for New Types and Species**	**TGP/13/1 ADOPTED**																	
TGP/14	**Glossary of Terms Used in UPOV Documents**	**TGP/14/3 ADOPTED**																	
TGP/15	**Guidance on the Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)**	**TGP/15/1 ADOPTED**																	

[End of document]