F

TC/45/14
ORIGINAL : anglais
DATE : 10 mars 2009

UNION INTERNATIONALE POUR LA PROTECTION DES OBTENTIONS VÉGÉTALES
GENÈVE

COMITÉ TECHNIQUE

Quarante-cinquième session
Genève, 30 mars – 1er avril 2009

DOCUMENTS TGP : SECTIONS À ÉLABORER SÉPARÉMENT

Document établi par le Bureau de l'Union

Comme expliqué dans le document TC/45/5 intitulé "Documents TGP" (voir les paragraphes 24 à 29 et 36 à 39), le présent document contient des informations (en anglais seulement) sur les sections des documents TGP qui ne figurent pas dans les projets de documents TGP soumis au Comité technique (TC) pour examen à sa quarante-cinquième session, afin de permettre leur élaboration séparée et leur insertion dans une future version révisée des documents TGP concernés. Ces informations sont présentées de la manière suivante :

Annexe I : sections ne figurant pas dans le document TGP/8/1 Draft 12;
Annexe II : Section 2, Sous-section 3 "Couleur", ne figurant pas dans le document TGP/14/1 Draft 8;
Annexe III : a) conclusions de l'atelier concernant le document TGP/14/1 Section 2, Sous-section 3 "Couleur"; et
b) observations formulées par les groupes de travail techniques concernant le document TGP/14/1 Draft 6 Section 2, Sous-section 3 "Couleur" et les conclusions de l'atelier concernant le document TGP/14/1 Section 2, Sous-section 3 "Couleur".

[L'annexe I suit]

TC/45/14

ANNEX I

SECTIONS FROM TGP/8/1 DRAFT 11 PART I

TABLE OF CONTENTS PAGE

3.1.3 Statistical tests central to the statistical methods

A number of different statistical methods have been developed to assess distinctness. These methods use statistical tests to assess whether differences between variety means are significant. The choice of the statistical test that is used by a statistical method has implications in terms of risks or chances of making statistical errors. This section describes two statistical tests that are commonly used. These are the Least Significant Difference and the Multiple Range Test.

3.1.3.1 The Least Significant Difference (LSD)

3.1.3.1.1 The Least Significant Difference (LSD) is a statistic used to compare variety means from analysis of variance (ANOVA) of a characteristic and to make decisions about whether the varieties are significantly different from each other in that characteristic. In other words it represents the minimum difference between two variety means that the crop expert may declare to be different at a given significance level. The LSD is calculated using an estimate of random variation from the ANOVA.

3.1.3.1.2 It would be inconsistent with the rest of this document to describe the LSD in detail as descriptions can be found in many statistical text books. However, enough detail will be given to place it in context with the following sections on Multiple Range Tests and their comparison with LSDs.

3.1.3.1.3 The LSD is chosen to give a particular size or significance level of test (α%) when comparing two means using a single characteristic, e.g. 5% or 1%. It means that if an LSD is used to make an a priori comparison, i.e. without knowledge of the data, then there is an α% chance of making a Type I error, i.e. declaring the means of two varieties to be significantly different when, if all plants of the two varieties could be examined, the means would not be different.

3.1.3.1.4 Although the LSD controls the comparison-wise Type I error chance, it does not control the experiment-wise Type I error chance, i.e. the chance that in all the comparisons made, the means of at least one pair of varieties are significantly different when, if all plants of the varieties could be examined, the means would not be different. The more comparisons that an LSD is used to make, the greater the experiment-wise Type I error chance. For example, if a 5% LSD is used to compare 14 independent pairs of means, then there is a 51% chance (=100% x (1- (1-0.05)^{14})) of declaring at least one of the pairs of variety means to be significantly different when, if all plants of the varieties could be examined, the means would not be different.

3.1.3.1.5 The standard formula for an α% LSD to compare two means made up of n_1 and n_2 observations respectively is:-

$$LSD\,\alpha\% = t_{(\alpha\%,rdf)} \times \sqrt{s^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where s^2 is an estimate of random variation taken from the ANOVA, rdf is the degrees of freedom of s^2, and $t_{(\alpha\%,\ rdf)}$ is either the two-sided or the one-sided α% critical value of the Student's t-statistic on rdf degrees of freedom, depending on whether the test is two-tailed or one-tailed.

3.1.3.1.6 It is important to note that in using an LSD to test the differences between variety means, the crop expert is assessing whether the difference in the variety means is larger than the difference that might reasonably have arisen due to chance or random variation affecting the observations making up the variety means when there was no difference between the varieties. Thus, the source of variation used to estimate random variation (s^2) in the LSD is very important in terms of the conclusions or inferences that can be drawn about the consistency of any differences between varieties declared to be significantly different.

3.1.3.2 The Multiple Range Test (MRT)

3.1.3.2.1 A Multiple Range Test (MRT), also known as a multiple comparison test, is similar to an LSD in that it is:

a statistic used to compare variety means from analysis of variance (ANOVA) of a characteristic and decide about the significance of variety differences;

calculated using an estimate of random variation from the ANOVA.

3.1.3.2.2 An MRT differs from an LSD in that it is chosen to give a particular size (α%) of test over all the comparisons for which it is intended for a characteristic. In other words, it controls to an extent the experiment-wise Type I error chance. It does this by reducing the comparison-wise Type I error chance and, as a result, the critical value of an α% MRT is larger than that of an α% LSD. This means that the MRT is usually more conservative than the LSD in that it is less likely to declare as significantly different two variety means where, if all plants of the two varieties could be examined, the means would not be different. On the other hand, the MRT is less powerful than the LSD as its Type II error chance is larger. In other words, there is a smaller chance with the MRT than with the LSD of declaring as significantly different two variety means where, if all plants of the two varieties could be examined, the means would be different.

3.1.3.2.3 There are a number of different MRTs. The choice of which to use depends partly on the comparisons to be made: for example, if one particular variety mean is to be compared with all others, or if all variety means are to be compared with all others. Descriptions of MRTs can be found in many statistical text books.

3.1.3.2.4 As with the LSD, the source of variation used to estimate random variation (s^2) in the MRT is very important in terms of the conclusions or inferences that can be drawn about the consistency of any differences between varieties declared to be significantly different.

3.1.3.3 Comparison of the use of the LSD and the MRT in distinctness testing

3.1.3.3.1 Comparison of the use of the LSD and the MRT in distinctness testing hinges on a risk that is different to both the experiment-wise and the comparison-wise Type I error chances. It is a risk of particular interest to testers and is called here the ‘test-wise Type I error chance’. It is the chance of one or more candidates being significantly different from all other varieties in at least one characteristic when, if all plants of the varieties could be examined, the means would not be different. In other words, it is the chance of one or more candidates being wrongly declared as distinct when they are not distinct.

3.1.3.3.2 The test-wise Type I error chance increases with the number of candidates and with the number of characteristics used in the comparisons. It decreases with the number of reference varieties and proportionally with the significance level used when comparing varieties on a characteristic by characteristic basis, i.e. the comparison-wise Type I error

chance. It is generally a very small chance, except however when there are few reference varieties, and in particular when there are many characteristics being used for the comparisons. Consequently, when trials are small, i.e. few reference varieties, and particularly when many characteristics are being used for variety comparisons, it is advantageous to use an MRT in place of an LSD, as the MRT serves to reduce the comparison-wise Type I error chance and hence reduce the test-wise Type I error chance. This effectively protects the tester's interests, as it reduces the chance of incorrectly declaring varieties distinct when they are not. In doing so it marginally penalizes the breeders, as the lower power of the MRT makes it harder to detect differences when they do exist. An alternative in these circumstances would be to use an LSD with a smaller significance level. This would also reduce the comparison-wise Type I error chance and hence reduce the test-wise Type I error chance.

3.1.3.3.3 With larger trials the test-wise Type I error chance is very small, and so the advantage of the MRT over the LSD in controlling it does not exist and the LSD should be used in preference to the MRT as being the more powerful test.

3.1.3.3.4 Depending on which MRT is used, the minimum difference between two variety means represented by a MRT depends either on the total number of varieties in the trial or on the relative position rank-wise of other varieties with respect to the pair being compared. In either case, the acceptance of a candidate would be affected by the other candidates included in the trial, which may not be considered a fair system for testing. However, as the degree to which the acceptance of a candidate is affected is proportionate to the size of the trial, providing the MRT is used with small trials, its impact on testing is likely to be minor. Nonetheless, if the small trial has relatively many candidate varieties, it may be necessary to take steps in order not to penalize one breeder because another breeder has entered many candidates. In this regard, grouping of varieties as described in document TGP/9 may be useful.

3.1.3.3.5 When using an MRT for distinctness testing not all the comparisons on which the MRT is based are necessarily made. Hence the MRT critical value is larger and the comparison-wise Type I error chance is smaller than are needed to achieve the intended experiment-wise Type I error chance. However, this is not a disadvantage when the MRT is used in small trials, as it still serves to reduce the test-wise Type I error chance.

[i] ~~3.1.5 Parametrical and non-parametrical statistical methods~~

3.1.5.1 Introduction

The statistical methods included in this document are grouped into parametric and non-parametric statistical methods. Parametric statistics refers to any statistical method that assumes the population fits a probability distribution (usually the normal distribution). Non-parametric statistics refers to any statistical methods that do not make assumptions about the underlying population distribution. Non-parametric statistical methods may be used for quantitative scaled data where assumptions for parametric methods are not met. They are often used for categorical data.

3.1.5.2 Use of non-parametric statistical methods

3.1.5.2.1 Introduction

3.1.5.2.1.1 Non-parametric methods are useful tools for DUS testing particularly when either:

- Observations are made using qualitative scales where the intervals between states of expression are not known or not necessarily equal (e.g. ordinal or nominal scales [cross ref); or
- The underlying statistical assumptions needed by the parametric methods are not met or are untested.

3.1.5.2.1.2 Ordinal and nominal scaled data contain less information than interval or ratio data, and their analysis is by definition, less sensitive. This leads to the conclusion that non-parametric methods are less powerful because, for the same sample size, they are less likely to confirm small differences between varieties. However where properly used, this may be an acceptable outcome which contributes to the maintenance of minimum distance and assists determination of "clearly distinct" as compared with "distinct by the smallest of differences".

3.1.5.2.1.3 Non-parametric methods are well suited to the analysis of characteristics assessed by "notes" such as for pseudo-qualitative and qualitative data and in situations where objective rigor is required in the development of national descriptors.

3.1.5.2.1.4 While non-parametric methods are usually applied to the analysis of ordinal and nominal scaled data, they can also be used to analyze interval or ratio data. Nominal scaled data can only be analyzed using non-parametric methods.

3.1.5.3.1.5 Where sample size is small, (say less than 6 observations), there is no alternative to using non-parametric methods unless the distribution of the states of expression of the candidate variety are known exactly (a rare circumstance for DUS testing authorities).

3.1.5.2.2 Role of non-parametric analysis for analyzing quantitative data

3.1.5.2.2.1 Generally, for quantitative measured data, such as plant length in centimeters or number of stamens (see TGP8 Part I, Section 2.5.4) *[cross reference]*, parametric statistical methods are preferred. The use of parametric methods relies on underlying assumptions of the population distribution. They are usually robust and powerful even if there is moderate departure from the statistical assumptions (such as departure from a normal distribution). If assumptions are badly violated, non-parametric tests could be employed, however, before doing so, it is necessary to first investigate whether experimental error is the cause (see TGP8 Part I, Section 4.2) *[cross reference]* or establish that the type of data collected does not fit the

parametric assumptions. There are many non-parametric tests (e.g. Kruzkal-Wallis one way analysis of variance and Mann-Whitney U test) that could be used and these are well documented and described. The use of non-parametric statistics for quantitative measured data from DUS trials is the exception rather than the rule and it is not necessary to describe these further here. Instead it is sufficient to note that these methods are documented in statistical literature and can be considered if necessary.

3.1.5.2.3 Role of non-parametric analysis for analyzing qualitative data

3.1.5.2.3.1 Some characteristics routinely used in DUS testing do not usually satisfy the assumptions required for parametric methods.. Qualitatively scaled data are usually obtained from visually assessed characteristics using ordinal or nominal scales (see TGP/8 Part I, section 2.5.4.2) *[cross reference]*. For example, where individual plants are scored on 1 to 10 scale of increasing resistance to a particular disease, the position within the scale is important (i.e. it is an ordinal scale). If one plant is assessed as having a higher level of resistance than another then it is scored with a higher number on the scale. However, it is usually difficult to precisely identify the limit of each interval of the scale. Consequently, the exact interval size is unknown and is likely to vary. For this reason the scores cannot be treated as quantitative data with an assumed normal distribution which would allow the use of parametric methods. Instead it is appropriate to use non-parametric methods that do not rely on equally spaced intervals. Another example is scoring of results from an iodine starch test in assessing the maturity of apples using an ordinal scale.

3.1.5.2.3.2 Sometimes individual plants can be placed in "categories" where the order does not matter (i.e. a nominal scale) e.g. scoring plants as shattering or non-shattering in *Phalaris*.

3.1.5.2.3.3 Where all or most plants of a variety fall into one category it is unnecessary to apply a statistical method to decide on distinctness. However, in some cases, particularly for cross pollinated varieties, the allocation to categories is not absolute and there will be a certain amount of heterogeneity in the population due to the breeding system of the species. The consequence is that large numbers of plants of the variety may be allocated to different categories. This is acceptable provided the degree of heterogeneity is within that for comparable varieties of the species. A decision has to be made as to whether there is sufficient separation to establish distinctness between varieties.

3.1.5.2.3.4 In these cases, non-parametric statistical methods can be used as they do not rely on assumptions about the underlying population distribution of the data.

3.1.5.2.3.5 Whilst there are many non-parametric methods that can be used for qualitative data, two methods commonly used in plant variety testing are Chi-square (χ^2) and Fishers Exact Test. For convenience these are briefly described below.

2. DATA TO BE RECORDED

[TWC: Mr. Uwe Meyer (Germany) to revise and restructure the section starting from the perspective of characteristics as viewed by DUS experts e.g. using Tables 2 and 3 and to include examples for clarification.]

2.1 Introduction

Document TGP/9 Examining Distinctness, sections 4.4 and 4.5 provide the following guidance on the type of observation for distinctness in respect to the type of characteristic and the method of propagation of the variety:

> **"4.4 Recommendations in the UPOV Test Guidelines**
>
> The indications used in UPOV Test Guidelines for the method of observation and the type of record for the examination of distinctness, are as follows:
>
> Method of observation
>
> M: to be measured (an objective observation against a calibrated, linear scale e.g. using a ruler, weighing scales, colorimeter, dates, counts, etc.);
>
> V: to be observed visually (includes observations where the expert uses reference points (e.g. diagrams, example varieties, side-by-side comparison) or non-linear charts (e.g. color charts). "Visual" observation refers to the sensory observations of the expert and, therefore, also includes smell, taste and touch.
>
> Type of record(s)
>
> G: single record for a variety, or a group of plants or parts of plants;
>
> S: records for a number of single, individual plants or parts of plants
>
> For the purposes of distinctness, observations may be recorded as a single record for a group of plants or parts of plants (G), or may be recorded as records for a number of single, individual plants or parts of plants (S). In most cases, "G" provides a single record per variety and it is not possible or necessary to apply statistical methods in a plant-by-plant analysis for the assessment of distinctness.

4.5 Summary

The following table summarizes the common method of observation and type of record for the assessment of distinctness, although there may be exceptions:

	Type of expression of characteristic		
Method of propagation of the variety	QL	PQ	QN
Vegetatively propagated	VG	VG	VG/MG/MS
Self-pollinated	VG	VG	VG/MG/MS
Cross-pollinated	VG/(VS*)	VG/(VS*)	VS/VG/MS/MG
Hybrids	VG/(VS*)	VG/(VS*)	**

* Records of individual plants only necessary if segregation is to be recorded.
** To be considered according to the type of hybrid."

The following sections consider the data in relation to the type of record and type of trial design:

2.2 Side-by-side visual comparison[b]

2.2.1 When distinctness is assessed by side-by-side visual comparison, uniformity is assessed by off-types. In these cases, the trial design is a single plot, there is a single record per variety, which is obtained from visual observations of a group of plants or part of plants (VG), which provide notes (see sections 1.6.1.6 and 1.6.2) *[cross ref.]*.

2.3 Notes/Single variety records[c]

2.3.1 When distinctness is assessed by notes/single variety records, uniformity is assessed by off-types. In these cases, the trial design consists of single plots[d]. There is a single record per variety which is obtained from visual observation of a group of plants or part of plants (VG), providing a note, or a measurement of a group of plants or parts of plants (MG) (see sections 1.6.1.6 and 1.6.2) *[cross ref.]*.

2.4 Variety mean/statistical analysis of records of group of plants [variety mean statistical analysis of records of group data][e]

2.4.1 In general, when distinctness is assessed, for at least some characteristics, by a variety mean or by statistical analysis of groups of plants, uniformity is assessed by off-types. In these cases, the trial design is replicate plots (see sections 1.6.1.7 and 1.6.3.2) *[cross ref.]*.

2.4.2 Records from visual observation of a group of plants or part of plants provide notes which belong to qualitative scale data. It is important to note that, in general, it is not possible to calculate means with qualitative scale data (see section 2.5.4.2) *[cross ref.]*.

2.5 Statistical analysis of individual plant data

2.5.1 Introduction

2.5.1.1 When distinctness is assessed, for at least some characteristics, by statistical analysis of individual plant data, uniformity is assessed by standard deviation for relevant characteristics.

2.5.1.2 In order to understand how statistical analysis can be appropriate to trial data it is necessary to answer the following questions:

1. What is a characteristic?
2. What is a process level?
3. What is a scale level of a characteristic?
4. What is the influence of the scale level on the :
   - planning of a trial,
   - recording of data,
   - determination of distinctness and uniformity and
   - description of varieties.

2.5.2 Different levels to look at a characteristic

2.5.2.1 Introduction

2.5.2.1.1 Characteristics can be considered in different levels of process (Table 1). The characteristics as expressed in the trial (type of expression) are considered as process level 1. The data taken from the trial for the assessment of distinctness, uniformity and stability are defined as process level 2. These data are transformed into states of expression for the purpose of variety description. The variety description is process level 3.

Table 1: Definition of different process levels to consider characteristics

Process level	Description of the process level
1	characteristics as expressed in trial
2	data for evaluation of characteristics
3	variety description

From the statistical point of view, the information level decreases from process level 1 to 3. Statistical analysis is only applied in level 2.

2.5.2.1.2 Sometimes for crop experts it seems that there is no need to distinguish between different process levels. The process level 1, 2 and 3 could be identical. However, in general, this is not the case.

2.5.2.2 Understanding the need for process levels

2.5.2.2.1 The crop expert may know from UPOV Test Guidelines or his own experience that, for example, ‘Length of plant’ is a good characteristic for the examination of DUS. There are varieties which have longer plants than other varieties. Another characteristic could be ‘Variegation of leaf blade’. For some varieties, variegation is present and for others not. The crop expert has now two characteristics and he knows that ‘Plant length’ is a quantitative characteristic and ‘Variegation of leaf blade’ is a qualitative characteristic (definitions: see

Part I: section 2.5.3.2 to 2.5.3.4 *[cross ref.]* below). This stage of work can be described as **process level 1.**

2.5.2.2.2 The crop expert then has to plan the trial and to decide on the type of observation for the characteristics. For characteristic 'Variegation of leaf blade', the decision is clear. There are two possible expressions: 'present' or 'absent'. The decision for characteristic 'Plant length' is not specific and depends on expected differences between the varieties and on the variation within the varieties. In many cases, the crop expert will decide to measure a number of plants (in cm) and to use special statistical procedures to examine distinctness and uniformity. But it could also be possible to assess the characteristic 'Plant length' visually by using expressions like 'short', 'medium' and 'long', if differences between varieties are large enough (for distinctness) and the variation within varieties is very small or absent in this characteristic. The continuous variation of a characteristic is assigned to appropriate states of expression which are recorded by notes (see document TGP/9, section 4*)[cross ref.]*. The crucial element in this stage of work is the recording of data for further evaluations. It is described as **process level 2**.

2.5.2.2.3 At the end of the DUS test, the crop expert has to establish a description of the varieties using notes from 1 to 9 or parts of them. This phase can be described as **process level 3**. For 'Variegation of leaf blade' the crop expert can take the same states of expression (notes) he recorded in process level 2 and the three process levels appear to be the same. In cases where the crop expert decided to assess 'Plant length' visually, he can take the same states of expression (notes) he recorded in process level 2 and there is no obvious difference between process level 2 and 3. If the characteristic 'Plant length' is measured in cm, it is necessary to assign intervals of measurements to states of expressions like 'short', 'medium' and 'long' to establish a variety description. In this case, for statistical procedures, it is important to be clearly aware of the relevant level and to understand the differences between characteristics as expressed in the trial, data for evaluation of characteristics and the variety description. This is absolutely necessary for choosing the most appropriate statistical procedures in cooperation with statisticians or by the crop expert.

2.5.3 Types of expression of characteristics

2.5.3.1 Characteristics can be classified according to their types of expression. The consideration of the type of expression of characteristics corresponds to process level 1. The following types of expression of characteristics are defined in the General Introduction to the Examination of Distinctness, Uniformity and Stability and the Development of Harmonized Descriptions of New Varieties of Plants, (document TG/1/3, the "General Introduction", Chapter 4.4):

2.5.3.2 Qualitative characteristics" are those that are expressed in discontinuous states (e.g. sex of plant: dioecious female (1), dioecious male (2), monoecious unisexual (3), monoecious hermaphrodite (4)). These states are self-explanatory and independently meaningful. All states are necessary to describe the full range of the characteristic, and every form of expression can be described by a single state. The order of states is not important. As a rule, the characteristics are not influenced by environment.

2.5.3.3 "Quantitative characteristics" are those where the expression covers the full range of variation from one extreme to the other. The expression can be recorded on a one-dimensional, continuous or discrete, linear scale. The range of expressions is divided into a number of states for the purpose of description (e.g. length of stem: very short (1), short (3), medium (5), long (7), very long (9)). The division seeks to provide, as far as practical, an

even distribution across the scale. The Test Guidelines do not specify the difference needed for distinctness. The states of expression should, however, be meaningful for DUS assessment.

2.5.3.4 In the case of "pseudo-qualitative characteristics" the range of expression is at least partly continuous, but varies in more than one dimension (e.g. shape: ovate (1), elliptic (2), circular (3), obovate (4)) and cannot be adequately described by just defining two ends of a linear range. In a similar way to qualitative (discontinuous) characteristics – hence the term "pseudo-qualitative" – each individual state of expression needs to be identified to adequately describe the range of the characteristic.

2.5.4 Types of scales of data

The possibility to use specific procedures for the assessment of distinctness, uniformity and stability depends on the scale level of the data which are recorded for a characteristic. The scale level of data depends on the type of expression of the characteristic and on the way of recording this expression. The type of scale may be quantitative or qualitative.

2.5.4.1 Quantitatively scaled data (metric or ordinal scaled data)

2.5.4.1.1 Introduction

2.5.4.1.1.1 Quantitatively scaled data are all data which are recorded by measuring or counting. Weighing is a special form of measuring. Quantitatively scaled data can have a continuous or a discrete distribution. Continuous data result from measurements. They can take every value out of the defined range. Discrete quantitative data result from counting.

Examples

Quantitatively scaled data	Example	Example number
- continuous	Plant length in cm.	1
- discrete	Number of stamens	2

For description of the states of expression, see Table 6.

2.5.4.1.1.2 The continuous quantitatively scaled data for the characteristic "Plant length" are measured on a continuous scale with defined units of assessment. A change of unit of measurement e.g. from cm into mm is only a question of precision and not a change of type of scale.

2.5.4.1.1.3 The discrete quantitatively scaled data of the characteristic "Number of stamens" are assessed by counting (1, 2, 3, 4, and so on). The distances between the neighboring units of assessment are constant and for this example equal to 1. There are no real values between two neighboring units but it is possible to compute an average which falls between those units.

2.5.4.1.1.4 In biometrical terminology, quantitative scales are referred to as metric scales or cardinal scales. Quantitative scales can be subdivided into ratio scales and interval scales.

2.5.4.1.2 Ratio scale

2.5.4.1.2.1 A ratio scale is a quantitative scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete.

The absolute zero point:

2.5.4.1.2.2 The definition of an absolute zero point makes it possible to define meaningful ratios. This is a requirement for the construction of index numbers (e.g. the ratio of length to width). An index is the combination of at least two characteristics. In the General Introduction, this is referred to as a combined characteristic (see document TG/1/3, section 4.6.3).

2.5.4.1.2.3 It is also possible to calculate ratios between the expression of different varieties. For example, in the characteristic 'Plant length' assessed in cm, there is a lower limit for the expression which is '0 cm' (zero). It is possible to calculate the ratio of length of plant of variety 'A' to length of plant of variety 'B' by division:

[TWC Chairperson: To review if this paragraph is relevant for DUS testing]

Length of plant of variety 'A' = 80 cm
Length of plant of variety 'B' = 40 cm
Ratio = Length of plant of variety 'A' / Length of plant of variety 'B'
= 80 cm / 40 cm
= 2.

2.5.4.1.2.4 So it is possible in this example to state that plant 'A' is double the length of plant 'B'. The existence of an absolute zero point ensures an unambiguous ratio.

2.5.4.1.2.5 The ratio scale is the highest classification of the scales (Table 2). That means that ratio scaled data include the highest information about the characteristic and it is possible to use many statistical procedures (section 2.5.7 *[cross ref.]*).

2.5.4.1.2.6 The examples 1 and 2 (Table 6) are examples for characteristics with ratio scaled data.

2.5.4.1.3 Interval scale

2.5.4.1.3.1 An Interval scale is a quantitative scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scaled data may be distributed continuously or discretely.

2.5.4.1.3.2 An example for a discrete interval scaled characteristic is 'Time of beginning of flowering' measured as date which is given as example 6 in Table 6. This characteristic is defined as the number of days from April 1. The definition is useful but arbitrary and April 1 is not a natural limit. It would also be possible to define the characteristic as the number of days from January 1.

2.5.4.1.3.3 It is not possible to calculate a meaningful ratio between two varieties which should be illustrated with the following example:

Variety ‘A’ begins to flower on May 30 and variety ‘B’ on April 30

Case I) Number of days from April 1 of variety ‘A’ = 60
Number of days from April 1 of variety ‘B’ = 30

$$Ratio_I = \frac{\text{Number of days from April 1 of variety ‘A’}}{\text{Number of days from April 1 of variety ‘B}} = \frac{60 \text{ days}}{30 \text{ days}} = 2$$

Case II) Number of days from January 1 of variety ‘A’ = 150
Number of days from January 1 of variety ‘B’ = 120

$$Ratio_{II} = \frac{\text{Number of days from January 1 of variety ‘A’}}{\text{Number of days from January 1 of variety ‘B}} = \frac{150 \text{ days}}{120 \text{ days}} = 1.25$$

$Ratio_I = 2 > 1.25 = Ratio_{II}$

2.5.4.1.3.4 It is impossible to state that the time of flowering of variety ‘A’ is twice that of variety ‘B’. The ratio depends on the choice of the zero point of the scale. This kind of scale is defined as an “Interval scale”: a quantitative scale without a defined absolute zero point.

2.5.4.1.3.5 The interval scale is lower classified than the ratio scale (Table 2). Fewer statistical procedures can be used with interval scaled data than with ratio scaled data (see Part I: section 2.5.7 *[cross ref.]*). The interval scale is theoretically the minimum scale level to calculate arithmetic mean values.

2.5.4.2 Qualitatively scaled data

Qualitatively scaled data are data which can be arranged in different discrete qualitative categories. Usually they result from visual assessment. Subgroups of qualitative scales are ordinal and nominal scales.

2.5.4.2.1 Ordinal scale

2.5.4.2.1.1Ordinally scaled data are qualitative data of which discrete categories can be arranged in an ascending or descending order. They result from visually assessed (notes) quantitative characteristics.

Example:

Qualitative data	Example	Example number
- ordinal	Intensity of anthocyanin	3

For description of the states of expressions, see Table 6.

2.5.4.2.1.2 An ordinal scale consists of numbers which correspond to the states of expression of the characteristic (notes). The expressions vary from one extreme to the other and thus they have a clear logical order. It is not possible to change this order, but it is not important which numbers are used to denote the categories. In some cases ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data (section 2.5.6 *[cross ref.]*).

2.5.4.2.1.3 The distances between the discrete categories of an ordinal scale are not exactly known and not necessarily equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale.

2.5.4.2.1.4 The ordinal scale is lower classified than the interval scale (Table 2). Less statistical procedures can be used for ordinal scale than for each of the higher classified scale data (see Part I: section 2.5.7 *[cross ref.]*).

2.5.4.2.2 Nominal scale

2.5.4.2.2.1 Nominal scaled qualitative data are qualitative data without any logical order of the discrete categories. They result from visually assessed (notes) pseudo-qualitative and qualitative characteristics.

Examples:

Qualitative data	Example	Example number
- nominal	Sex of plant	4
- nominal with two states	Leaf blade: variegation	5

For description of the states of expressions, see Table 6.

2.5.4.2.2.2 A nominal scale consists of numbers which correspond to the states of expression of the characteristic, which are referred to in the Test Guidelines as notes. Although numbers are used for designation there is no inevitable order for the expressions and so it is possible to arrange them in any order.

2.5.4.2.2.3 Characteristics with only two categories (dichotomous characteristic) are a special form of nominal scales.
2.5.4.2.2.4 The nominal scale is the lowest classification of the scales (Table 2). Few statistical procedures are applicable for evaluations (section 2.5.7 *[cross ref.]*).

2.5.4.2.2.5 The different types of scales are summarized in the following table.

Table 2: Types of scales and scale levels

[TWC Chairperson: To modify the table for consistency with the subsequent paragraphs]

Type of scale		Description	Distribution	Data recording	Scale Level
quantitative data (measured or counted)	ratio	constant distances with absolute zero point	Continuous	Absolute Measurements	High
			Discrete	Counting	
	interval	constant distances without absolute zero point	Continuous	Relative measurements	↑
			Discrete	Date	
qualitative data (visually observed QN)	ordinal	Ordered expressions with varying distances	Discrete	Visually assessed notes	↑
qualitative data (visually observed notes without logic order from PQ or QL)	nominal	No order, no distances	Discrete	Visually assessed notes	Low

2.5.4.2.2.6 From the statistical point of view a characteristic is only considered at the level of data which has been recorded, whether for analysis or for describing the expression of the characteristic. Therefore, characteristics with quantitative data are denoted as quantitative characteristics and characteristics with ordinal and nominal scaled data as qualitative characteristics.

2.5.5 Scale levels for variety description

The description of varieties is based on the states of expression (notes) which are given in the Test Guidelines for the specific crop. In the case of visual assessment, the notes from the Test Guidelines are usually used for recording the characteristic as well as for the assessment of DUS. The notes are distributed on a nominal or ordinal scale (see Part I: section 2.5.4.2 *[cross ref.]*). For measured or counted characteristics, DUS assessment is based on the recorded values and the recorded values are transformed into states of expression only for the purpose of variety description.

2.5.6 Relation between types of expression of characteristics and scale levels of data

2.5.6.1 Records taken for the assessment of qualitative characteristics are distributed on a nominal scale, for example “Sex of plant”, “Leaf blade: variegation” (Table 6, examples 4 and 5).

2.5.6.2 For quantitative characteristics the scale level of data depends on the method of assessment. They can be recorded on a quantitative (when measured) or ordinal (when visually observed) scale. For example, “Length of plant” can be recorded by measurements resulting in ratio scaled continuous quantitative data. However, visual assessment on a 1 to 9 scale may also be appropriate. In this case, the recorded data are qualitatively scaled (ordinal

scale) because the size of intervals between the midpoints of categories is not exactly the same.

Remark: In some cases visually assessed data on quantitative characteristics may be handled as measurements. The possibility to apply statistical methods for quantitative data depends on the precision of the assessment and the robustness of the statistical procedures. In the case of very precise visually assessed quantitative characteristics the usually ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data.

2.5.6.3 A pseudo-qualitative type of characteristic is one in which the expression varies in more than one dimension. The different dimensions are combined in one scale. At least one dimension is quantitatively expressed. The other dimensions may be qualitatively expressed or quantitatively expressed. The scale as a whole has to be considered as a nominal scale (e.g. "Shape", "Flower color"; Table 6, examples 7 and 8).

2.5.6.4 In the case of using the off-type procedure for the assessment of uniformity the recorded data are nominally scaled. The records fall into two qualitative classes: plants belonging to the variety (true-types) and plants not belonging to the variety (off-types). The type of scale is the same for qualitative, quantitative and pseudo-qualitative characteristics.

2.5.6.5 The relation between the type of characteristics (process level 1) and the type of scale of data recorded for the assessment of distinctness and uniformity is described in Table 3. A qualitative characteristic is recorded on a nominal scale for distinctness (state of expression) and for uniformity (true-types vs. off-types). Pseudo-qualitative characteristics are recorded on a nominal scale for distinctness (state of expression) and on a nominal scale for uniformity (true-types vs. off-types). Quantitative characteristics are recorded on an ordinal, interval or ratio scale for the assessment of distinctness depending on the characteristic and the method of assessment. If the records are taken from single plants the same data may be used for the assessment of distinctness and uniformity. If distinctness is assessed on the basis of a single record of a group of plants, uniformity has to be judged with the off-type procedure (nominal scale).

Table 3: Relation between type of characteristic and type of scale of assessed data

Procedure	Type of scale (level 2)	Distribution	Type of characteristic (level 1)		
			Quantitative	Pseudo-qualitative	Qualitative
Distinctness	ratio	Continuous	✔		
		Discrete	✔		
	interval	Continuous	✔		
		Discrete	✔		
	ordinal	Discrete	✔		
	~~combined~~	~~Discrete~~		~~✔~~	
	nominal	Discrete		✔	✔
Uniformity	ratio	Continuous	✔		
		Discrete	✔		
	interval	Continuous	✔		
		Discrete	✔		

	ordinal	Discrete	✔		
	~~combined~~	~~Discrete~~	~~✔~~		
	nominal	Discrete	✔	✔	✔

2.5.7 Relation between method of observation of characteristics, scale levels of data and recommended statistical procedures
[TWC Chairperson: To update these paragraphs in accordance with any changes to documents TGP/7 and TGP/9]

2.5.7.1 Established statistical procedures can be used for the assessment of distinctness and uniformity considering the scale level and some further conditions such as the degree of freedom or unimodality (Tables 4 and 5).

2.5.7.2 The relation between the expression of characteristics and the scale levels of data for the assessment of distinctness and uniformity is summarized in Table 6.

Table 4: Statistical procedures for the assessment of distinctness

Type of scale	Distribu-tion	Observa-tion method	Procedure[1)] and further Conditions	Reference document
ratio	continuous	MS MG (VS) [1)]	COYD Normal distribution, df >=20 long term LSD Normal distribution, df<20 2 out of 3 methods (LSD 1%) Normal distribution, df>=20	TGP/9
	discrete			
interval	continuous			
	discrete			
ordinal	discrete	VG VS	See explanation for QN characteristics in TGP/9 sections 5.2.2 and 5.2.3, See explanation for QN characteristics in TGP/9 section 5.2.4	TGP/9 TWC/ 14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete	VG (VS) [32]	See explanation for PQ characteristics in TGP/9 sections 5.2.2 and 5.2.3	TGP/9
nominal	discrete	VG (VS) [2)]	See explanation for QL characteristics in TGP/9 sections 5.2.2 and 5.2.3	TGP/9

1) see remark in section 2.5.6.2 *[cross ref.]*
2) normally VG but VS would be possible

Table 5: Statistical procedures for the assessment of uniformity

Type of scale	Distribu-tion	Observa-tion method	Procedure[1)] and Further Conditions	Reference document
ratio	continuous	MS	COYU Normal distribution	TGP/10
	discrete	MS	2 out of 3 method (s^2_c<=1.6s^2_s)) Normal distribution	
interval	continuous	VS	LSD for untransformed percentage of off-types	
	discrete			
ordinal	discrete	VS	threshold model	TWC/ 14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete		There is no case where uniformity is assessed on combined scaled data	
nominal	discrete	VS	off-type procedure for dichotomous (binary) data	TGP/10

Table 6: Relation between expression of characteristics and scale levels of data for the assessment of distinctness and uniformity

Example	Name of characteristic	Distinctness			Uniformity		
		Unit of assess-ment	Description (states of expression)	Type of scale	Unit of assess-ment	Description (states of expression)	Type of scale
1	Length of plant	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data
					True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
2	Number of stamens	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data
3	Intensity of anthocyanin	1 2 3 4 5 6 7 8 9	very low very low to low low low to medium medium medium to high high high to very high very high	ordinally scaled qualitative data (with an underlying quantitative variable)	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
4	Sex of plant	1 2 3 4	dioecious female dioecious male monoecious unisexual monoecious hermaphrodite	nominally scaled qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data

Example	Name of characteristic	Distinctness			Uniformity		
		Unit of assess-ment	Description (states of expression)	Type of scale	Unit of assess-ment	Description (states of expression)	Type of scale
5	Leaf blade: variegation	1 9	absent present	nominally scaled qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
6	Time of beginning of flowering	date	e.g. May 21, 51st day from April 1	interval scaled discrete quantitative data	Date	e.g. May 21, 51st day from April 1	interval scaled discrete quantitative data
					True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
7	Shape	1 2 3 4 5 6 7	deltate ovate elliptic obovate obdeltate circular oblate	combination of ordinal and nominal scaled discrete qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
8	Flower color	1 2 3 4 5 6 7 8 9 10	dark red medium red light red white light blue medium blue dark blue red violet violet blue violet	combination of ordinal and nominal scaled discrete qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data

3 CONTROL OF VARIATION DUE TO DIFFERENT OBSERVERS

[To be developed on the basis of sections I and II of document TWC/25/12] [f]

[The TWV noted that it had encouraged the development of that section and agreed that it should provide suitable text for aspects which were not adequately covered in document TWC/25/12.

TWC: Mr. Gerie van der Heijden (Netherlands) will consult his Naktuinbouw colleagues in the Netherlands to see if they could contribute a draft for this section.]

6 DATA PROCESSING FOR THE ASSESSMENT OF DISTINCTNESS AND FOR PRODUCING VARIETY DESCRIPTIONS

[The TWC agreed that the information provided in TWC/26/15 and TWC/26/23, presented by Mr. Vincent Gensollen (France) and Mr. Uwe Meyer (Germany), respectively, and an oral presentation by Ms. Mariko Ishino (Japan) included in document TWC/26/15 Add. provided valuable guidance on data processing for the assessment of distinctness and for producing variety descriptions and noted that UPOV did not have guidance on that matter in the TGP documents. It agreed that a new section should be created in document TGP/8/1, Part I as "Data processing for the assessment of distinctness and for producing variety descriptions"]

SECTIONS FROM TGP/8/1 DRAFT 11 PART II FOR FURTHER DEVELOPMENT

TABLE OF CONTENTS PAGE

PART II: TECHNIQUES USED IN DUS EXAMINATION

3.4 SECTION ON SINGLE GROWING CYCLE METHOD~~S~~

~~3.4.1 Single Growing Cycle Method~~

~~[TWP's are invited to provide information on this method.]~~

~~3.4.1.1 In the absence of information on this method it is thought likely that the following applies:~~

- ~~For two varieties to be distinct using the Single Growing Cycle method, the varieties need to be significantly different at a given significance level in one or more measured characteristics. Differences can be assessed using a statistical test based on a two-tailed LSD to compare the variety means with standard errors calculated using the plot residual mean square from the analysis of the variety x replicate plot means as the estimate of random variation (s^2).~~

- ~~The source of variation used to estimate random variation (s^2) in the LSD determines what can be inferred from using the LSD. The Single Growing Cycle method estimates random variation (s^2) in the LSD using the plot residual mean square, which represents the plot-to-plot variation within a variety (allowing for any block effects if blocks are present). As a result, using this LSD the crop expert can conclude that varieties with significantly different means are different relative to the plot-to-plot variation within a variety.~~

3.5 STATISTICAL METHODS FOR VERY SMALL SAMPLE SIZES[h]

Note

The TC agree to invite the Technical Working Parties to consider including statistical methods for very small sample sizes, subject to suitable methods which are in use by members of the Union being provided.

The TC requested that for each statistical method an explanation of the requirements for its application and the situations where it would be appropriate to apply the method be included

3.5.1 One of the main problems when applying a statistical test on small trials is that we do not have enough data available to limit the risk of making a wrong decision to an acceptable level. Every statistical test has a probability/risk of making wrong decisions: there is a Type I error, i.e. the risk of declaring two varieties different where in reality they are not significantly different, and a Type II error: declaring two distinct varieties not significantly different.

3.5.2 In general we control the Type I error by fixing the significance level (α). However, especially with small trials, a low risk of Type I (low α) considerably increases the Type II error, or alternatively stated, such a test has a considerable lack of discriminating power. Another problem with small sample sizes is that we do not have enough data to test our assumptions.

3.5.3 From a statistical point of view it is possible to statistically compare the mean of a candidate variety after a single measurement on a single plant in a single year with a set of reference varieties, if at least several reference varieties are being measured in the same year as well as in one or more other years. For this, one could use any statistical package capable of analysing unbalanced two-way designs with the factors years and varieties. This analysis can be seen as an extension of the long-term LSD but is not standard UPOV practice. The test is based on the usual assumptions, which can however not be tested with such a small dataset. If we are willing to accept assumptions like normality, homogeneity of variance and additivity, e.g. from previous knowledge, the test is in principal valid, although lack of power is still a problem.

3.5.4 In general, small sample size may refer to different aspects of the variety trial:

(a) limited number of plants/measurements in a plot,
(b) limited number of replications,
(c) limited number of varieties,
(d) limited number of years,

or any combination of these aspects.

3.5.5 Ad (a). For any experiment, sound experimental design principles should be kept in mind at all times. With regard to the number of plants per plot, it is bad practise to use so few plants in a plot that measured plants are considerably influenced by their neighbours. A plant of a small variety next to a plant of a tall variety may lead to both plants having a more extreme expression than under the condition of neighbouring plants of similar height. This interaction effect hampers unbiased comparisons. To overcome this neighbouring effect, one often uses border plants. Alternatively one can group varieties in different height classes such

that these effects are minimised within the groups. Also refer to document TGP/8 Part 1, section 1.6.3.7 for further details.

Ad (b). The number of replications in a trial is often at least 2. Strictly speaking, for the COYD or long-term LSD we only use the variety means of the year for the analysis, so from a theoretical point of view a single replication per variety per year is sufficient. Of course having no replications within a year may lead to a significant increase of the uncertainty of the estimate of the variety mean and it limits the testing of assumptions for the analysis.

Ad (c). With regard to the number of varieties in the test, from a theoretical point as few as three or four varieties are sufficient if two or three years of data are used. However, in most cases, experience has taught us that such small experiments with just a few degrees of freedom are not really useful, as the discriminating power of the test is too low. A low power may be less of a problem, if we have just a few varieties and large and consistent differences between them.

Ad (d). Theoretically spoken, it is possible to make a decision based on a single year's observation of a candidate variety, when reference varieties are also observed and data from the reference varieties over several years are available. Several assumptions need to be made and these assumptions can not be tested. An important assumption is that the candidate variety to be tested does not exhibit a strong interaction from year to year with close reference varieties for the characteristic under study. However, the most important drawback is that the power of the test is very limited, i.e. the chance that a truly significant difference between a pair of varieties will indeed be declared significant in the analysis is very small. In that case, the conclusion would be that the two varieties are not sufficiently different to obtain a significant result given the small sample size. If this information is sufficient for rejection of the candidate variety is an open question, but probably not.

3.5.6 Historical data can be used to gain insight in the lack of power of the experiment, i.e. the risk of accidentally rejecting a distinguishable variety. One can also use these data to get an impression of the best way to improve the experimental design.

3.5.7 The power of the test can be increased in several ways. If a reference variety is not tested in the same years as the candidate variety, the standard error of this difference is rather large. By putting the varieties in the same trial in the next year, the standard error for this difference can be reduced considerably.

3.5.8 Another way to increase the power of the test is by increasing the number of degrees of freedom for the residual term. This can be done by using more data from previous years, which is exactly what is done in the long-term LSD.

3.5.9 Note that small trials are troublesome for distinctness testing, but even more so for uniformity testing. The COYU requires a considerable number of plants per plot for a reasonable estimate of the standard deviation.

3.5.10 Another problem when we use small and unbalanced designs is that some variety differences are tested with greater power than others. The comparison of candidate varieties with reference varieties which are less frequent (or even absent) in the years of testing of the candidates will have a much larger standard error of difference. This might lead to rejecting a candidate which can not be declared sufficiently distinct, but which is due to bad luck since it is close to a reference not in the collection of reference varieties on the field. The procedure

is in itself statistically valid and sound, but might be unwanted from a fair policy point of view.

5. EXAMINING DUS IN BULK SAMPLES

5.1 Introduction and abstract

In some crops samples are bulked before certain characteristics are examined. The term "bulk sampling" is used here for the process of merging some or all individual plants before recording a characteristic. There are different degrees of bulking ranging from: (1) merging pairs of plants, (2) merging 3 or 4 up to all plants within a plot up to (3) merging all plants within a variety. The degree of bulking may play an important role in the efficiency of the tests. Bulking is usually only applied where the measurement of the characteristic is very expensive or very difficult to obtain for individual plants. Some examples are seed weight in cereals and peas and beans, and erucic acid content in rapeseed. This section describes some of the consequences of bulk sampling. It is shown that the test of distinctness (using COYD, see Part II: Section 3.2 *[cross ref.]*) may be expected to be relatively insensitive to the degree of bulking, but that the efficiency of the tests for uniformity (using COYU, see Part II: Section 4.2 *[cross ref.]*) must be expected to decrease when the data are bulked. The COYU test for uniformity cannot be carried out if all plants within a plot are bulked.

5.2 Distinctness

5.2.1 In the COYD method for examining distinctness the basic values to be used in the analyses are the annual variety means. As bulk sampling also gives at least one value for each variety per year, it will usually still be possible to use the COYD method for distinctness purposes for any degree of bulking, as long as at least one value is recorded for each variety in each year and that the bulk samples are representative for the variety. However, some problems may be foreseen: the assumption of data being normal distributed may be better fulfilled when the mean of many individual measurements are analyzed instead of the mean of fewer measurements or, in the extreme, just a single measurement.

5.2.2 The efficiency of the test of distinctness may be expected to be lower when based on bulked samples than when it is based on the mean of all individual plants in a year. The loss will be from almost zero upwards, depending on the importance of the different sources of variations. The variation which is relevant for the efficiency of variety comparisons is formulated in the following model:

$$\sigma^2_{total} = \sigma^2_{vy} + \sigma^2_{p} + \sigma^2_{i} + \sigma^2_{m}$$

where

σ^2_{total} is the total variance of a characteristic used for comparing varieties. The total variance is regarded as being composed of four sources of variation:

1: σ^2_{vy} the variance component due to the year in which the variety is measured

2: σ^2_{p} the variance component due to the plot in which the measurement was taken

3: σ^2_{i} the variance component due to the plant on which the measurement was taken

4: σ^2_{m} the variance component due to the inaccuracy in the measurement process

5.2.3 In cases where the data are not bulked the variance of the difference between two variety means, σ^2_{diff} , becomes:

$$\sigma^2_{diff} = 2\left\{\frac{\sigma^2_{vy}}{a} + \frac{\sigma^2_p}{ab} + \frac{\sigma^2_i}{abc} + \frac{\sigma^2_m}{abc}\right\}$$

where

a is the number of years used in the COYD method

b is the number of replicates in each trial

c is the number of plants in each plot

5.2.4 Assuming that each bulk sample has been composed in such a way that it represents an equal amount of material from all the individual plants which have been bulked into that sample, the variance between two varieties based on *k* bulked samples (each of *l* plants) becomes:

$$\sigma^2_{diff} = 2\left\{\frac{\sigma^2_{vy}}{a} + \frac{\sigma^2_p}{ab} + \frac{\sigma^2_i}{abkl} + \frac{\sigma^2_m}{abk}\right\}$$

where

k is the number of bulk samples

l is the number of plants in each bulk sample

5.2.5 Thus if all plants in each plot are divided in *k* groups of *l* plants each and an average measurement is taken for each of the k groups, then only the last term in the expression for σ^2_{diff} has increased (as *kl* is equal to *c*). For many characteristics it is found that the variance caused by the measurements process is small and hence the bulking of samples will only have a minor effect on the conclusions reached by the COYD method. Only if the variance caused by the measurement process is relatively large can bulking have a substantial effect on the distinctness tests using COYD.

Example 1

Variances for comparing varieties were estimated (by the use of estimated variance components) for different degrees of bulking. The calculations were based on the weight of 100 seeds of 145 pea varieties grown in Denmark during 1999 and 2000. In this example, the contribution to the variance caused by the measurement process was relatively very small, which means that bulking will have a low influence on the test for distinctness. In a 3 year test with 30 plants in each of 2 blocks, the variance of a difference between two varieties was estimated to be 2.133 and 2.135, for no bulking and a single bulk sample per plot, respectively.

For other variables the variance component due to the measurement process may be relatively more important. However, it is likely that in most practical cases this variance component will be relatively small.

5.2.6 In some cases each bulk sample is not drawn from a specific set of plants (say, plant 1 to 5 in bulk sample 1, plant 6 to 10 in bulk sample 2 etc.), but bulk samples are formed from mixed samples of all plants in a plot. This means that different bulk samples may contain material from the same plants. It must be expected that similar results apply here, although, in this situation, the effect of bulking may have an increased effect because there is no guarantee that all plants will be equally represented in the bulk samples.

5.3 Uniformity

5.3.1 Bulking within plot

5.3.1.1 In COYU the test is based on the standard deviation of the individual plant observations (within plots) as a measurement of uniformity. The log of the standard deviations plus one are analyzed in an over-years analysis; i.e. the values $Z_{vy} = \log(s_{vy} + 1)$ are used in the analyses. The variance on these Z_{vy} values can be regarded as arising from two sources, a component that depends on the variety-by-year interaction and a component that depends on the number of degree of freedom used for estimating the standard deviation, s_{vy} (the fewer degrees of freedom the more variable the standard deviation will be). This can be written (note that the same symbols as used in the distinctness section will be used here with different meaning):

$$Var(Z_{vy}) = \sigma_{vy}^2 + \sigma_f^2$$

where this variance can be regarded as being composed of two sources of variation:

1: σ_{vy}^2 the variance component due to the year in which the variety is measured

2: σ_f^2 the variance component due to the number of degrees of freedom used in estimating s_{vy}

σ_f^2 is approximately $\frac{1}{2v}\left(\frac{\sigma}{\sigma+1}\right)^2$ when the recorded variable is normally distributed and the standard deviations do not vary too much. This last expression reduces to $0.5/v$ when $\sigma >> 1$. Here σ is the mean value of the s_{vy} values and v is the number of degrees of freedom used in the estimation of s_{vy}.

5.3.1.2 The variance caused by the year in which the variety is measured may be assumed to be independent of whether the samples are bulked or not, whereas the variance caused by the number of degrees of freedom will be increased when bulked samples are used because a lower number of degrees of freedom is available.

5.3.1.3 The variance of a difference between a Z_{vy} for a candidate variety and the mean of the reference varieties' Z_{vy} values may be written:

$$\sigma_{diff}^2 = \left(\sigma_{vy}^2 + \sigma_f^2\right)\left(\frac{1}{a} + \frac{1}{ar}\right)$$

where

a is the number of years used in the test

r is the number of reference varieties

Example 2

The effect of bulking in the test for uniformity, an estimate was made using the same data as for Example 1 I Part II, section 5.2.5 [cross ref.]. For a test using 50 reference varieties in 3 years with 30 plants per variety in each of 2 plots per trial the variance for comparing the Z_{vy} value for a candidate variety and the mean of the reference varieties' Z_{vy} will be 0.0004 if no bulking is done. This can be compared to 0.0041, 0.0016 and 0.0007 when 2, 4 and 10 bulk samples per plot were used. Thus, in this example, the effect of bulking has a great influence on the test for uniformity. The variance increased,

approximately by a factor of 10 when changing from individual plant records to just 2 bulk samples per plot. This means that the degree of non-uniformity must be much higher for it to be detected when 2 bulk samples are used instead of individual plant records.

5.3.2 Bulking across plots

Bulking across plots means that part of the between plot (and block) variation will be included in the estimated standard deviation between bulked samples. If this variation is relatively large it will tend to mask any differences in uniformity between varieties. In addition some noise may also be added because the ratio of material from the different plots may vary from bulk to bulk. Finally the assumptions for the present recommended method, COYU, may not be fulfilled in such cases. Therefore it is recommended to bulk only within plots.

5.3.3 Taking just one bulk sample per plot

In general, if all plants in a plot are bulked such that only a single sample is available for each plot, it becomes impossible to calculate the within plot variability and in such cases no tests for uniformity can be performed. In rare cases, where non-uniformity may be judged from values that can only be found in mixtures, non-uniformity may be detected even where a single bulk sample for each plot is used. For example, in the characteristic "erucic acid" in oil seed rape, values between 2% and 45% can only arise because of a lack of uniformity. However this only applies in certain special cases and even here the non-uniformity may only show up under certain circumstances.

6. EXAMINATION OF CHARACTERISTICS USING IMAGE ANALYSIS[1]

6.1. Introduction

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.

6.2. Combined characteristics

6.2.1 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'."

6.2.2 Thus, the General Introduction clarifies that the use of image analysis is one possible method for examining characteristics which fulfil 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.

6.3. Guidance on the use of image analysis

[to be developed by the Technical Working Party on Automation and Computer Programs (TWC)]

[The TWC, at its Twenty-sixth Session, agreed as follows:

(a) for existing characteristics: to explain the need to compare the results of the characteristics examined by the old method and by image analysis. The TWC noted that it might, in some cases, lead to a modification of the existing characteristic, in which case it would be necessary for the Test Guidelines to provide a clear definition of the characteristic, including an outline of the algorithm which defined the characteristic;

(b) for new characteristics: to provide guidance on the need to meet the requirements for a characteristic to be used for DUS, as set out in the General Introduction, and the need to check for independence from other characteristics, in the same way as for other characteristics

In response to an observation from an expert from China, the TWC agreed that the guidance to be developed in document TGP/8 on image analysis should provide guidance on how to consider calibration of images, particularly images containing more than one object, to account for the differing distances of the objects from the camera.]

The TWC also agreed that Mr. Gerie van der Heijden (Netherlands) should prepare a draft text for Section III, Subsection 3, taking into account the comments made above.]

[the TWA, at its thirty seventh session, agreed that for existing characteristics: to explain the need to compare the results of the characteristics examined by old method and by image analysis; for new characteristics: to provide guidance on the need to meet the requirements for a characteristic to be used for DUS, as set out in the General Introduction, and the need to check for independence from other characteristics]

7. METHODS FOR DATA PROCESSING FOR THE ASSESSMENT OF DISTINCTNESS AND FOR PRODUCING VARIETY DESCRIPTIONS

[The TWC agreed that the information provided in documents TWC/26/15 and TWC/26/23, presented by Mr. Vincent Gensollen (France) and Mr. Uwe Meyer (Germany), respectively, and an oral presentation by Ms. Mariko Ishino (Japan) included in document TWC/26/15 Add. provided valuable guidance on data processing for the assessment of distinctness and for producing variety descriptions and noted that UPOV did not have guidance on that matter in the TGP documents. It agreed that a new section should be created in document TGP/8/1, Part I as "Data processing for the assessment of distinctness and for producing variety descriptions for producing variety descriptions" and that the methods used by France, Germany and Japan should be included in a new section in document TGP/8/1, Part II as "Methods for data processing for the assessment of distinctness and for producing variety descriptions. [..]The TWC agreed that Finland, France, Germany, Japan, Kenya and the United Kingdom should prepare information on their methods for inclusion in the next draft of document TGP/8]

7.1 Handling measured, quantitative characteristics for vegetable and herbage crops tested in the United Kingdom

7.1.1 This document provides an explanation of how measured, quantitative characteristics are handled and used to develop variety descriptions in the United Kingdom for vegetable and herbage crops.

7.1.2 In vegetable and herbage crops, which are mostly cross-pollinated except for pea which is self-pollinated, the trials are conducted according to the UPOV Test Guidelines.

7.1.3 For the measured, quantitative characteristics, as part of the determination of distinctness, COYD is applied on the original scale of the characteristics.

7.1.4 To develop variety descriptions, over-year variety means are calculated on the original scale of the characteristics. These over-year means are then converted to notes.

7.1.5 For each crop the over-year variety means of the varieties in trial are calculated from their yearly means in trials. For herbage crops the past 10 years are used, whereas for vegetable crops all years are included in which the reference collection varieties have been tested. As not all varieties are present in all years, a fitted constants analysis is used to adjust the over-year means for the different years varieties were present in. This is done using the DUSTNT module FITC in conjunction with the module FIND.

7.1.6 The over-year means are converted to notes using the DUSTNT module VDES. This permits two methods of division of the range of expression into states and notes as follows:-

a) By division of the range of expression of the over-year means for the reference collection varieties into equal-spaced states. The number of states is as given in the UPOV Test Guideline.

b) By use of delineating varieties to divide the range of expression into states.

7.1.7 For vegetable crops excluding potato method (a) is used to divide the range of expression into states and notes, and for herbage crops method (b) is used.

7.1.8 For herbage crops the DUSTNT module SAME is used to check whether there are varieties with the same variety description.

7.1.9 For herbage crops the DUSTNT module MOST, is used in conjunction with the modules SSQR and DIST to find most similar varieties based on multivariate distances.

7.2 DETERMINATION OF NOTES FOR MEASURED CHARACTERISTICS OF CROSS-POLLINATED SPECIES

7.2.1 The method how measured, quantitative characters of cross-pollinated species are transformed into notes in Finland is described in this document.

7.2.2 In Finland, the Combined Over Years method (DUST Package) is used to assess the distinctness and uniformity of measured characters of cross-pollinated plants. The specific Test Guidelines indicates the required amount of individuals in the test. Usually measurements of quantitative characters are done from 60 single plants.

7.2.3 If the candidate variety fulfills the DUS criteria, the transformation of characters into notes is done for the variety description. The transformation of characters to notes is done for each year separately by using the least significant difference LSD 1% value from the DUST9 module of the single years test. The final note is the fusion of these values attained from two or three testing periods.

7.2.4 The value for LSD 1% is considered as a two note difference. Two note difference is considered as a clear difference in 'General introduction'. This rule is advised for interpretation of observations of quantitative characteristics without the application of statistical methods. This principle is applied here, though it is not an absolute standard. If LSD 1% would be used as a one note difference, the width of one note would be two times wider and the values would mostly be five or close to it. LSD 1% value is divided by two to get the one note 'width' for counting the scale.

7.2.5 The variety which is located at the middle of the ranked variety list is used to present the note five. By using this value as an anchor point for the scale, the limits for other notes are counted. It is important that first the note five 'spreads' over the key value, because this value is considered as a middle point of the scale.

7.2.6 As an example, the determination of notes for the characteristic length of longest stem in timothy (*Phleum pretense*, UPOV TG/34/6 characteristic no. 9.) is presented. In the scale for Growing cycle 1, the ranked value for note five is 1131.75 mm. The limits for note five are from 1105.68 mm to 1157.83 mm (span of one note is 52.15 mm which is spread over value 1131.75, i.e. 26.075 mm on both sides). Three candidate varieties and three reference varieties are included in the example. In Table A, all the reference varieties are presented.

TABLE A

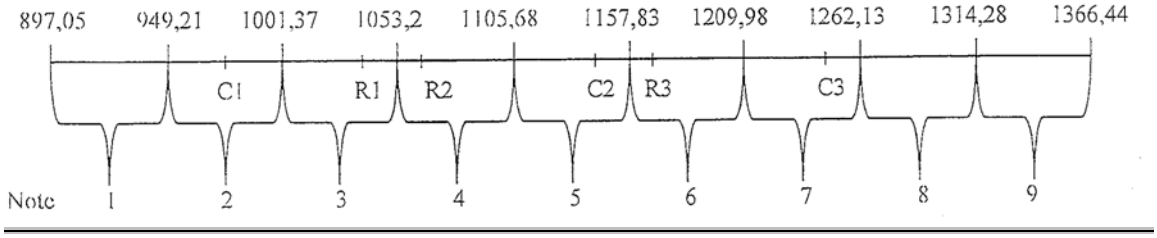
Means and notes of the character length of longest stem for 58 reference varieties and 3 candidate varieties. Notes are given according to the scales from Figure 1.

Variety	mean cycle 1	mean cycle 2	note cycl1	note cycle2	final note
Alexander	1187.963	1079.937	6	6	6
Bilbo	1184.482	1073.060	6	6	6
Comtal	1182.298	1121.723	6	6	6
Haukila	1193.220	1124.715	6	6	6
Hja Tiiti	1177.401	1099.171	6	6	6
Niilo	1163.676	1087.547	6	6	6
Nokka	1169.178	1084.585	6	6	6
Saga	1162.915	1115.309	6	6	6
Saguenay	1241.350	1075.108	7	6	6
Candidate3	1247.418	1085.617	7	6	6
Szarvasi-60	1160.033	1102.463	6	6	6
Alma	1129.250	1055.518	5	5	5
Barmidi	1145.067	1009.506	5	5	5
Billy	1140.145	1107.154	5	6	5
Bodin	1157.192	1080.214	5	6	5
Candidate2	1142.896	1052.053	5	5	5
Bottnia II	1120.209	1074.090	5	6	5
Carola	1202.265	1031.624	6	5	5
Comer	1149.167	1022.001	5	5	5
Engmo	1105.789	1079.133	5	6	5
Erecta	1113.320	1090.889	5	6	5
Goliath	1073.732	1002.738	4	5	5
Grindstad	1064.367	1017.783	4	5	5
Iki	1122.657	1130.778	5	6	5
Jonatan	1153.056	1092.797	5	6	5
Jouliette	1155.228	1096.119	5	6	5
Jögeva 54	1179.194	1067.970	6	5	5
Kämpe II	1141.096	1039.338	5	5	5
Linus	1126.054	1081.078	5	6	5
Noreng	1064.765	1032.674	4	5	5
Otto	1147.578	1095.721	5	6	5

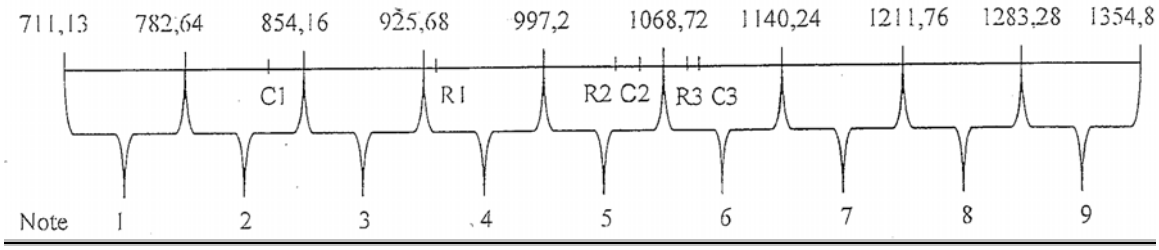
Variety	mean cycle 1	mean cycle 2	note cycle1	note cycle2	final note
Promesse	1106.941	1028.101	5	5	5
Ragnar	1148.307	1075.430	5	6	5
Sobol	1195.870	1051.092	6	5	5
Snorri	1139.985	1085.358	5	6	5
SW Janus	1064.656	1057.217	4	5	5
Tammisto II	1169.548	1027.743	6	5	5
Tarmo	1119.613	997.216	5	5	5
Tenho	1151.037	999.464	5	5	5
Tika	1154.765	1017.026	5	5	5
Topas	1170.965	1017.741	6	5	5
Tundra	1118.927	973.412	5	4	5
Turku	1136.744	1008.277	5	5	5
Tuukka	1136.614	1035.912	5	5	5
Tuure	1116.041	940.266	5	4	5
Tryggve	1117.383	1044.938	5	5	5
Uula	1071.082	1032.955	4	5	5
Vähäsöyrinki	1152.661	1055.362	5	5	5
Argus	1065.424	949.176	4	4	4
Farol	1050.870	946.707	3	4	4
Forus	1085.912	1017.135	4	5	4
Jari	1061.794	961.897	4	4	4
Liglory	1048.505	952.761	3	4	4
Liscka	1095.313	987.829	4	4	4
Nuutti	1033.737	929.759	3	4	4
Peti	1090.900	926.385	4	4	4
Phlewiola	1040.769	963.806	3	4	4
Tammisto	1096.183	979.941	4	4	4
Tia	1063.439	996.023	4	4	4
Vega	1084.838	995.675	4	4	4
Candidate 1	979.862	839.060	2	2	2

Figure 1. Scales for two different growing cycles for the characteristic length of longest stem in timothy. The situation of three candidate (C) and reference (R) varieties are indicated.

Growing cycle 1 (scale in mm)



Growing cycle 2



7.2.7 Table 1 presents the different values for the calculation of the scale for both growing cycles. In the Growing Cycle 1 the LSD 1% value was 104.31 mm and in the Cycle 2 143.02 mm. This variation is due to environmental effect (e.g. different water or temperature conditions during growing periods, variation in the soil). Mean value for note 5 is 10 cm longer in the Growing cycle 2. Also the width of one note is 2 cm longer in the Growing cycle 2.

Table 1. LSD 1% values, width of one note and ranked value for the note five for two testing years.

	Growing cycle 1	**Growing cycle 2**
LSD 1% (mm)	104.31	143.02
width of one note (mm)	52.15	71.51
ranked value for note 5 (mm)	1131.75	1032.96

7.2.8 Different conditions in different testing periods cause variation to the variety means and to the LSD values. If two or three different testing periods give different notes for a character (as in Table 2 for candidate 3, growing cycle 1: 7, cycle 2: 6), the 'fusion' of notes is done towards the value 5. Therefore in the case of candidate 3, the final note is 6. If the data is obtained from three years and there is variation in the notes, the fusion is done similarly towards note 5. For example 5, 5 and 7 are transferred to 5, 6, 6 (1 note is given from 7 to 5) and the final note is therefore 6, which is the most abundant note. If there is an obvious reason for the odd note, for example extreme conditions during the growth period or severe area of testing field, it may be neglected.

Table 2. Means and notes of the character length of longest stem in timothy for three candidate and three reference varieties in two different growing cycles. Notes for the characters are given according to the scales in Figure 1.

	Mean for cycle 1	**Mean for cycle 2**	**Note for cycle 1**	**Note for cycle 2**	**Final Note**
Candidate 1	979.86	839.06	2	2	2
Candidate 2	1142.9	1052.05	5	5	5
Candidate 3	1247.42	1085.62	7	6	6
Reference 1	1033.74	929.76	3	4	4
Reference 2	1064.37	1017.78	4	5	5
Reference 3	1169.18	1084.59	6	6	6

7.2.9 In Table 2, Candidate variety 1 is considered as note 2 which means very short to short. This variety, being the shortest one from year to year, could be used as an example variety for this character. The use of example varieties for the determination of notes is difficult for this character, because most of the varieties tend to get the same value. In this timothy example 60% of the varieties have the value 5 for the character length of longest stem (see Table A). Also the continuous variation in the character makes it difficult to judge the note in the field.

Conclusions

7.2.10 This method provides an objective way to transform measured characters into notes for each individual year separately based on 1% LSD value and the ranking list of varieties. The final note is the fusion of these notes from individual years. This method is suitable for species where example varieties are difficult to use for character determination.

7.3 The Method to adjust the Table of Assessment for Quantitative Characteristics

Japan
National Center for Seeds and Seedlings (NCSS)

TABLE OF CONTENTS

1. Introduction

1.1 This provides an explanation of the Japanese methods to adjust the table of assessment for quantitative characteristics in characteristics table of TG.

1.2 The method is based on the premise as below.

a) This method is mainly used for ornamental plants and vegetable crops.

b) Basically, DUS growing trial for ornamental plants and vegetable crops is assessed in two independent growing cycles. When we decide it is satisfactory for the assessment of DUS, further growing trial will not be done. This document explains the adjusting method of the quantitative characteristics from the result of DUS growing trial of one growing cycle.

c) The term “the table of assessment” means the table to evaluate the notes from the data of quantitative characteristics.

2. Method with the Fundamental Table of Assessment (FAT)

2.1 [Background]

2.1.1 For the assessment of note in most quantitative characteristics, the relative assessment based on the data of the example variety in one time seems to be general method. Especially when we start DUS growing trial about new species, we use this method. But, we seek more effective method to reduce the yearly variation for concerned species which we have examined for many years.

.

2.1.2 The method with FAT is used for this purpose. We make FAT as the adjustable base only for the species that had examined in sufficient number of DUS growing trials. FAT is adjusted every year to correct yearly variations of data.

2.2 [What is FAT?]

2.2.1 FAT is the table of assessment that made from the enough experimental data about the species. In the concrete, one of the experimental data is “Proposition by experts”. It is the table that is based on the expert’s experience and knowledge, and the table covers the full ranges of variations that the species or variety groupings show under the normal growth. The other of the experience is “Accumulated statistical data.” It is the data accumulated about several example varieties in sufficient number of DUS growing trials. We try to accumulate the data from sufficient number of growing trials. But it needs long time to accumulate the data in one site for many times. Before we get enough data to make FAT, we set the notes based on example variety’s data from one growing trial and our experiences. If we estimate the data accumulated in certain place for one species are enough stable, we make FAT based on the data. FAT is available only for species that had examined for sufficient experience of DUS growing trial about several example varieties.

2.3 [Composition of FAT]

2.3.1 Table 1 shows the part of example FAT, the characteristic “length of leaf blade”. There are nine notes. In the note 5,
Range : 70-79 mm
Interval : 10 mm,
Median : 75 mm
Standard example variety of the note 5 : ‘EV-B’

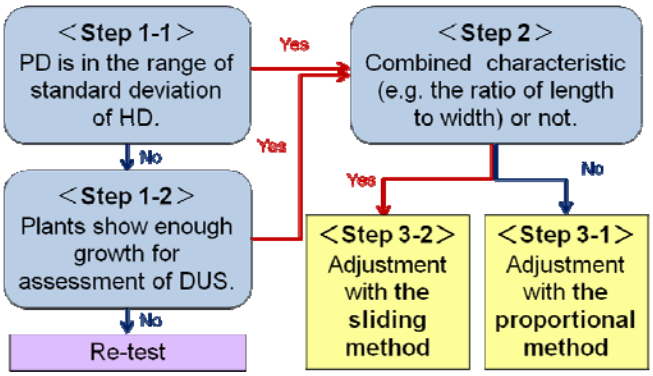
Table 1: Example FAT for the characteristic “ length of leaf blade”

Characteristics	Note	1	2	3	4	5	6	7	8	9
Length of leaf blade (mm)	Range	~ 39	40 ~ 49	50 ~ 59	60 ~ 69	70 ~ 79	80 ~ 89	90 ~ 99	100 ~ 109	110 ~
	Interval		10	10	10	10	10	10	10	
	Median		45	55	65	75	85	95	105.	
	Example variety			EV -A		EV -B				

2.4 [Practical adjusting methods for use of FAT]

2.4.1【 Overview of the methods】

2.4.1.1 There are two methods in adjustment of FAT. One is the proportional method, the other is the sliding method. PD indicates Present data, the data of the example variety measured in this time. HD indicates Historical data, the mean of the data of the example variety measured in sufficient times of DUS growing trial.



**PD: Present data = The data of Example Variety measured in this time*
HD: Historical data = Mean of the Data of Example Variety measured in sufficient number of DUS growing trial
Fig. 1: Flow chart of the practical adjusting method with FAT

2.4.1.2 Figure 1 shows the practical adjusting method.
Step 1-1: Check whether PD is in the range of standard deviation of HD
Step 1-2: Check whether plants show satisfactory growth for assessment of DUS
Step 2 : Check whether the characteristic is combined characteristic or not.
Step 3-1: Adjustment FAT with the proportional method
Step 3-2: Adjustment FAT with the sliding method

2.4.2【 Step 1-1: Check whether PD is in the range of standard deviation of HD】

2.4.2.1 We confirm the example variety’s normal growth by checking step 1-1. If step 1-1 is not satisfied, we should check whether the growing trial can be done reasonably and properly or not.

2.4.2.2 The examples are as follows.
Characteristic “length of leaf blade”
HD: 74.0mm
Standard deviation: 5.01
Range of the standard deviation: 69.0-79.0mm

2.4.2.2.1 If PD is 70.3mm, PD is in the range of standard deviation of HD. → Go to step 2

2.4.2.2.2 If PD is 83.6mm, PD is out of the range of standard deviation of HD. → Go to step 1-2.

2.4.3【 Step 1-2:Check whether plants show satisfactory growth for assessment of DUS】

2.4.3.1 The purpose of step 1-2 is to check whether the growing trial can be done reasonably and properly or not.

2.4.3.2 If the example variety we expect to use for adjustment doesn’t show satisfactory growth, we can use another example variety (which shows satisfactory growth and has enough experimental data) for adjustment of FAT. In this case, we estimate plants in this growing trial shows satisfactory growth for evaluation of DUS.→ Go to step 2

2.4.3.3 In the case other varieties also show unusual growth, we should try to make clear the reason with assistance of the plant species expert. After taking into account the distance from the range of standard deviation of HD and the advice of our expert and examiner, we estimate whether we can evaluate DUS in this growing trial.
We can evaluate DUS.→ Go to step 2
We can’t evaluate DUS. → Re-test

2.4.4【 Step 2: Check whether the characteristic is combined characteristic or not】

2.4.4.1 The purpose of step 2 is to decide which method, the proportional method or the sliding method, is more suitable for the characteristic. In the proportional method, range and interval of notes are adjusted at once. In the sliding method, range is adjusted on the one hand and interval is not changed. It means that the proportional method is not suitable for the characteristics that need fixed interval. In the concrete, the combined characteristics are generally stable than other characteristics and they need fixed interval. In such case, the sliding method is applied.

2.4.4.2 Characteristic “length of leaf blade”
It is not the combined characteristic. → Go to step 3-1

2.4.4.3 Characteristic “Leaf: ratio length/width”
It is the combined characteristic. → Go to step 3-2

2.4.5【 Step 3-1: Adjustment FAT with the proportional method】

2.4.5.1 We calculate the proportion of the measured data in this time to the mean of the historical data about an example variety. FAT multiplied by the proportion gives the adjusted table of assessment in this time.

2.4.5.2 The examples are as follows.
Characteristic “ length of leaf blade”

PD: 70.3mm
HD: 74.0mm
Proportion (PD/HD) =0.95

2.4.5.3 The upper line of Figure 2 is FAT expressed in a number line. FAT multiplied 0.95 gives the adjusted table of assessment of this time, the lower line.

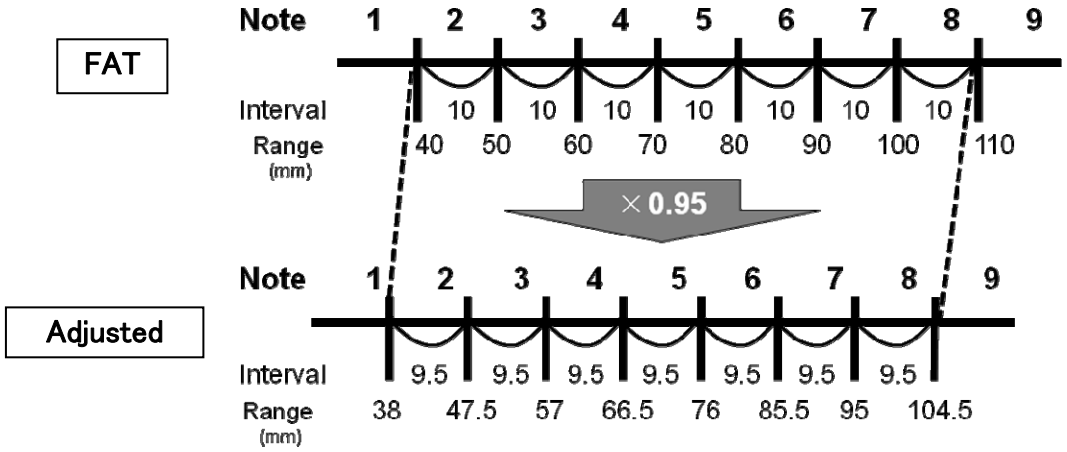


Fig.2: Adjustment FAT with the proportional method

2.4.5.4 We take the note 5 as an example,
The minimum of the range is 70. 70 multiplied by 0.95 make 66.5.
The maximum of the range is 80. 80 multiplied by 0.95 make 76.
The interval of the note 5 changes from 10 to 9.5.

2.4.6【 Step 3-2: Adjustment FAT with the sliding method】

2.4.6.1 We do subtraction the mean of the historical data from the measured data in this time about an example variety. FAT added to the difference is the adjusted table of assessment in this year.

2.4.6.2 The examples are as follows.
Characteristic “Leaf: ratio length/width”
PD of the example variety of the note 5 (EV) is 1.16.

2.4.6.3 The upper line of Figure 3 is FAT expressed in a number line. PD of EV, 1.16 is allocated in the note 4 in FAT. We should adjust FAT as the median of the note 5 becomes the same value to PD of EV, 1.16. FAT subtracted 0.19 gives the table of assessment of this time, the lower line.

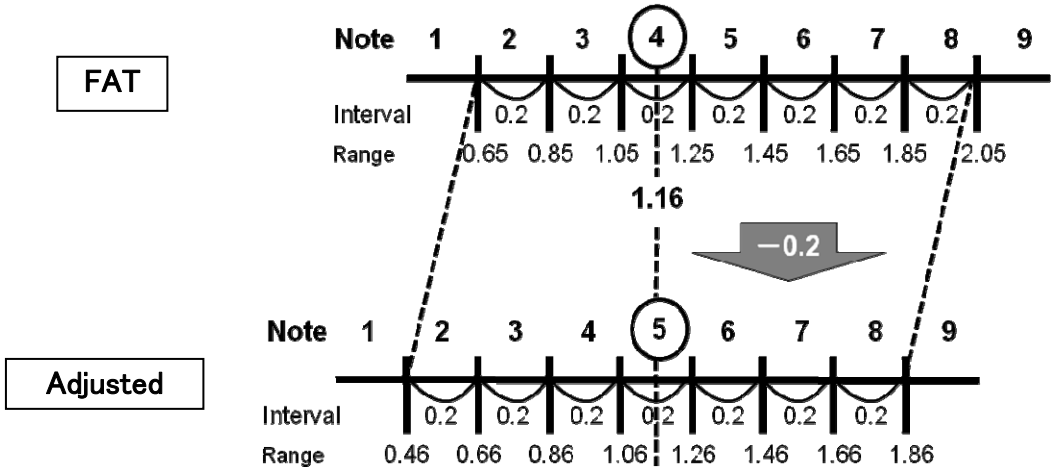


Fig.3: Adjustment FAT with the sliding method

2.4.6.4 We take "the note 5" as an example.
The minimum of the range 1.25−0.19 = 1.06.
The maximum of the range 1.45−0.19 = 1.26.
The interval is not adjusted.
The median of the note 5 = PD of EV, 1.16.

2.4.6.5 Generally, there are several example varieties in a characteristic. But we select one example variety from them for adjustment of FAT. We basically use the least variable example variety during many years' DUS growing trials about each characteristic.

2.5 [Difference between self-pollinated varieties and cross-pollinated varieties]

2.5.1 We use the same method to self-pollinated varieties and cross-pollinated varieties. But the adjustable range changes according to dispersion of HD of example variety. Because our methods are based on the data of example variety, the propagation type of example variety is automatically reflected in the adjustable range.

2.5.2 Table 2 shows the example data. In general, there is tendency that the dispersion of the self-pollinated varieties is lower than that of the cross-pollinated varieties. In this example, HD of two varieties is the same. But the dispersion of self- pollinated varieties example variety is lower than that of cross-pollinated varieties.

Table 2: Example data of self-pollinated example variety and cross-pollinated example variety

Trial number	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Historical Data(HD)	Dispersion	Standard deviation	Coefficient of variance
Self E.V.	80	84	81	83	86	88	83	80	87	88	84.0	9.78	3.13	11.64
Cross E.V.	75	84	74	83	87	96	84	75	88	94	84.0	59.11	7.69	70.37

*E.V.is example variety

2.5.3 Figure 4 shows the normal curve of two varieties of different propagating type. The curve of self-pollinated example variety is narrower than that of cross-pollinated example variety. As I said earlier, if the data of this year is in the range of standard deviation, we can adjust FAT. Therefore, the adjustable range of self-pollinated varieties becomes narrower than that of cross-pollinated ones automatically.

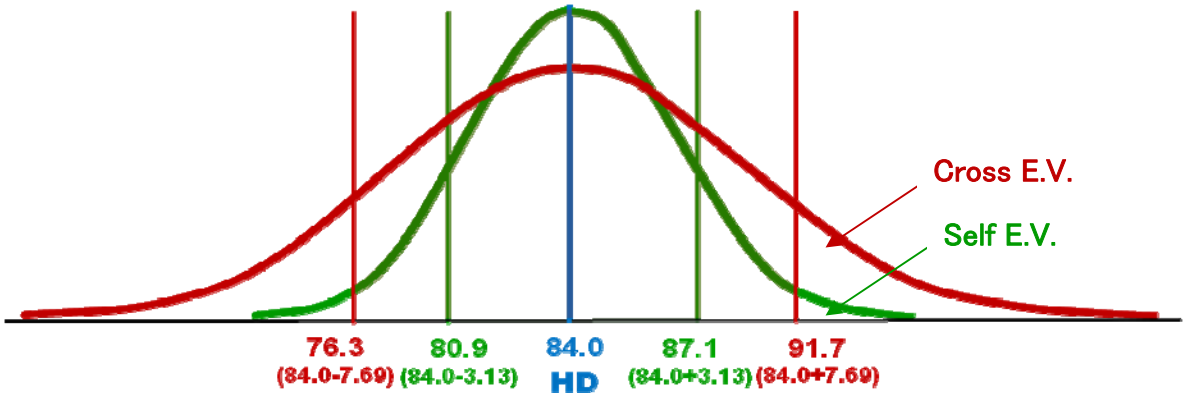


Fig.4: Normal curve of self-pollinated example variety (Self EV) and cross-pollinated example variety (Cross EV)

3. Conclusions

3.1 We have two methods to adjust FAT. One is the proportional method, and the other is the sliding method. In the proportional method, we calculate the proportion of the measured data in this time to the mean of the historical data (HD) about example variety. FAT multiplied by the proportion is the adjusted table of assessment in this time. The sliding method is applied to the characteristics that need fixed interval. We do subtraction the mean of the HD from the measured data in this time about example variety. We can get the adjusted table of assessment in this time by adding the difference to FAT.

3.2 We use the same method to self-pollinated varieties and cross-pollinated varieties to assess the quantitative characteristics. The difference between self-pollinated varieties and cross-pollinated varieties is the allowable range of the value of PD to estimate whether we can adjust the FAT or not. The adjustable range changes according to dispersion of HD of an example variety. Generally, the adjustable range of self-pollinated varieties becomes narrower than that of cross-pollinated varieties because the dispersion of the former is narrower than that of latter. Because our methods are based on the enough experimental data of example variety, the dispersion of HD according to the propagation type of example variety is automatically reflected in the adjustable range.

[Annex II follows]

ANNEX II

CONCLUSIONS OF THE WORKSHOP ON DOCUMENT TGP/14 SECTION 2, SUBSECTION 3 "COLOR"

The following is a report on the conclusions of the workshop on document TGP/14 Section 2, Subsection 3 "Color" (TGP/14 Workshop), which was held on May 30 and 31, 2008, in Lisbon, Portugal, under the chairmanship of Mr. Ton Kwakkenbos (Community Plant Variety Office of the European Community (CPVO)). A copy of the documents and information discussed at the Workshop can be found at *http://www.upov.int/restrict/en/two/index_two41.htm.*

STRATEGIES FOR COLOR CHARACTERISTICS

1. The TGP/14 Workshop agreed that the most appropriate strategy for describing color would need to be considered on a characteristic-by-characteristic basis. The following are strategies which might be appropriate:

(a) Number of Colors

2. The TGP/14 Workshop concluded that the use of characteristics for "number of colors" should be avoided as the starting point for describing color distribution and patterns. Instead, it was agreed that the colors should first be described, followed by characteristics explaining the area, distribution, pattern etc. of each color

Example

1.1	**Petal: color 1**				
(+)					
	Option 1		~~*Option 2**~~		*Option 3*
PQ	green	1	~~UPOV Color Group~~	(1-50)	RHS Colour Chart
	yellow	2			
	red	3			
		etc	* option subsequently deleted		

3. The TGP/14 Workshop agreed that the following standard explanation should be included in the Test Guidelines when using this approach for describing color:

> "Ad. 1 (Option 1). The order of colors in Char. 1.1, 1.2 etc. should be according to the order in the states of expression (green, yellow etc.). The RHS Colour Chart should be used to allocate the color to the appropriate state. In some cases, it may be appropriate to create particular groupings of RHS Colour Chart refernces in the Test Guidelines."
>
> ~~"Ad. 1 (Option 2). The order of colors should follow the UPOV Color Group order"~~
>
> "Ad. 1 (Option 3). The order of colors should follow the RHS Colour Chart order."

- with the following paragraph added for all options above:

> "A photograph of the [relevant organ] should be provided in conjunction with the description in order to clarify the color pattern. However, a warning should be added to this photograph, explaining that the first intention is to represent the distribution of colors on flowers of the varieties more than the colors themselves. Such colors can be affected by the technology of the camera and the facilities used to display the photograph (printer, overhead projector, etc.)."

1.2	**Petal: color 2**				
(+)					
	Example 1		~~*Example 2*~~		*Example 3*
PQ	green	1	~~UPOV Color Group~~	(1-50)	RHS Colour Chart
	yellow	2			
	red etc.	3			

etc.

2.1	**Petal: area of color 1** (small/medium/large)
2.2	**Petal: area of color 2** (small/medium/large)
3.1	**Petal: distribution of color 1** (at margins etc..
3.2	**Petal: distribution of color 2** (at margins, at base etc..
4.1	**Petal: shape of color 1** (continuous base color (1), spots (2); stripes (3)
4.2	**Petal: shape of color 2** (continuous base color (1), spots (2); stripes (3)
5.1	**Petal: border of color 1** (clearly defined to slightly diffused (1); moderately diffused (2); strongly diffused or continuous (3))
5.2	**Petal: border of color 2**(clearly defined to slightly diffused (1); moderately diffused (2); strongly diffused or continuous (3))

(b) "Ground" / "Over" color

4. The TGP/14 Workshop agreed that for organs which have two independent layers of tissue containing color pigmentation (e.g. apple), the two layers could be described as follows:

GROUND COLOR: the color of the inner tissue layer, which in most cases develops first.

OVER COLOR: the color of the outer tissue layer, where this pigmentation is developed. In most cases this color appears after the ground color.

(c) "Main" Color

5. The TGP/14 Workshop concluded that the term "MAIN COLOR" should only be used where, for all varieties, there would always be a clearly identifiable main color, with a continuous distribution across the relevant organ, with other colors in the form of isolated spots, patches etc. e.g.

1. (+)	**Organ: main color**			
	Option 1		*Option 2*	
PQ	green	1	RHS Colour Chart	
	yellow	2		
	red	3		
		etc		

6. The TGP/14 Workshop agreed that the following standard explanation should be included in the Test Guidelines when using this approach for describing color:

"Ad. 1 The main color is the color which has a continuous dispersion across the surface of the organ; in general, it will also be the largest surface area."

- with the addition of the following paragraph if considered appropriate for the characteristic:

"A photograph of the [relevant organ] should be provided in conjunction with the description in order to clarify the color pattern. However, a warning should be added to this photograph, explaining that the first intention is to represent the distribution of colors on flowers of the varieties more than the colors themselves. Such colors can be affected by the technology of the camera and the facilities used to display the photograph (printer, overhead projector, etc.)."

Example 1

2. **(+)**	**Organ: color of spots**	
PQ	green yellow etc.	

Example 2

2. **(+)**	**Organ: form of secondary color**	
PQ	spots spots and patches patches etc.	

7. The TGP/14 Workshop agreed that the scheme for determining color pattern terms, drafted by the experts from Japan, attached as the Appendix to this document, should be included in TGP/14 with any necessary modifications.

(d) Color Change Over Time

8. The Workshop noted the need to consider how to describe different color transition stages. The proposal below was discussed, but it was agreed that further discussion would be required in relation to that proposal, in particular by the Technical Working Party for Vegetables (TWV) in relation to the Test Guidelines for Pepper.

1.	**Fruit: number of different colors over time:**	
QN	one	1
	two	2
	three	3
	four	4

2.	**Fruit: succession of colors (only for varieties with more than two colors)**	
PQ	green-yellow-red	1
	green-yellow-orange-red	2
	white-yellow-red	3
	white-yellow-orange-red	4
	yellow-orange-red	5

(e) Describing color patterns where those are in addition to the variegation in variegated varieties

9. The TGP/14 Workshop agreed on the following definition:

VARIEGATION: well defined areas of different colors, with less or no chlorophyll, especially as irregular patches or stripes on one organ.

10. The TGP/14 Workshop agreed that the following approaches might be used as appropriate on a case-by-case basis:

In cases where there are more colors than a main green color and a less green variegated part:

(a) exclude variegation from the general color pattern by defining variegation and indicating "(excluding variegation)" in the general pattern characteristics (where appropriate); or

(b) consider variegation within the general color pattern and indicate "(including variegation)"

(f) Consideration of whether pigments, such as anthocyanin, should be considered as a color

11. The TGP/14 Workshop proposed:

(a) to refer to "anthocyanin coloration" where the pigment is known to be anthocyanin;

(b) to refer to "red pigment" in cases where the red pigment is not known or is not anthocyanin; or

(c) to refer to the name of the pigment if known.

12. With regard to describing anthocyanin/red coloration, the TGP/14 Workshop agreed, on a case-by-case basis, to decide whether coloration should be:

(a) considered as a color pattern; or

(b) excluded from the pattern observations, by indicating, e.g. "(excluding anthocyanin)".

13. The TGP/14 Workshop agreed that TGP/14 should provide guidance on anthocyanin coloration characteristics, including: an explanation that, unlike other color characteristics, anthocyanin coloration is, in general, a quantitative characteristic; the possibility to describe intensity of anthocyanin coloration (weak, medium strong) and / or distribution of anthocyanin coloration; the importance of light intensity, position on plant, temperature etc. in observing anthocyanin coloration characteristics.

(f) "Conspicuousness"

14. The TGP/14 Workshop proposed that a characteristic for conspicuousness might be used where the individual factors could not be usefully described, e.g. for small organs (e.g. veins, hairs), or because they are not consistently expressed across the organ.

15. The TGP/14 Workshop agreed that an explanation of the meaning of "conspicuousness" in terms of the individual factors (e.g. color contrast, relative size etc.) should be provided.

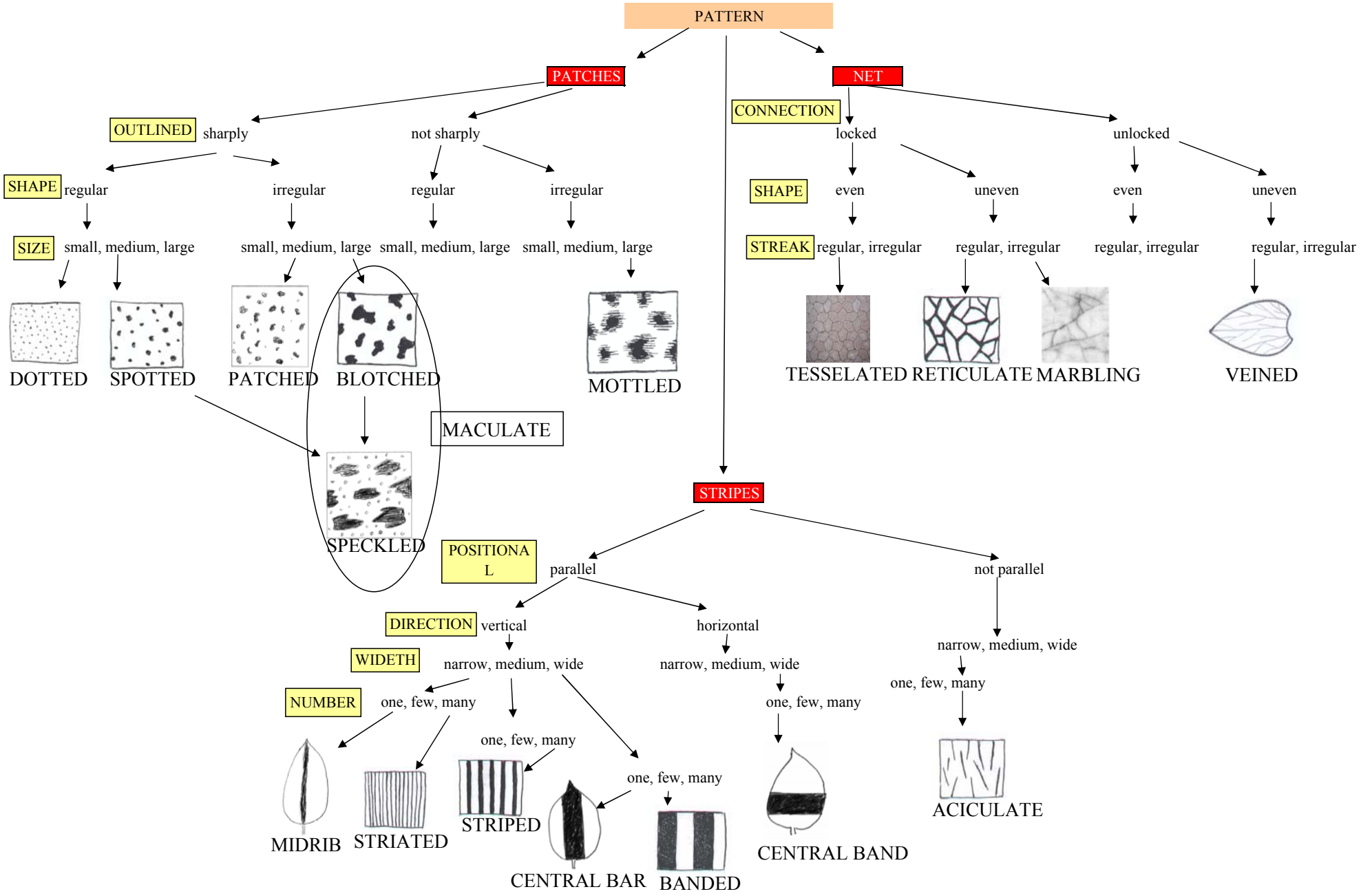
(g) COLOR CHART

16. The TGP/14 Workshop noted that a new version of the RHS Colour Chart had been published and that it was understood that some color charts had been added in the new (2007) version. It agreed that document TGP/14, including Annexes I and II should be updated accordingly.

[Appendix follows]

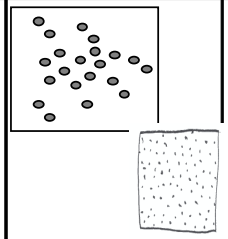
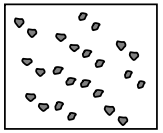
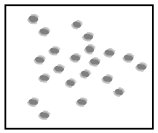
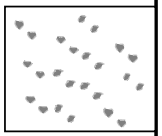
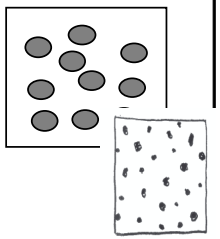
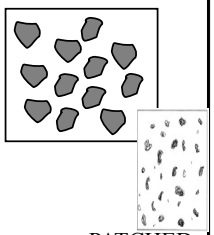
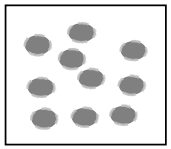
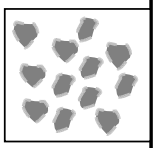
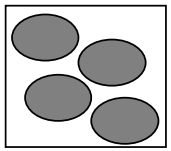
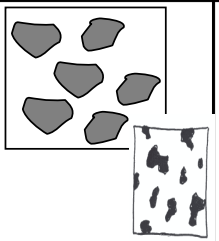
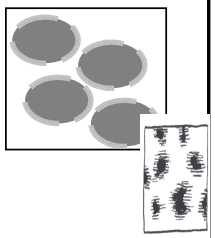
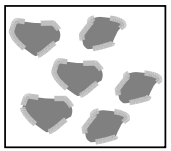
APPENDIX

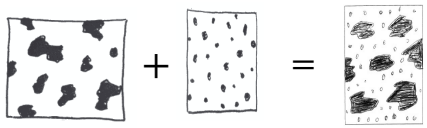
SCHEME FOR DETERMINING COLOR PATTERN TERMS



Annex 2

PATCHES	
outline	(sharply / not sharply)
shape	(regular / irregular)
size	(small / medium / large)

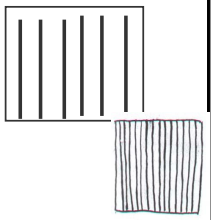
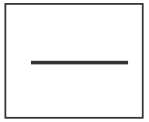
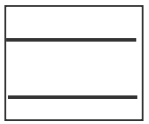
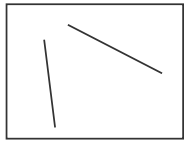
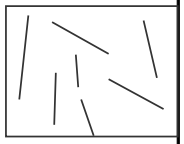
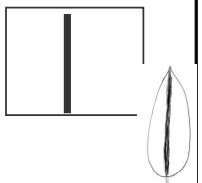
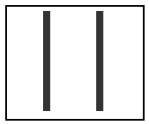
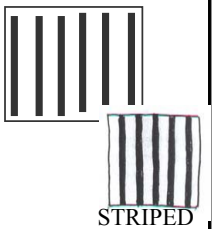
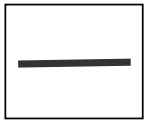
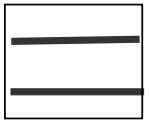
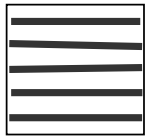
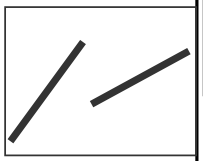
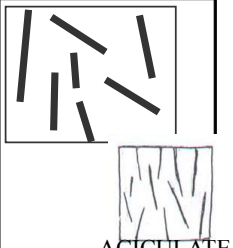
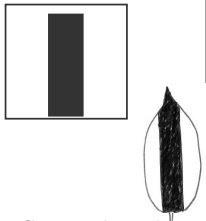
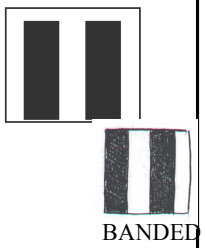
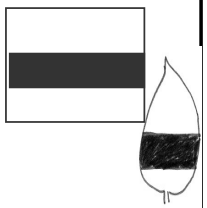
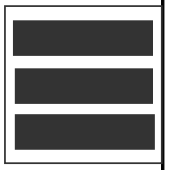
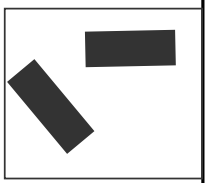
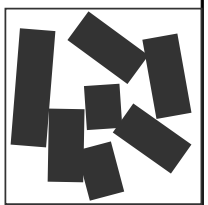
OUTLINED	sharply		not sharply	
SIZE＼SHAPE	regular	irregular	regular	irregular
small	 DOTTED			
medium	 SPOTTED	 PATCHED		
large		 BLOTCHED	 MOTTLED	


SPECKLED

Annex 3

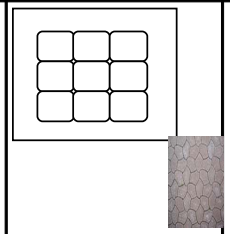
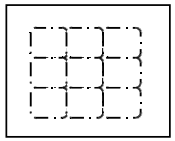
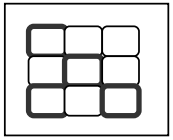
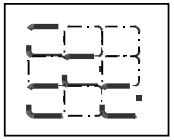
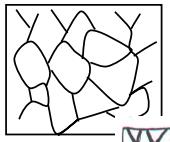
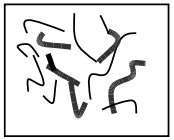
STRIPES

positional relation (parallel / not parallel)
direction (vertical / horizontal)
width (narrow / medium / wide)
number (one / few / many)

position	parallel						not parallel	
direction	vertical			horizontal				
width \ number	one	few	many	one	few	many	few	many
narrow			 STRIATED					
medium	 MIDRIB		 STRIPED					 ACICULATE
wide	 CENTRAL BAR	 BANDED		 CENTRAL BAND				

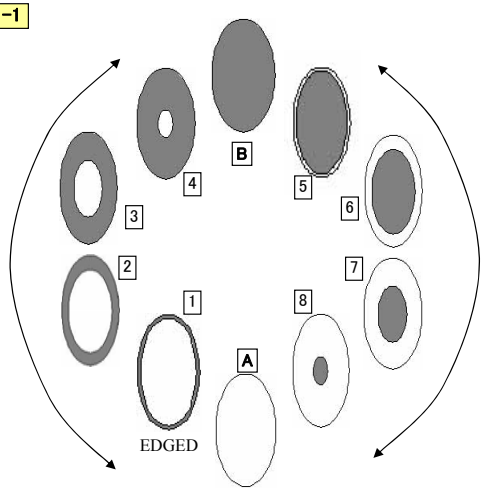
Annex 4

NET	
connection	(locked / unlocked)
shape/grid	(even/ uneven) / (regular / irregular)

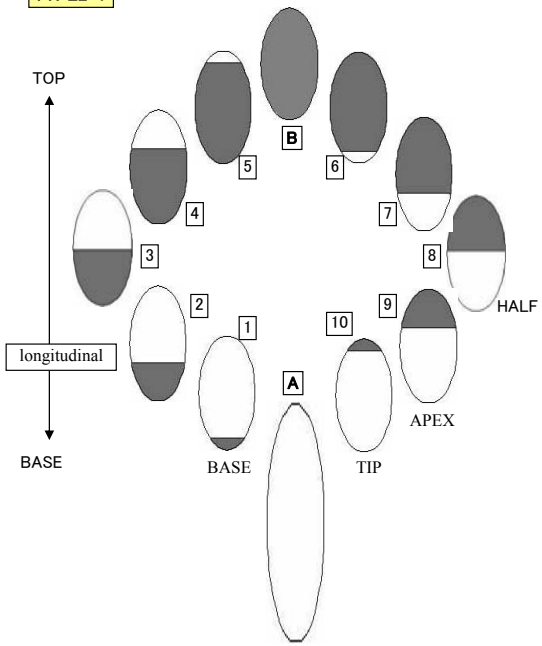
shape	grid	locked	unlocked
even	regular	 TESSELATED	
	irregular		
uneven	regular	  RETICULATE	  VEINED
	irregular	  MARBLING	

DISTRIBUTION

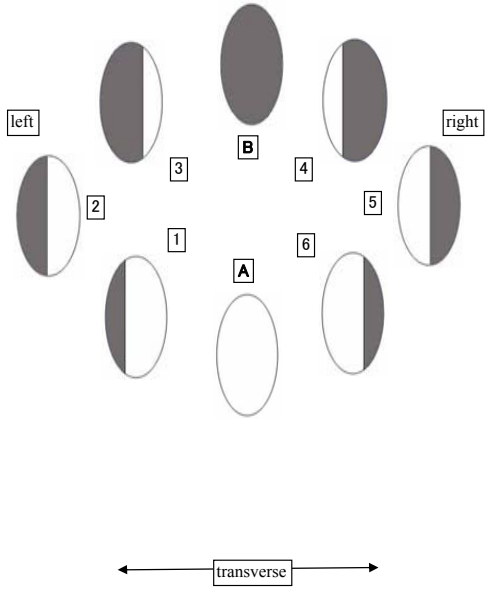
TYPE 1-1



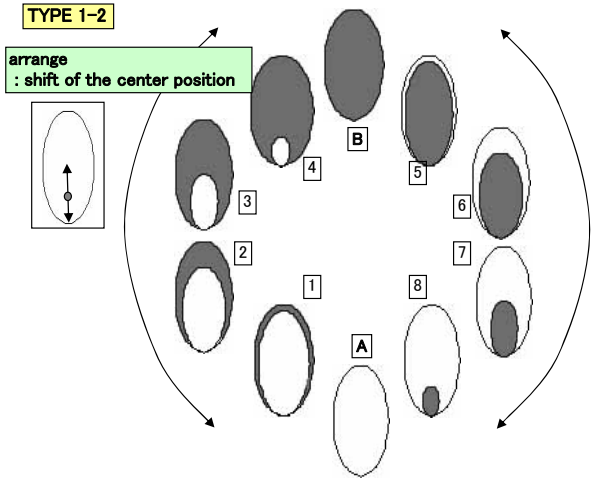
TYPE2-1



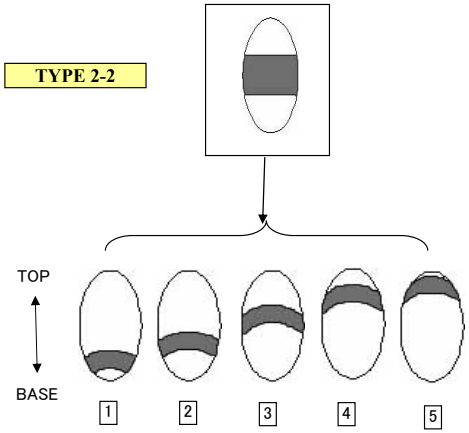
TYPE 3-1



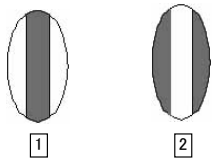
TYPE 1-2



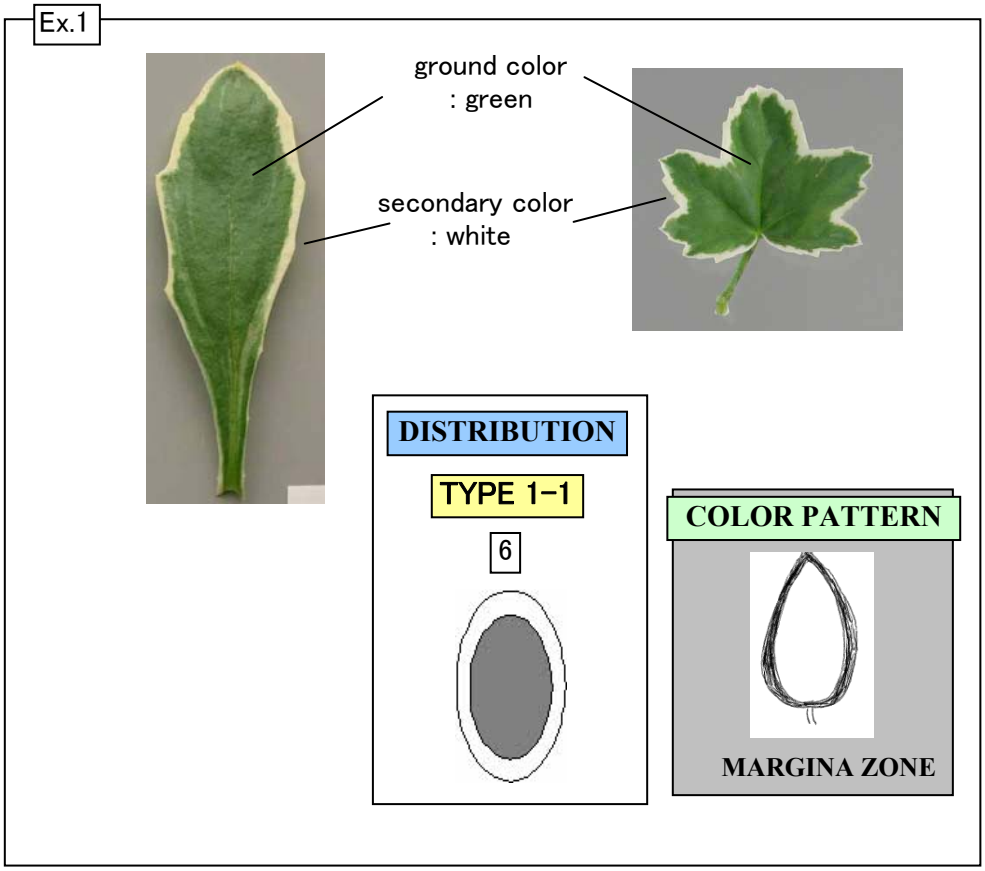
TYPE 2-2

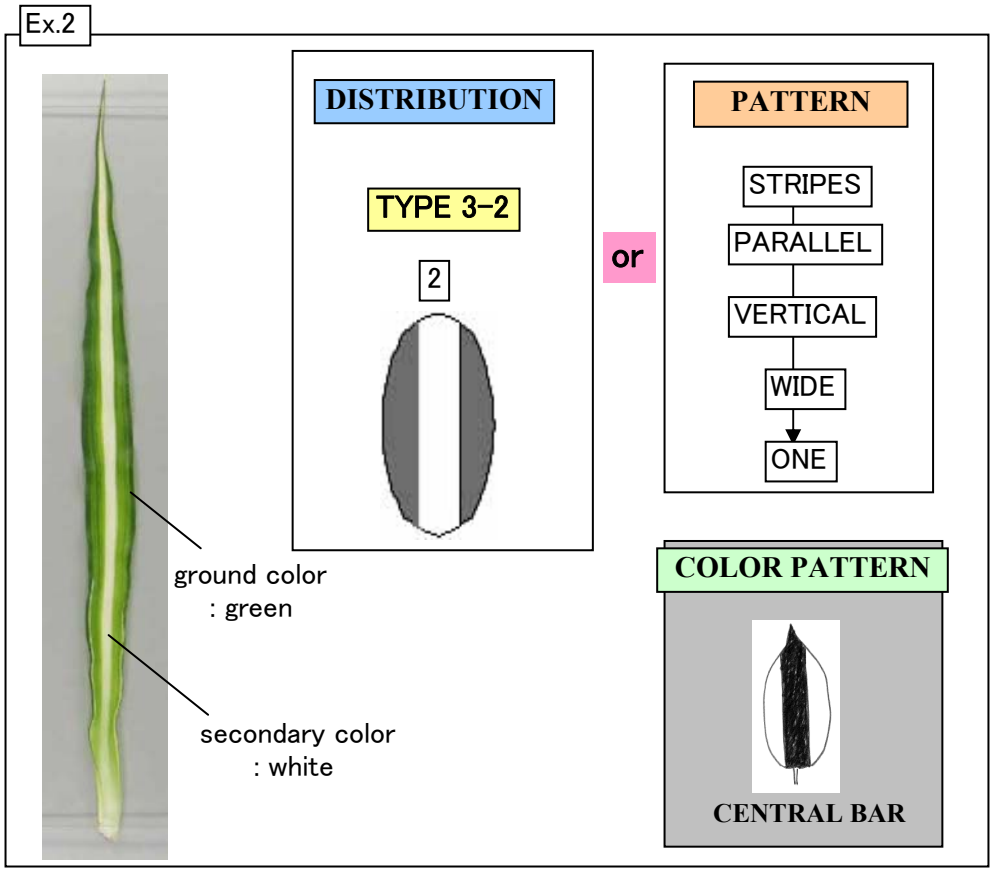


TYPE 3-2



EXAMPLE(LEAF)

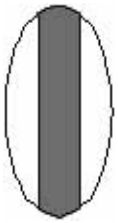




Ex.3

DISTRIBUTION

TYPE 3-2



1

ground color
: green

secondary color
: white

or

PATTERN

STRIPES

PARALLEL

VERTICAL

MIDIUM

FEW

Ex.4

ground color :
green

secondary color : white

PATTERN

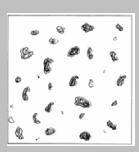
PATCHES

SHARPLY

IRREGULAR

MEDIUM

COLOR PATTERN



PATCHED

Ex.5

ground color
: green

secondary color
: yellow

tertiary color : red

DISTRIBUTION

TYPE 1–2

3

plus

PATTERN

STRIPES

PARALLEL

VERTICAL

NARROW

ONE

MID RIB : red

COLOR PATTERN

MIDRIB

Ex.1 Stripe

take bloom to petal →

secondary color

main color

COLOR PATTERN

PICOTEE

DISTRIBUTION

TYPE 1-2

1

and

PATTERN

- STRIPES
- PARALLEL
- VERTICAL
- NARROW
- MANY

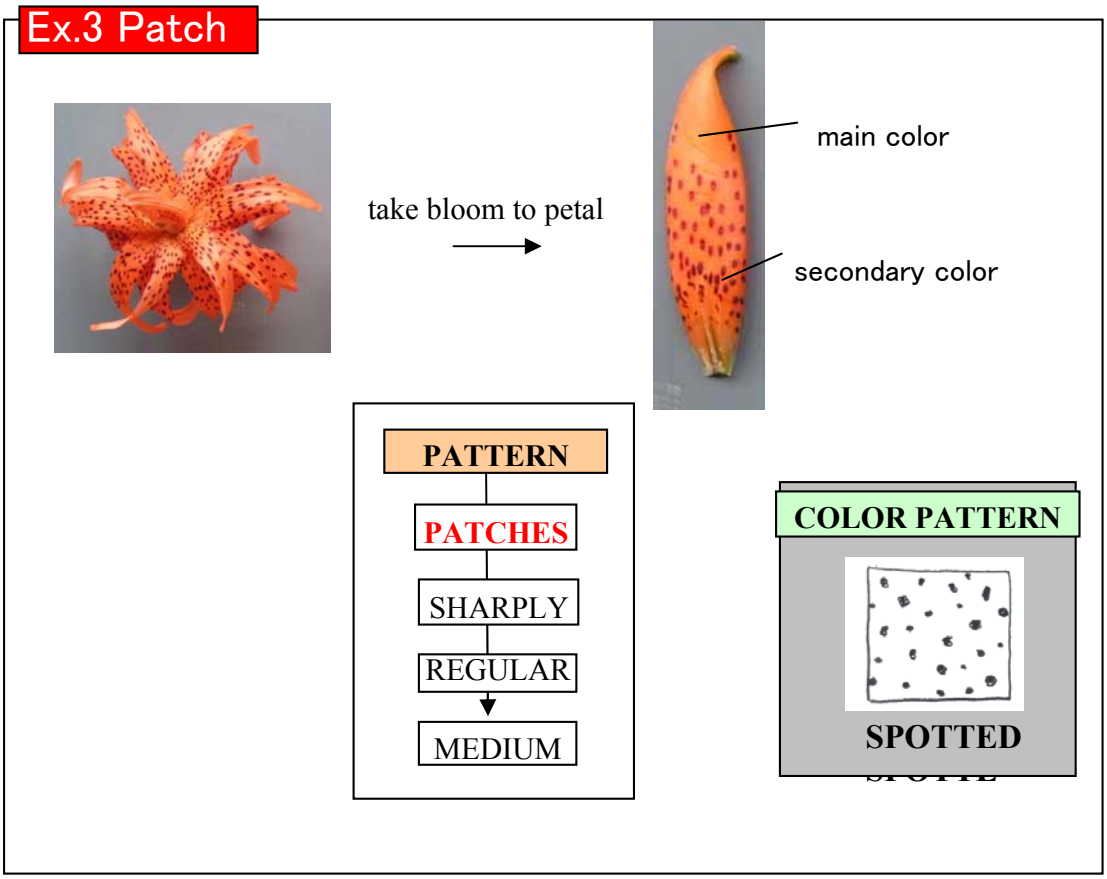
Ex.2 Net

PATTERN

- NET
  - LOCKED
    - UNEVEN
      - REGULAR
  - UNLOCKED
    - UNEVEN
      - REGULAR

COLOR PATTERN

VEINED



[Annex III follows]

ANNEX III

COMMENTS OF THE TECHNICAL WORKING PARTIES ON DOCUMENT TGP/14/1 DRAFT 6 SECTION 2, SUBSECTION 3 "COLOR" AND ON THE CONCLUSIONS OF THE WORKSHOP ON DOCUMENT TGP/14 SECTION 2, SUBSECTION 3 "COLOR"

TWA	The Technical Working Party for Agricultural Crops (TWA) proposed that document TGP/14/1 Section 2, Subsection 3 "Color" should include guidance on characteristics and states of expression for green color and, in particular, should avoid the creation of a separate characteristic for intensity of particular hues of green (c.f. draft Test Guidelines for Pea (document TG/7/10(proj.5): Chars. 7 and 8)

TWF	The Technical Working Party for Fruit Crops (TWF) supported the proposals set out in document TWF/39/3 Add. (Annex II to this document). With regard to characteristics for color changes over time, it noted that that matter would be discussed at its next session in relation to Peach. It was also noted that any such characteristics would need to fulfill the UPOV requirements for a characteristic. The TWF proposed that the example of anthocyanin coloration in the flesh of peach could be used to illustrate the need to consider both the intensity and distribution of anthocyanin coloration in some cases.

TWO	The Technical Working Party for Ornamental Plants and Forest Trees (TWO) agreed that the definitions of the components of color should be deleted from document TGP/14/1 Draft 6 Section 2, Subsection 3 "Color", II.1, 2.1.
	The TWO agreed to start using the proposals set out in document TWF/39/3 Add. (Annex II to this document) in the preparation of draft Test Guidelines for 2009. It noted that it would be necessary to develop a new state of expression in color pattern characteristics to describe the area of color which was previously described as the "main" color, (e.g. continuous dispersion). It was also agreed that the example in 4.2 of document TWF/39/3 Add. (Annex II to this document) should be amended to read "Petal: shape of color [1]/[2] area. The TWO agreed that it would still be important to retain the possibility to have a characteristic for number of colors in order to have a simple overall characteristic, but which was not used as the starting point to describe color pattern. It also agreed that it would be important to retain the option, where appropriate, to describe the color pattern by describing colors in specified parts of the plants (e.g. color of margin, color of basal zone etc.). With regard to anthocyanin coloration, it was agreed that an example of characteristics should be included in TGP/14. In order to develop and test the approach to color characteristics proposed

	in document TWO/41/3 Add. (Annex II to this document), the TWO agreed to have an exercise on color in Alstroemeria, Canna and Phalaenopsis to see if characteristics based on that approach would be more effective than the traditional approach. The TWO agreed that the European Community should coordinate a subgroup to develop proposals for an exercise to be conducted by the TWO, in which the two approaches would be evaluated. The experts present at the session, from Australia, France, Germany, Japan, Mexico, Netherlands (Kees Grashoff), New Zealand, United Kingdom and the Office of the Union agreed to participate in the subgroup. The first draft of characteristics, to be prepared by the European Community according to the proposed new approach, would be circulated to the subgroup for comment by October 31, 2008, with 4 weeks for comments. On the basis of the comments, a new draft would be prepared by the European Community and checked by the subgroup. A circular presenting the exercise would be sent by the Office of the Union to the TWO by the end of February 2008, with 6 weeks for completion[1]. The completed exercises by the TWO experts would be sent to the European Community, with a copy to the Office of the Union. The European Community would then prepare a TWO document, containing the compiled results of the exercise, 6 weeks before the forty-second session of the TWO.

`

TWV	The Technical Working Party for Vegetables (TWV) agreed that consideration should be given to including "flecking" as a color pattern in the scheme in the annex to document TWV/42/3 Add. (Annex II to this document). With regard to document TWV/42/3 Add. (Annex II to this document), "(d) Color Change Over Time", the TWV agreed that characteristic 2 "Fruit: succession of colors" should be considered as a possible option for consideration in relation to relevant Test Guidelines.

[i] TWC proposed to delete Section 3.1.5
[b] TWC: to be edited
[c] TWC: to be edited
[d] Rewording proposed by TWC
[e] Rewording proposed b TWC
[f] TWC: Mr. Gerie van der Heijden (Netherlands) will consult his Naktuinbouw colleagues in the Netherlands to see if they could contribute a draft for this section.
[g] TWC proposed to delete Section 3.4
[h] Section drafted by Mr. Gerie van der Heijden (Netherlands)
[i] TWA and TWC agreed to move Section III "Examination of characteristics using image analysis" from TGP/12 to TGP/8

[End of Annex III and of document]

[1] The exercise on color in Alstroemeria, Canna and Phalaenopsis can be found on the TWO/42 website (*http://www.upov.int/restrict/en/two/index_two42.htm*)