The systematic description of cacao clones and its significance for taxonomy and plant breeding

Promotoren : dr. ir. J.E. Parlevliet, hoogleraar
in de plantenveredeling

dr. ir. L.J.G. van der Maesen, hoogleraar
in de plantentaxonomie

J.M.M. ENGELS

THE SYSTEMATIC DESCRIPTION OF CACAO CLONES AND ITS SIGNIFICANCE FOR TAXONOMY AND PLANT BREEDING

Proefschrift
ter verkrijging van de graad van
doctor in de landbouwwetenschappen,
op gezag van de rector magnificus,
dr. C.C. Oosterlee,
in het openbaar te verdedigen
op vrijdag 12 september 1986
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen.

ISN 2[illegible]

STELLINGEN

1. In landen zoals Ethiopië wordt niet alleen de genetische variatie in cultuurplanten bedreigd, maar ook de culturele variatie. Voor beide geldt dat aan het behoud van deze variatie grote aandacht dient te worden geschonken.

2. Het conserveren van genenmateriaal van cultuurgewassen in zogenaamde natuurparken - in situ conservering - is uit theoretisch oogpunt bezien ideaal, echter uit praktische, economische en administratieve overwegingen nauwelijks realiseerbaar.

3. Het met desastreuse gevolgen optreden van "coffee berry disease" (CBD) en koffie roest in de Ethiopische koffie is vooral te wijten aan de voortschrijdende verstoring van het ecologische evenwicht in de koffiebossen. Deze verstoring bestaat met name uit de invoering van cultuurmethoden (plantage koffie) en door het planten van nieuwe, genetisch uniforme rassen in de staatsboerderijen leidend tot ernstige genetische erosie.

4. De term "cocoa" zou om taalkundige redenen in de engelstalige literatuur vermeden moeten worden, ofwel alleen maar consequent voor het eindproduct gebruikt moeten worden.

   . Chatt, E.M., 1953. Cocoa, cultivation, processing, analysis. Interscience Publishers, New York. 302 pp.

5. Wanneer botanici cultuurplanten proberen te classificeren, kan men toevallige resultaten verwachten.

   . Harlan, J.R., 1975. Crops and men. American Society of Agronomy, Madison, Wisconsin. 295 pp.

6. Een te lang verblijf van een zogenaamde "expert" in eenzelfde ontwikkelingsland bevordert in het algemeen niet zijn motivatie voor het werk, maar wel het aangepast zijn aan de omgeving, waarbij dit laatste zijn honkvastheid neigt te bevorderen.

7. Het systematisch beschrijven van collecties in genenbanken vormt een belangrijke voorwaarde voor een efficiënte benutting van deze geconserveerde genetische verscheidenheid in veredelingsprogramma's.

   . Dit proefschrift

8. Het feit dat cacao mogelijk een tetraploid en geen diploid gewas is, waarvoor in dit onderzoek sterke aanwijzingen zijn gevonden, zou ten dele de geringe vooruitgang die de plantenveredeling in dit gewas heeft geboekt, kunnen verklaren.

. Dit proefschrift

9. Het verlenen van medische hulp aan ontwikkelingslanden leidt vaak tot een verstoring van het subtiele evenwicht tussen mens en zijn omgeving. Derhalve dienen tegelijkertijd alternatieven te worder aangeboden voor de socio-economische zekerheden, die tot heden toe meestal bestonden uit grote nakomelingenschappen.

10. Het verlenen van voedselhulp aan door hongersnood geplaagde landen is in het algemeen niet meer dan symptoombestrijding.

11. Het uitvoeren van een promotieonderzoek en vervolgens het schrijven van een proefschrift in een wetenschappelijk isolement (bijvoorbeeld in een ontwikkelingsland) kan er gemakkelijker toe leiden, dat niet zo zeer de kracht van de omgeving (= fenotype) tot uiting komt, maar dat vooral de zwakte van de promovendus (=genotype) zichtbaar wordt.

J.M.M. Engels

The systematic description of cacao clones and its significance for taxonomy and plant breeding.
Wageningen, 12 september 1986.

ABSTRACT

ENGELS, J.M.M., 1986. The systematic description of cacao clones and its significance for taxonomy and plant breeding (x) + 125 p., 11 figures, 38 tables, 149 references. Doctoral thesis.
Agricultural University, Wageningen.

The value of germplasm collections depends to a large extent on the data accompanying the individual accessions. In order to facilitate the selection of the most useful characters for the systematic description of a cacao germplasm collection methods were developed to measure and to compare the discriminative values of both qualitative and quantitative characters. The relationships between cacao clones were studied as it was expected that they would influence the value of the discriminatory power of a given character for a given group of clones and some conclusions with regard to breeding of cacao were developed from the study of the relationships between characters.

The mode of inheritance of some qualitative and quantitative characters was studied using the data of a complete diallel cross, in order to determine the relationship between the discriminative value of a character and its inheritance. Although such a relationship could not be established there were strong indications that the qualitative characters examined follow tetraploid rather than diploid inheritance. Following the results and based on progeny testing and clonal propagation a breeding method was proposed for a more efficient utilization of cacao germplasm.

Classical taxonomic classifications were proven to be useful. However, to enable correct identification of cacao clones in a germplasm collection it was felt advisable to propose a new classification.

Free descriptors: cacao clones, discriminative value, quantitative and qualitative descriptors, taxonomic identifiers, numerical taxonomy, multivariate statistical methods, diallel cross, inheritance, general and specific combining ability, reciprocal effects, additive and dominance variation, cacao breeding, classification and identification.

PREFACE

The topic of this thesis originated in 1978 as a problem of routine documentation work at the Centro Agronomico Tropical de Investigacion y Ensenanza (CATIE) in Turrialba, Costa Rica. I am grateful to Dr Gilberto Paez formerly Director of CATIE and Dr Gustavo Enriquez, Head of the Perennial Crops Programme and cacao breeder respectively, for their support and encouragement during the collection of the data. The data of the diallel cross were collected in part by the staff of the cacao breeding programme. The kind acquiesce of Dr Jorge Soria, the initiator of the diallel cross project for the use of the data in this thesis is gratefully acknowledged.

Prof. Dr J. Sneep accepted the theme as a possible thesis topic and for his "long distance" supervision as well as for his suggestions, criticism and encouragements I am deeply grateful. Because of his retirement Prof. Dr J.E. Parlevliet and Prof. Dr L.J.G. van der Maesen were readily prepared to take over the task of supervision and I am very thankful for their advice, criticism and great interest in spite of the fact that the thesis was already partly completed. Also the suggestions, comments and creative discussions with Dr Anton Zeven and Dr Ies Bos were of great help.

The actual assistance in data collection, which was sometimes done under adverse conditions, was provided by Jorge Morera, Rigoberto Bonilla and Rodolfo Sanchez. Without the great and persistent help of my colleagues in the computing unit at CATIE (Mrs Heather Palmer and Dr Julio Henao) and later on in Wageningen (Ing. G. Heemstra) as well as in Addis Abeba (Jeff Durkin and Robin Sayers of ILCA and Solomon Zewdie of PGRC/E), and without Mrs Yadegdigu Belay's patient (re)typing of the numerous drafts the analyses and the preparation of the manuscripts could not have been finalized.

I would like to thank my colleagues and friends: the late Dr Bill Dyson, Mr John Palmer, Dr Mike Jackson and, especially, Dr John Lazier for their considerable effort in helping to correct the English text and for their many suggestions and comments given during the course of the preparation of the articles. The (indirect) support of the Deutsche Gesellschaft fuer Technische Zusammenarbeit (GTZ) GmbH made it possible for me to live in Costa Rica to collect the necessary data. I would like to thank the concerned colleagues in the Headquarters at Eschborn.

Last but not least, I am especially grateful to my wife, Eimelt, and my children Boris, Saskia and Claudia, who had to realise how difficult it is to "understand" cacao. However, their understanding and patience were real encouragements.

CONTENTS

1. INTRODUCTION

The genus Theobroma, to which the species T. cacao belongs, originated in South America, probably in the Amazon basin; a second centre of speciation is found on the Pacific slopes of the Andes mountains in Columbia (Purseglove, 1968; Wood, 1975). Of the 22 Theobroma species listed only T. cacao L. and T. bicolor Humb. and Bonpl. have been domesticated and/or used. Brucher (1977) listed six other species which are used to some extent, mainly for the appreciated pulp around the seeds. The domestication of T. cacao took place in Central America in pre-Columbian times (Thorold, 1975) and the words 'cacao' and 'chocolate' are of Central American origin (Nahuatl, Aztec). Cultivation was extended southwards by the Spaniards. The main centre of diversity is assumed to be in South America, somewhere in the Columbian-Venezuelan marshes (Thorold, 1975). The so-called 'Criollo' types were cultivated in Central America and the 'Forasteros' in South America. The hybrids between these two "genetic groups" are called 'Trinitarios'. The Forastero cacaos are at present the most important group commercially and are responsible for some 80% of world cocoa production (Wood, 1975).

For many years the genetic diversity of cacao has been threatened by factors such as genetic erosion due to deforestation on a vast scale, replacement of primitive cultivars by modern ones and heavy disease pressure. There is, however, a great demand for more genetic diversity in breeding programmes, since: a) the genetic base of parents used for the production of hybrid cacao cultivars is very limited, and b) the progress in the control of major diseases is slow mainly due to lack of genetically diverse germplasm (Anonymous, 1981a). Thus, the systematic collection of cacao germplasm, its conservation in collections and a better evaluation of the existing collections have been recommended by international organizations and plant breeders.

The need for proper documentation and information for any collected germplasm sample has been generally stressed (Anonymous, 1983; Swaminathan, 1983; Sneep et al., 1979; Engels, 1985), for the quality of the documentation work determines directly the usefulness of the germplasm collections for plant breeders (Anonymous, 1981b), and for proper management by genebanks (Chang, 1976; Engels, 1983).

Since the exchange of information generally preceeds the exchange of germplasm it has been recommended by the International Board for Plant Genetic Resources (IBPGR) and others that the descriptors* and their states be standardized in order to facilitate communication between scientists (Rogers et al., 1975; Howes, 1981). For many crops descriptor lists have been developed by panels of experts and published by IBPGR (Anonymous, 1984). The list for cacao (Anonymous, 1981a) is principally based on the descriptors proposed in Engels et al. (1980).

One of the aspects of standardization of the descriptors was the need for a clear definition of the various categories in genetic resources documentation (Howes, 1981; Simmonds, 1981). IBPGR classifies genebank descriptors into:

1) **passport data** (accession identifiers and information recorded by collectors);

2) **characterization** (those characters are utilized which have a reasonable constant expression over environments, which can be easily observed, and which are expressed in all environments);

3) **preliminary evaluation** (the recording of a limited number of additional traits which show continuous variation and which are recommended by the users of the particular crop);

4) **further evaluation** (consists of recording economic characters which need specific conditions during the evaluation).

The descriptors of passport data and characterization form a "minimal descriptor list" (Simmonds, 1981). In the present study characters which belong to the systematic description (= characterization plus preliminary evaluation) were used. Since the descriptors recorded during the characterization of a crop will be of primary importance for the identification or characterization of a germplasm accession, it is clear that

---

* The terms descriptor, character and characteristic are used interchangably.

these descriptors have to reflect highly heritable (qualitative) characters. These descriptors involve many morphological characters which generally do not possess great agronomic importance. Thus quantitative characters (i.e. less highly heritable) were examined for their usefulness in characterizing cacao clones and for their use in breeding programmes.

The reliability of scoring and the repeatability of the observations are crucial. These aspects have been studied and the results are presented for the quantitative characters in Chapter 2. The discriminative value of qualitative characters (Chapter 3) and quantitative characters (Chapter 4) have been calculated to enable a selection of the most discriminative characters. The relationships between the clones involved was given special attention as well as the relationships between characters (Chapter 5) and some conclusions regarding the breeding work in cacao were drawn. The heritability of qualitative and quantitative characters were studied (Chapters 6 and 7) using a complete diallel of seven clones. The results of these studies were then used to classify and identify cacao clones; and the implications for germplasm management were discussed (Chapter 8). A general discussion concludes the study (Chapter 9).

REFERENCES

ANONYMOUS, 1981a. Genetic resources of cacao. International Board for Plant Genetic Resources, Rome. 25pp.

ANONYMOUS, 1981b. Crop genetic resources. International Board for Plant Genetic Resources, Rome. 12pp.

ANONYMOUS, 1983. Practical constraints affecting the collection and exchange of wild species and primitive cultivars. International Board for Plant Genetic Resources, Rome. 11pp.

ANONYMOUS, 1985. Annual report 1984. International Board for Plant Genetic Resources, Rome. 122pp.

BRUCHER, H., 1977. Tropische Nutzpflanzen. Ursprung, Evolution und Domestikation. Springer-Verlag, Berlin. 529pp.

CHANG, T.T., 1976. Manual on genetic conservation of rice germplasm for evaluation and utilization. The International Rice Research Institute, Los Banos, Philippines. 77pp.

ENGELS, J.M.M., B.G.D. BARTLEY & G.A. ENRIQUEZ, 1980. Cacao descriptors, their states and modus operandi. Turrialba 30: 209-218.

ENGELS, J.M.M., 1983. Documentation and utilization at the Plant Genetic Resources Centre/Ethiopia. In: Neddenriep, K.J and D. Wood, (eds). Genetic resources and the plant breeder: the next ten years. The German Agency for Technical Cooperation (GTZ) Ltd. (mimeographed). 140pp.

ENGELS, J.M.M., 1985. Descripcion sistematica de colecciones de germoplasma. Serie: Lectura sobre recursos fitogeneticos. No. 6. Caracterizacion y documentacion. Consejo Internacional de Recursos Fitogeneticos (CIRF). Regional Coordinator, CIAT, Colombia. 21pp.

HOWES, C., 1981. Guidelines for developing descriptor lists. Plant Genetic Resources-Newsletter 45: 26-32.

PURSEGLOVE, J.W., 1968. Tropical crops: dicotyledons. John Wiley & Sons, New York. 719pp.

ROGERS, D.J., B. SNOAD and L. SEIDEWITZ, 1975. Documentation for genetic resources centres. In: O.H. Frankel and J.G. Hawkes (eds). Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge. p.399-405.

SIMMONDS, N.W., 1981. Principles of characterization and evaluation. In: Proceedings of the FAO/UNEP/IBPGR Technical Conference on Crop Genetic Resources in Rome, April 1981. FAO-UNEP-IBPGR, Rome. 80pp.

SNEEP, J. and A.J.T. HENDRIKSEN (Eds.), O. HOLBEK (Coed.), 1979. Plant breeding perspectives. Pudoc, Wageningen. 435pp.

SWAMINATHAN, M.S., 1983. Genetic conservation: microbes to Man. Presidential Address. XV Int. Congress of Genetics. New Delhi. 32pp.

THOROLD, C.A., 1975. Diseases of cocoa. Claredon Press, Oxford. 423pp.

WOOD, G.A.R., 1975. Cocoa, (Third edition). Longman, London. 292pp.

Reprinted from *Turrialba* Vol. 30, N° 2, pp. 209-218. Quarter April-June 1980

Cacao descriptors, their states and **modus operandi** *1/

J. M. M. ENGELS**, B. G. D. BARTLEY***, G. A. ENRIQUEZ****

COMPENDIO

Para estandarizar la descripción de los clones y poblaciones de cacao, se presenta una lista de descriptores y sus respectivas clases. También se dan instrucciones cortas para su uso, incluyendo el tamaño de la muestra mínima para las características cuantitativas.

Introduction

THE reasons to compile an extensive list of cacao descriptors are many. First of all is the standardization of the descriptive terminology to permit an exchange of information between scientists working with cacao genetic resources. Secondly, to facilitate an inventory of what is available worldwide in existing cacao collections and, consequently, determine what valuable accessions should be duplicated in other places. Thirdly, to help the breeder in selecting better material, not present in his breeding programmes. Fourthly, the methods of computer-assisted data processing need information about individual accessions related to descriptors (descriptive terms, in general of plant characteristics) and their states (gradation in the expression of a descriptor). This ensures a quick and adequate transfer of the collected data into machine-readable form, and its efficient storage and retrieval. Fifthly, to enable efficient management and maintenance of the collection. Since the latter is not uniformly handled, an additional set of descriptors should be developed locally. This set should also include descriptors for germplasm distribution and use. A final but important aspect is that the development and standardization of crop specific descriptors is the basis for a systematic description of germplasm collections. Chang (1) listed some of the advantages of such a systematic description:

a) characterization of cultivars or breeding lines of national and international interest;

b) differentiation between accessions with identical or similar names;

c) identification of accessions with desired characteristics;

d) classification of cultivars based on reliable data;

e) development of interrelationships between characteristics and also between geographical groups of cultivars; and

f) estimation of the variation available within the collection.

Most descriptors presented here are based on an unpublished revision of the literature (2), and an evaluation of the selected descriptors in the Genetic Resources Project at CATIE, Costa Rica.

Methodology

In so far as the descriptors are not self-explanatory, short instructions or comments are given to facilitate their use and to make them unequivocal. The mean and stardard deviation of the measured sample should be given for quantitative characteristics of a clone, which are expressed on a continous scale. The minimum

* Received for publication February 15th, 1980.

1/ Part of the present work was supported by the German Agency of Technical Cooperation (GTZ) with funds from the German Ministry of Economic Cooperation. We wish to express our thanks to Drs. M. Jackson and J. R. Palmer for carefully reading the English manuscript.

** Plant Breeder/Documentalist, Plant Genetic Resources Project CATIE/GTZ, Turrialba, Costa Rica.

*** Geneticist, IICA, CEPLAC, Itabuna, Bahía, Brazil.

**** Head of the Perennial Plants Programme, CATIE, Turrialba, Costa Rica.

sample size 'r' is indicated after descriptors of quantitative characteristics. Values for 'r' have been calculated in a preliminary study so that the sample mean would fall within five percent of the population mean ninety five times in a hundred; using the formula (3):

$$r \geq 0.16 \left[\frac{(S)\ (100)}{\bar{x}} \right]^2$$

in which 'r' represents the "minimum sample size", 0.16 a constant, S the standard deviation, and $\bar{x}$ the mean of a sample. If the calculated minimum sample size (r) is large for practical purposes, a smaller sample size, a fraction of r may be chosen arbitrarily. When according to the circumstances and the authors experience, this arbitrary value is used, the notation n is employed.

In the preliminary study (2) differences for fruit and flower characteristics between trees of the same clone, grown at the same site, did not reach statistical significance. Thus the fruits and flowers of all the trees of one clone can be used and mixed without indicating the number of trees. In spite of this, it is recommended to study fruits and flowers of several trees of one clone to ensure the determination of possible mixtures within an accession.

If a particular accession represents a population, the ranges of the phenotypic expression of the characteristics should be given, if possible, with a note on the frequency distribution.

Three types of descriptor states will be found. The first type is an open one, such as 'accession number' or 'leaf length in cm'. The second type consists of fixed state descriptors which do not have a continuous expression. These are arbitrarily coded, generally commencing with '1'. Examples of this type are population state', 'collecting source' and 'leaf base shape'. If the descriptor states of characteristics with a continuous expression are classified, the third type, a scale from 1 to 9 is used. Class '1' always represents the lowest, smallest, etc., expression, and '9' the highest, greatest, etc. In general, only some classes of the whole scale are given, e.g. 3 = weak, 5 = intermediate and 7 = vigorous, from the scale ranking from 1 = very weak to 9 = very vigorous, for the descriptor 'vigour'. This does not imply that those not mentioned cannot be used. The presence of an unclassified characteristic is indicated by '+'. When the expression of a characteristic is not measured, or the information is lacking, a dash '—' should be used.

Some of the descriptors are marked 'optional'. Further investigation is needed on the correct use and classification of their states.

Instruments and apparatus indispensable for a systematic description are: stereo microscope (for flower characteristics), magnifying glass, different sizes of vernier calipers, tweezers, dissecting needles, scalpels and glycerine. A camera can be very useful for the recording and determination of shapes and colour patterns; a scale must be included in each photograph.

To increase the information content of data garthered under determined environmental conditions, the use of at least one world-wide accepted standard clone is strongly recommended. The results of the description of this (these) clone(s) should be used to adjust, if necessary, the classes of these descriptors, whose expressions are strongly influenced by the environment. For purpose of comparison, detailed information on the climatic and soil conditions of the germplasm collection site should be added to the descriptive data.

List of descriptors

1. Accession identifier

This identifier is recorded when an accession enters a genetic resources centre or germplasm collection. It consists of three descriptors: the first is a unique number, the second and third represents the country and locality of the genetic resources centre or collection, respectively. The combination of these three descriptors is unique world-wide.

1.1 *Accession number*

This is a number intended to serve as a unique identifier for each accession. This number once assigned can never be reassigned, even when an accession becomes extinct.

1.2 *Country genetic resource centre*

If the complete name is not used, one of the following abbreviations should be given:

ANG = Angola
BEL = Belize
BOL = Bolivia
BRZ = Brazil
CAM = Cameroon
CAR = Caribbean islands (not specified)
CDR = Congo
CIV = Ivory Coast
CLB = Colombia
CRI = Costa Rica
CUB = Cuba
DOM = Dominican Republic
ECD = Ecuador
EGU = Equatorial Guinea
ELS = El Salvador
GHA = Ghana
GRE = Grenada
GUA = Guatemala
GUI = Guiana
HAI = Haiti
HON = Honduras
IND = India
JAM = Jamaica
MAI = Malawi
MAL = Malaysia
MEX = Mexico

NIC = Nicaragua
NIG = Nigeria
OCA = Oceania Islands (not specified)
PAN = Panama
PER = Perú
PNG = Papua, New Guinea
PRI = Puerto Rico
RIN = Rep. of Indonesia
STP = St. Tomé & Principe
SUR = Surinam
TRT = Trinidad & Tobago
USA = United States of America
VEN = Venezuela

2. Nomenclature

Since all the descriptors refer to cacao *(Theobroma cacao* L.) the genus and species name can be disregarded. However, the clonal name and its synonyms are very important for the identification of cacao cultivars.

2.1 *Accession name*

The current name for clone, cultivar, population, etc., is given by the 'original' experimental station. These names are generally alpha-numeric or alphabetic identifiers.

2.2 *Synonyms*

These include any previous identification other than the current name, collection number, newly assigned station name or number and/or vernacular name(s) are frequently used as identifier.

3. Origin

A set of data that specifies the genetic origin of the accession, including the techniques used in breeding work.

3.1 *Population state*

The 'breeding' state of a population from where an accession was taken can be: 1) spontaneous –a population not cultivated and which is unexploited by man; 2) primitive, but cultivated– the original, spontaneous population is unknown; 3) derived –the original population from which it is derived is known. This group includes all types of breeding material. The code is expressed as:

1 = spontaneous
2 = primitive cultivated
3 = derived

3.2 *Descent*

This code refers to the way an accession is derived from an ancestral population. This can be by natural or open pollination or by artificial pollination.

1 = natural pollination
2 = artificial pollination

3.3 *Breeding method*

Represented by a coded specification describing the way in which the artificial pollination was conducted in a breeding programme, expressed as:

1 = selfing (S)
2 = hybridization (F)
3 = backcross (BC)

3.4 *Generation*

The actual generation of an accession in a breeding programme. The generation number should be preceded by a corresponding abbreviation given in 3.3.

3.5 *Pedigree*

A register recording a line of ancestors. As much information as possible should be given; when the male parent is unknown, information on the female can be very useful.

3.6 *Utilization data*

Descriptor attemps to classify accession in accordance with use of the clone in respect to commercial planting or breeding properties; information on previous or present use.

1 = cultivar
2 = foundation parent
3 = disease resistant source
4 = mutation

4. Geographical origin

A set of data which specifies the geographic origin and precise site from where a certain accession was collected, selected or bred.

4.1 *Country*

The full name or an abbreviation — as given under 1.2 — for the country in which a particular germplasm accession was collected, selected or bred.

4.2 *Political subdivision*

The name representing the political or administrative subdivision of the country in which a particular accession was collected. Examples are the names of a state, province, county, etc.

4.3 *Locality*

The specific name of the town, village or, if relevant, area in which the germplasm accession was collected. If necessary a short description of the exact site should be given, for instance 10 km north of..., along river... An alternative is the geographical coordinates of the collection site.

4.4 *Collecting source*

Self - explanatory. In case '4' is used as code, this should be specified.

1 = natural habitat
2 = farm
3 = experimental station
4 = other

4.5 *Name of source*

The name of owner of the farm, experimental station or 'other' should be given.

5. Donor identifier

A set of data which identifies the donor of an accession.

5.1 *Donor name*

The name of the person or institution responsible for donating germplasm to a collector.

5.2 *Donor number*

A number or an alpha-numeric identifier assigned to an accession by the donor.

6. Taxonomic and morphological data

Data of plant characteristics which are mainly collected for the characterization and identification of a population or clone, which are usually not directly related to the yield of the crop. However, relevant information for breeders is also included.

6.1 Plant characteristics

Data which describe the vegetative parts of the cacao tree.

6.1.1. *Architecture*

An average observation of several trees of a clone or a population should be given. The observations can be made by estimating the vertical angle between two opposite main branches. If the angle(s) is $\leq 90°$, the type is called erect; between 91 and 135°, intermediate; and $\geq 136°$, pendulous. The code is the following:

1 = erect
2 = intermediate
3 = pendulous

(If the angle refers to the trunk of the tree, appropriate adjustment should be done).

6.1.2 *Branch formation*

The classification is based on the existence of a single main branch and three or more branches (= verticillate) per ramification at the same height of the trunk

1 = single
2 = intermediate
3 = verticillate

6.1.3 *Vigour*

Code refers to the general appearance (growth) of an accession, and should be based on observations of several trees.

3 = weak
5 = intermediate
7 = vigorous

6.2 Leaf characteristics

6.2.1 Leaf shape

Numeric data are used to describe leaf shape. The minimum sample size for these descriptors has to be calculated; the mean and standard deviation should be given.

6.2.1.1 *Length from base, in cm* (L), $\bar{x}$ and S ($n = 15$)*

6.2.1.2 *Width at widest point, in cm* (W), $\bar{x}$ and S ($n = 15$).

6.2.1.3 *Length/width ratio* (L/W), $\bar{x}$ and S ($t = 15$).

6.2.1.4 *Length from base to widest point, in cm* (LBW), $\bar{x}$ and S ($n = 15$).

* n is the recommended 'sample size'; however, the calculated 'minimum sample size' (t) is larger.

6.2.1.5 *Ratio length/length from base to widest point* (L/LBW), $\bar{x}$ and S (n = 15). Sample as 6.2.1.1 and 6.2.1.4.

3 = ratio L/LBW < 2, shape is ovate

5 = ratio L/LBW = 2, shape is elliptic

7 = ratio L/LBW > 2, shape is obovate

6.2.2 *Leaf base*

The shape of base can be expressed in terms of the angles which the margins form with the central vein at its point of insertion. If the total angle is $\leq$ 90°, the leaf base is 'acute;' $\geq$ 90° 'obtuse' and $\pm$ 180°, 'rounded.' If leaf base is embayed in a sinus whose sides are straight or convex, 'cordate' is used. This observation should be based on several mature leaves of a tree and the code is expressed as:

1 = acute
2 = obtuse
3 = rounded
4 = cordate

6.2.3 *Leaf apex*

The shape of that portion of the leaf which is bounded by approximately the upper 15% of leaf margin. If the angle of the margins is $\leq$ 90°, both with straight and convex margins, the apex is 'acute'. If the tip is with margins markedly concave, the apex is acuminate. The tip can be short or long. This observations should be based on several leaves of a tree.

1 = acute
2 = short acuminate
3 = long acuminate

6.2.4 *Leaf petiole*

Petioles of some clones do not have very distinct pulvini or thickening, other clones have noticeable pulvini.

0 = without noticeable pulvini
1 = with noticeable pulvini

6.2.5 *Leaf texture*

If the appearance of mature leaves is opaque, like writing paper, the term chartaceous is used; when leathery, thick and stiff, the leaves are coriaceous. In case of 'other' observations, details should be given.

1 = chartaceous
2 = coriaceous
3 = other

6.2.6 *Young leaf colour*

Data which describe the absence or presence of anthocyanin in the young flush.

6.2.6.1 *Anthocyanin absent*

3 = light green
5 = intermediate
7 = intense green

6.2.6.2 *Anthocyanin present*

3 = light reddish
5 = intermediate
7 = intense reddish

6.3 Flowering characteristics

The flowering habit of cultivars can be described in terms of:

6.3.1 *Flowering intensity* (optional)

The number of flowers per cushion and the number of cushions per tree are involved in this descriptor.

6.3.1.1 Number of flowers per cushion, $\bar{x}$ and S (n = 35).

6.3.1.2 Number of cushions per tree, $\bar{x}$ and S (n = 10).

6.3.2 *Flowering pattern* (optional)

This refers to the distribution of the flowering activity during the year; it may be continuous or with one or more peaks per year.

6.4 Flower characteristics

Data are taken from two to four recently opened flowers of each of five trees.

6.4.1 *Peduncle colour*

Because there is much variation within trees, depending on light conditions, only three classes are established:

1 = green
2 = green with reddish
3 = reddish

6.4.2 *Anthocyanin in outer sepal*

Several flowers of different trees should be observed:

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.3 *Sepal length, in mm.,* $\bar{x}$ *and S* (r = 20)

6.4.4 *Sepal width at widest point, in mm,* $\bar{x}$ and S (r = 20)

6.4.5 *Sepal length/width ratio,* $\bar{x}$ and S (n = 20)

6.4.6 *Orientation of sepals*

Several flowers should be observed due to variation. Only two classes are distinguished: 'reflexed' with sepals bent backward, and with sepals more or less 'horizontal';

1 = reflexed
2 = horizontal

6.4.7 *Length of petal ligule, in mm,* $\bar{x}$ and S (r = 20).

Distance between point of insertion of isthmus in hood and apex of ligule.

6.4.8 *Width of petal ligule at widest point, in mm,* $\bar{x}$ and S (r = 20)

6.4.9 *Petal ligule length/width ratio,* $\bar{x}$ and S (r = 20)

6.4.10 *Anthocyanin in petal ligule*

0 = absent
1 = present

6.4.11 *Anthocyanin in stamen filament*

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.12 *Staminode length, in mm,* $\bar{x}$ and S (r = 10)

6.4.13 *Anthocyanin in staminode*

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.14 *Ovary length, in mm,* $\bar{x}$ and S (r = 15)

6.4.15 *Ovary width at widest point, in mm,* $\bar{x}$ and S (r = 10)

6.4.16 *Anthocyanin in upper part of ovary*

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.17 *Anthocyanin in lower part of ovary*

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.18 *Maximum ovule number per ovary* (r = 5).

Since the ovary has five loculi, 40, 45, 50, etc. ovules will generaly be found with only slight deviation within each clone. Complete counts should be taken of at least 5 ovaries.

6.4.19 *Anther disposition*

In some clones anthers are missing, in others the anthers are not covered by the hood as in the cultivar 'P-11', this fact should be noted under 'other'.

0 = anthers absent
3 = normal
7 = other types

6.4.20 *Style length, in mm,* $\bar{x}$ and S (r = 10)

6.4.21 *Anthocyanin in lower half of style*

0 = absent
3 = slight
5 = intermediate
7 = intense

6.4.22 *Self-incompatibility*

0 = absent
1 = present

6.5. Fruit characteristics

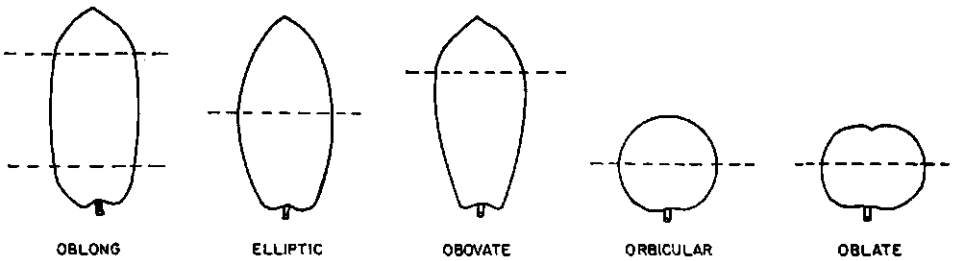
Qualitative and quantitative characteristics of unripe (4 months) and ripe cacao fruits or pods are described here.

6.5.1 *Fruit shape*

This coded information is based on several observations of mature fruits. For 'oblong' fruits the margins of the middle part of the fruit are parallel or nearly so with the long axis. If the perpendicular axis of the greatest width is close to the midpoint of the fruits axis and the margins are convex, the shape is 'elliptic'. If the axis

of greatest width cuts the long axis of the fruit apical to the midpoint, the shape is 'obovate'. More or less round fruits are called 'orbicular' and if the fruit width exceeds the length, 'oblate'.

1 = oblong
2 = elliptic
3 = obovate
4 = orbicular
5 = oblate



6.5.2 *Basal constriction*

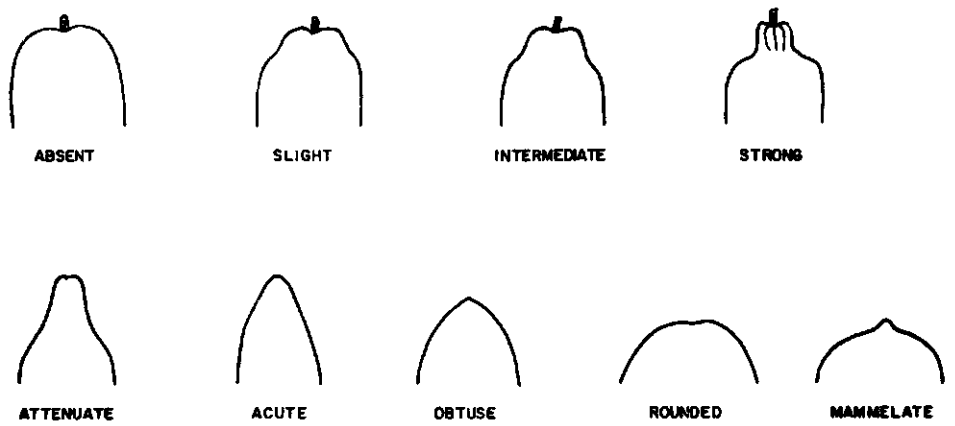
The code representing the constriction or 'bottle neck' of the basal part of the matute fruit is expresed as:

0 = absent
3 = slight
5 = intermediate
7 = strong

6.5.3 *Apex form*

The code representing the form of the apical part of the mature fruit is expressed as:

1 = attenuate
2 = acute
3 = obtuse
4 = rounded
5 = mammelate



6.5.4 *Fruit length from base, in cm,* $\bar{x}$ and S (r = 35)

6.5.5 *Fruit width at widest part, in cm,* $\bar{x}$ and S (r = 35)

6.5.6 *Fruit length/width ratio,* $\bar{x}$ and S (r = 35)

6.5.7 *Distance base to widest part, in cm,* $\bar{x}$ and S (r = 20)

6.5.8 *Length/distance base to widest part ratio,* $\bar{x}$ and S (r = 10)

6.5.9 *Weight of whole fruit in g,* $\bar{x}$ and S (n = 35)

6.5.10 *Fruit surface rugosity*

This codes refers to the visual observation of the absence or presence of protuberances on fruit surface

0 = absent
3 = slight
5 = intermediate
7 = intense

6.5.11 *Ridge pair appearance*

A code for the degree of separation of a pair of ridges. Extremes are fused pairs (e.g. 'Pentagona') and equidistant ones (e.g. 'Laranja,). A pair of ridges is always situated above a carpel.

0 = fused
3 = slightly separated
5 = intermediate
7 = well separated
9 = equidistant (individually)

6.5.12 *Primary furrow depth*

A code for the depth of the furrow between a pair of ridges is expressed as:

3 = superficial
5 = intermediate
7 = deep

6.5.13 *Fruit wall thickness*

Fruit wall is defined as comprising both the exocarp and mesocarp. The endocarp should not be considered.

6.5.13.1 *Thickness at ridge, in mm,* $\bar{x}$ and S (n = 35)

6.5.13.2 *Thickness at primary furrow, in mm,* $\bar{x}$ and S (n = 35)

65.13.3 *Thickness at secondary furrow, in mm,* x and S (n = 35)

(The furrow within a pair of ridges is defined as secondary furrow).

6.5.14 *Mesocarp hardiness*

A code representing the hardiness of mesocarp of fruits at least four months old. An objective method of measurement is recommended.

3 = soft
5 = intermediate
7 = hard

6.5.15 *Basic surface colour*

Only green exists as basic colour in unripe fruits, although the intensity can vary

3 = light
5 = intermediate
7 = dark

6.5.16 *Anthocyanin intensity of ridges*

The intensity of anthocyanin of the ridges of unripe fruits can be expressed as:

0 = absent
3 = slight
5 = intermediate
7 = intense

6.5.17 *Anthocyanin intensity in primary furrows*

The intensity of anthocyanin in the furrows of unripe fruits can be expressed as:

0 = absent
3 = slight
5 = intermediate
7 = intense

6.5.18 *Anthocyanin in ripe fruits*

The code representing absense (=yellow fruit) or presence (= reddish fruit) in different intensities of the anthocyanin in ripe fruits can be expressed as:

0 = absent
3 = slight
5 = intermediate
7 = intense

6.6 Seed characteristics

Data are taken from peeled seeds. The seeds should be taken at random.

6.6.1 *Wet weight of seed, in g,* $\bar{x}$ and S ($r = 100$)

Five seeds from each of 20 pods are used.

6.6.2 *Dry weight of seed, in g,* $\bar{x}$ and S ($r = 100$)

Five seeds from each of 20 pods are used.

6.6.3 *Seed length, in mm,* $\bar{x}$ and S ($r = 100$)
Five seeds from each of 20 pods are used.

6.6.4 *Seed width, in mm,* $\bar{x}$ and S ($r = 100$)
Sample as 6.6.3

6.6.5 *Seed thickness, in mm,* $\bar{x}$ and S ($r = 100$)
Sample as 6.6.3

6.6.6 *Seed form in longitudinal section*

Although the form of seeds can vary highly within a pod, the average form should be selected by using several seeds per pod with the embryo as the reference point. The code consists of the three following classes:

1 = oblong
2 = elliptic
3 = ovate

6.6.7 *Cotyledon colour, as a percentage* ($r = 10$)

Since the cotyledon colour depends also on the genotype of the male parent, controlled crosses (either selfing or a test cross) should be used to determine the coloration of the seeds. The code representing the colour or combination of colours (e.g. spotted) of the cotyledon is marked by giving the percentage of a colour class from the whole.

1 = ... % white
2 = ... % grayish-white
3 = ... % light purple
4 = ... % intermediate purple
5 = ... % dark purple
6 = ... % spotted

6.6.8 *Pulp colour*

The code representing the colour of fresh pulp is expressed as:

1 = white
2 = yellowish

6.6.9 *Fat content of cotyledons, as a percentage* ($r = 3$)

Only pods resulting from selfing or crosses with a standard clone should be used.
Content should be determined with a standard method and expressed as a percentage of the fresh seed weight.

7. Agronomic evaluation data

The information contained in this section summarizes the data obtained during the evaluation of the agronomic characteristics of the item, either with respect to its perfomance as a clone or to the seed-derived progenies. Since the data obtained refer to specific locations, it is necessary to specify the location and the conditions under which the evaluation was conducted. In view of the volume of information which could be ascribed to individual items and the fact that much of the information is relative in nature, the data should be supplemented by reference to appropriate publications.

7.1 Location of evaluation

7.1.1 *Name of country*

Alphabetic code or complete name (see 1.2)

7.1.2 *Name of institution*

7.1.3 *Period of report*

Specifies time or period during which evaluation was conducted.

7.2 Propagation characteristics

Includes information about the relative ease by which the accession may be multiplied vegetatively. Expressed on basis of the percentage of cuttings rooted or buddings taken.

7.2.1 *Cuttings*

1 = very difficult	< 30%
3 = difficult	31-40%
5 = intermediate	41-50%
7 = easy	51-60%
9 = very easy	> 61%

7.2.2 *Buddings*

3 = difficult	< 60%
5 = intermediate	61-80%
7 = easy	> 81%

7.3 *Period to fruit maturity*

Expressed as the number of days between flower fertilization and the physiological ripening of the fruit during normal cropping periods.

3 = short (< 154 days)
5 = intermediate (154 to 170 days)
7 = long (> 170 days)

7.4 Production as clones

Summary of the data relative to production of the accession as clone during defined and specified periods for the trials in which the clone has been evaluated, according to the way of establishing the plantation by:

7.4.1 *Cuttings*
7.4.2 *Buddings*
7.4.3 *Marcottings*

7.5 Production of progenies

Summary of the performance of the progenies of the accession obtained by sexual methods. The type of population, period of evaluation and location of the trial should be specified.

8. Environmental adaptability

8.1 *Reaction to drought*

Coded for observations about relative behaviour under unfavourable moisture (drought) regimes. Observations will include survival, production, speed of recuperation and leaf retention characteristics.

3 = tolerant
5 = intermediate
7 = susceptible

8.2 *Reaction to excessive soil moisture*

Coded for observations about relative behaviour under unfavourable moisture (excessive rainfall) regimes. Observations will include survival, production, speed of recuperation.

3 = tolerant
5 = intermediate
7 = susceptible

9. Disease and pest reaction data

Data of the reaction to particular organisms are recorded during evaluation at the site where the collection is maintained. Because of the variation in races of pathogens from country to country and even between locations, careful registration of reaction pattern and, if available, the source of information should be ensured. This category can be divided into subgroups: reaction to fungi, bacteria, viruses, nemotodes, insects, etc. In the following, only an example is given. Each germplasm centre should decide which are the locally important diseases and pests.

9.1 Reaction to fungal disease

This code describes the degree of reaction of an accession to a particular fungal pathogen, recorded by the reaction of a plant organ infected, and expressed as:

1 = very susceptible
3 = moderately susceptible
5 = moderately resistant
7 = very resistant
9 = extremely resistant

9.1.1 *Phytophthora palmivora*

9.1.2 *Crinipellis perniciosus*

9.1.3 etc.

9.2 Reaction to bacterial diseases

9.3 Reaction to virus diseases

9.4 etc.

Summary

In order to standardize the description of cacao clones and populations, a list of descriptors with their respective states is presented. Short instructions are given for their use, including the minimum sample size for the quantitative characteristics.

Literature cited

1. CHANG, T.T. Manual on genetic conservations of rice germplasm for evaluation and utilization. Los Baños, Philippines, International Rice Research Institute, 1976. 77 p.

2. ENGELS, J.M.M. Descriptores de cacao *(Theobroma cacao* L.). CATIE/GTZ, Turrialba, Programa Regional de Recursos Genéticos. 1977. 23 p.

3. POUND, J. F. The genetic constitution of the cocoa crop. *In* Imperial College of Tropical Agriculture, Trinidad. Annual Report on Cocoa Research 1:10-24. 1931.

Euphytica 32 (1983) 377–385

A SYSTEMATIC DESCRIPTION OF CACAO CLONES. I. THE DISCRIMINATIVE VALUE OF QUANTITATIVE CHARACTERISTICS

J. M. M. ENGELS[1]

Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica

Received 15 September 1981

INDEX WORDS

Theobroma cacao, cacao clones, discriminative value, quantitative descriptors, taxonomic identifiers.

SUMMARY

Data from 23 possibly distinctive quantitative flower and pod characteristics were collected from 32 cacao clones and analysed statistically. The discriminative value of each of them was calculated by using Duncan's new multiple range test to obtain a discriminatory value 'D' which was adjusted to allow for correlations with other characteristics. It is concluded that quantitative characteristics, which usually have a higher agronomic interest than qualitative ones, are also helpful in characterizing and identifying varieties or clones.

INTRODUCTION

An important activity of the Plant Genetic Resources Unit at Turrialba is the systematic description of its collections. The cacao collection at Turrialba with more than 475 accessions was only partly described in 1967 and, as a native crop of the Central American region, has been accorded high priority in the descriptive work.

A provisional list was composed of 87 quantitative and qualitative descriptors (ENGELS et al., 1980). Data for each of these were collected and analysed to select the most reliable and discriminative characteristics. The present study continues that work by a more detailed examination of the discriminatory value of 23 quantitative flower and pod descriptors. Plant habit and quantitative leaf characteristics are not considered as highly discriminating (OSTENDORF, 1956) and are of lesser agronomic importance.

ENRÍQUEZ & SORIA (1964, 1968) studied flowers and beans. They selected descriptive characteristics, based on the variance between and within clones, to facilitate the preparation of an international catalogue of cacao clones. ENRÍQUEZ (1966) included also leaf characteristics and found high variation within clones. OSTENDORF (1956) reviewed the characteristics most frequently used to describe clones in order to decide their discriminative value.

When the description of the whole collection has been completed, the data accumulated will provide a basis for: 1) characterizing the clones; 2) distinguishing accessions

[1] Present address: Plant Genetic Resources Centre/Ethiopia, P.O. Box 30726, Addis Abeba, Ethiopia.

with homonyms or similar names, and recognising duplicates; 3) identifying clones with a desired combination of characteristics; 4) taxonomically classifying clones; 5) detecting groups of correlated characteristics which may have immediate practical value, or which may give clues to genetic relationships among the clones; 6) estimating the variation within the collection, as indicated by CHANG (1976).

MATERIAL AND METHODS

The data for this study were from the first 32 clones to be described from the cacao germplasm collection planted at Turrialba. More details of the clones will be found in a later article in this series. The 32 clones were divided into two sets to check on the consistency of the statistical magnitudes for each of the descriptors and for a set of clones. The arbitrary division into sets was based on the planting sequence in the field.

The sets contain the following clones:

Set I	*Set II*
CC numbers 9, 38, 41, 44, 46, 74, 107, 121, 137, 139, 143 UF number 701	Diamantes -800, EEG-25 EET numbers 59, 75, 80, 164 Laranja, P-10, SGU-82 UF numbers 4, 11, 36, 168, 601, 654, 667, 672, 676, 677, 715

The average number of trees per clone was 10.4, ranging from 5 to 16 trees. All the trees of each clone are planted in one plot. The phenotypic variation between clones due to site differences could not be separated from variation due to hereditary factors. However, this is not relevant to the results from the method presented here, which involves calculation on mean values per clone.

Random samples of pods were collected from all the trees of each clone and mixed together, since there were no significant differences between trees-within-clones. The following features were selected for their agronomic importance and were measured on each pod: length, maximum width, weight and the thickness of the pod wall at two points (ridges, and furrows). The seeds were extracted, peeled, and counted, and their length, width, thickness, fresh weight and oven dry weight measured. Each of these variables was tested as a descriptor, as were the following derived variables: width/length ratio of pods, width/length ratio of seeds, seed per pod/weight of pod, and fresh weight/dry weight of seeds. Samples of flowers were collected at random from five different trees of each of the clones. In each fresh flower, the length (including the isthmus) and maximum width of one petal, one sepal and the ovary were measured and the length only of the style and one staminode. The flower characteristics were chosen for their possible taxonomic value. They were arranged on a glass slide with a drop of glycerol and carefully flattened under a cover slip and measured under a stereo microscope. The pod dimensions were measured with vernier calipers (mm) and the weights on a torsion balance (g).

A preliminary study was made to determine the sample size appropriate for each descriptor such that the sample mean would fall within five percent of the population

mean ninety five times in a hundred. If the mean of the population is μ and the standard deviation σ, the standard error of the mean of a sample of 'n' observations is $\frac{\sigma}{\sqrt{n}}$, and if repeated samples of 'n' observations are taken, their means will fall within the limits $\mu \pm \frac{2\sigma}{\sqrt{n}}$ in rather more than 95 per cent of cases. If $\frac{2\sigma}{\sqrt{n}}$ should not be more than 0.05 μ, whence 'n' must not be less than $\left(\frac{2\sigma}{0.05\mu}\right)^2 = 0.16\left(\frac{100\sigma}{\mu}\right)^2$ (POUND, 1931).

For practical reasons, the sample size for pods could not exceed 40 per clone in set I and 35 per clone in set II. For some descriptors these numbers are lower than the minimum sample size calculated with the above formula.

The number of measurements actually used for each descriptor is shown in the second column of Tables 1 and 2, and in the third column the calculated N_{min} values are given.

The data obtained for the several descriptors were subjected to analysis of variance and to correlation analysis. The ratio of the variance between and within clones ('P') gives an indication of how variable a certain characteristic is between clones. The higher the ratio, the more variation exists between the clones for this characteristic. ENRÍQUEZ & SORIA (1968) used the magnitude of this ratio as a discriminative value for typifying characteristics.

The ratio P does not, however, provide any information about the distribution of the clone average for a particular descriptor over the total range found within a group of clones. P will tend to be larger when individual clone means are clustered towards the extremes of the group range, than when they are normally distributed, assuming that within-clone variance remains the same. Significance of differences between clonal means for any one variable can be shown by multiple comparison tests such as Duncan's new multiple range test (DUNCAN, 1955; see also THOMAS, 1974), and the more strict Tukey's test (TUKEY, 1949). Such tests are unaffected by clustering of means.

Descriptors which yield the greatest number of significant differences will be the most efficient for identifying clones. If the number of significant differences, detected by Duncan's multiple range test, is expressed as a fraction of the total number of comparisons possible within the clone group, an index value 'D' is obtained, which measures the discriminatory power of the descriptor concerned. Comparison of D values among a set of descriptors will permit selection of those having greater discriminative value. The more clones that are included and the more they are related to each other, the lower the D values will be. With increasing number of clones, the absolute number of mis-classifications due to Type I and Type II errors will increase. However, the user is buffered against error by the use of several variables, in any one of which a mis-classification is statistically unlikely to be parallelled in the other variables. The D value will increase if the interclone variation for a given descriptor increases.

The discriminatory power D of the several descriptors may be further refined by correcting each for that part of the total information it shares with other descriptors. An overall 'D_w' value was calculated for each descriptor as the mean of the corresponding D values found for each of the clone sets I and II, weighted by number of clones

in a set. Then the descriptor with the highest D_w was chosen as the most promising and all the other D_w values were recalculated by subtracting that part of their distinctive value, which was shared with the most promising one, as is indicated by a significant positive or negative correlation.

The formula to re-calculate D_w was:

$$D'_w = D_w(1 - r^2),$$

where $1 - r^2$ is the coefficient of nondetermination, expressing the proportion of variance of a characteristic that has not been explained by another characteristic (SOKAL & ROHLF, 1969).

RESULTS AND DISCUSSION

Table 1 and 2 present the results for the flower and pod characteristics respectively, for the two sets of clones separately. Differences between the means for each of the descriptors in the two sets of clones, and also the coefficients of variability, are small. This indicates that the variation of a given descriptor is specific to that descriptor and more or less independent of which clones are measured. By comparing Tables

Table 1. Flower descriptors (in mm) and some of their statistical magnitudes.

Statistical magnitudes / Descriptors	Number of measurements per clone		9 clones, set I without CC-46, CC-107 and CC-143[1]				19 clones, set II without clone EEG-25[1]				set I & II	P values of ENRÍQUEZ & SORIA (1964, 1968)
	used	calculated	$\bar{x}$	C.V.	P	D	$\bar{x}$	C.V.	P	D	D_w	P
1. Sepal length	10	5	8.7	5.36	1.71	0.67	8.8	5.62	1.60	0.61	0.62	4.53
2. Sepal width	20	19	2.1	8.18	0.11	0.25	2.3	13.72	0.33	0.42	0.39	1.91
3. Petal length	15	11	7.4	7.54	0.27	0.36	7.8	6.19	1.75	0.63	0.58	2.16
4. Petal width	15	16	2.1	9.91	0.88	0.36	2.8	10.19	1.91	0.65	0.60	3.22
5. Staminode length	10	7	7.5	6.14	3.02	0.69	7.4	5.11	2.12	0.65	0.66	9.05
6. Ovary length	20	14	1.3	9.17	0.20	0.42	1.4	8.66	1.58	0.75	0.69	6.25
7. Ovary width	20	11	1.0	7.73	0.08	0.25	1.0	6.95	0.47	0.35	0.33	12.06
8. Style length	10	8	2.1	6.72	2.01	0.75	2.2	6.57	5.09	0.82	0.81	1.38
Mean				7.59	1.04	0.48		7.88	1.86	0.67	0.59	5.07

[1] The exclusions were due to the lack of sufficient flowers at the time of measurement.
Number of measurements = total number of flowers measured per clone, collected at random from five trees.
In each flower, one sepal, petal and staminode were randomly chosen and measured.
$\bar{x}$ = mean of all the means of the clones for each descriptor.
C.V. = mean of coefficients of variation of all the clones for each descriptor.
P = ratio of the variance between clones to the variance within clones.
D = number of significant differences, at 5% probability level, found between clones expressed as a fraction of the maximum possible number of significant differences, based on the new multiple range test of Duncan.
D_w = the weighted mean of the number of significant differences among clone combinations calculated over both groups.

Table 2. Pod and seed descriptors measured in mm or grams and some of their statistical magnitudes.'

Descriptor \ Statistical magnitudes	Number of measurements per clone: used	calculated	12 clones, set I: $\bar{x}$[1]	C.V.	P	D	20 clones, set II: $\bar{x}$	C.V.	P	D	I & II: D_w	P values of ENRÍQUEZ & SORIA (1964)
1. Seed length	Average of 5 seeds/ pod & 20 pods[3]	5 seeds/ pod and 15 pods	25.5	5.52	1.08	0.62	25.5	6.27	3.62	0.85	0.79	3.84
2. Seed width			14.8	5.26	1.87	0.65	14.0	6.37	3.70	0.83	0.78	9.47
3. Seed thickness			9.4	9.66	0.62	0.53	–	–	–	–	0.53	4.02
4. Seed width/seed length			0.58	6.47	0.90	0.59	0.55	6.46	0.55	0.35	0.41	–
5. Pod length	35 pods[2]	28	156.7	11.72	0.38	0.48	166.0	13.26	2.05	0.61	0.57	0.66
6. Pod diameter	20 pods	14	83.5	8.62	0.31	0.44	85.0	9.24	0.51	0.46	0.46	0.95
7. Pod diameter/pod length	20 pods	27	0.53	12.88	0.18	0.36	0.53	8.03	10.08	0.67	0.59	–
8. Seed number per pod	35 pods[2]	201	31.3	26.79	0.22	0.36	29.7	35.40	0.27	0.54	0.49	0.28
9. Pod weight	35 pods[2]	145	500.6	29.79	0.35	0.39	532.1	30.07	0.75	0.52	0.48	1.02
10. Seed number per pod/pod weight	35 pods[2]	175	0.064	27.12	0.45	0.67	0.063	33.06	2.23	0.74	0.72	–
11. Wet seed weight	15 seeds per pod and 10 pods per clone[3]	8 seeds/ pod of 10 pods per clone	33.4	13.31	0.90	0.62	32.4	19.77	2.27	0.68	0.67	7.02
12. Dry seed weight			21.2	14.16	0.80	0.52	20.3	21.35	1.97	0.66	0.62	5.10
13. Dry seed weight/ wet seed weight			0.63	11.26	0.06[5]	0.06	0.63	7.78	0.70	0.45	0.35	–
14. Pod wall thickness (furrow)	40 pods	37	7.6	15.18	0.63	0.59	–	–	–	–	0.59	–
15. Pod wall thicknesss (ridge)	40 pods	26	11.7	12.61	0.83	0.45	–	–	–	–	0.45	25.55
Mean[4]				14.41	0.63	0.48		16.42	2.39	0.61	0.57	5.79

[1] For the explanations of $\bar{x}$, C.V., P and D see Table 1.
[2] The sample size per clone for the clones of set I is 40 pods.
[3] The 5 or 15 seeds per pod measured for various descriptions were randomly chosen.
[4] In the calculation of the means the descriptors number 3, 14 and 15 are not included, since they were omitted in set II.
[5] The F-value is not significant at the 5% level.

Table 3. Comparison of the discriminatory power D_w of several descriptors calculated with Duncan's New Multiple range Test and Tukey's Test.

Descriptor	Duncan's D_w	Tukey's D_w	Tukey/Duncan	$\bar{P}$[1]
pod diameter	0.46	0.30	0.65	0.41
pod length	0.61	0.54	0.89	1.22
seed number/pod	0.54	0.35	0.64	0.25
seed length	0.85	0.81	0.95	2.36
seed width	0.83	0.81	0.98	2.79

[1]Average ratio of both sets of clones of the variance between clones to the variance within clones.

1 and 2, it can also be observed that the average variation of flower descriptors within clones is about half of the average variation of pod descriptors. The variation of the different pod descriptors, with C.V.'s ranging from 5.26 to 35.40, is greater than that of the flower descriptors (5.11 to 13.72). In general therefore, a smaller sample size can be used for flower characteristics than is needed for most pod characteristics.

Comparison of the P values given in the right and left halves of tables 1 and 2 shows that, for almost all descriptors, the P value for set I clones is smaller than the

Table 4. The significant correlation coefficients between descriptor pairs, based on the standardized means of 32 clones (set I and II).

Descriptors positive correlations	Negative correlations descriptor numbers									
	1	2	3	4	5	6	7	8	9	10
1. Seed length							−0.51**	−0.48**		−0.58**
2. Seed width	0.88**							−0.48**		−0.58**
3. Seed thickness[1]		0.58*					−0.77**	−0.88**		−0.60*
4. Seed width/length		0.35*								
5. Pod length	0.50**						−0.87**			−0.78**
6. Pod diameter	0.37*				0.54**					−0.64**
7. Pod diameter/length				0.44*					−0.61**	
8. Seed number per pod										
9. Pod weight	0.50**	0.44**			0.87**	0.77**				−0.81
10. Seed no. per pod/pod weight							0.70**	0.45*		
11. Wet seed weight	0.92**	0.88**	0.75**		0.51**	0.42*			0.58**	
12. Dry seed weight	0.92**	0.89**	0.71**		0.54**	0.40**			0.60**	
13. Dry seed weight/wet seed weight										
14. Pod wall thickness (furrow)[1]					0.70**	0.60*				
15. Pod wall thickness (ridge)[1]				0.53*	0.87**	0.61*			0.55*	
16. Sepal length					0.47**	0.49**			0.57**	
17. Sepal width							0.35*			
18. Petal length					0.56**	0.39**			0.50**	
19. Petal width										
20. Staminode length		0.39*								
21. Ovary length										
22. Ovary width										
23. Style length										

[1]The correlation coefficients of these descriptors are calculated only with the means of the 12 clones of set I.
* = Significant at 5% level; ** = significant at 1% level.

corresponding value in set II. This indicates that set I is more uniform than set II with respect to both flowers and pod characteristics and suggests that set I clones may be more closely related genetically than the clones of set II. In a later article in this series it will be shown that phenotypically the clones of set I are much more similar to each other than the clones of set II. From the Tables 1 and 2 it can be estimated that D values are higher in clone set II for almost all the descriptors, despite the higher number of clones in set II. This supports the above conclusion, that the clones of set II are less related to each other than those of set I.

The mean values of D of all the flower and all the pod descriptors of the two sets are remarkably similar and may indicate that D is a good measure of the discriminatory value of the descriptors used.

As noted before, the size of D depends on the number of clones or varieties used in the test and on the relationship between these clones. Therefore, the calculated D values are relative and will only give information on the discriminative value of descriptors for that group of varieties or clones which were included in the calculations. However, there are objections to the use of Duncan's test with large numbers of multiple comparisons (THOMAS, 1974). Results from Tukey's test (TUKEY, 1949) on the same data are given in Table 3. D_w values are similar from the two tests if the ratios (P) of the variance between clones to the variance within clones are high. Characteristics

11	12	13	14	15	16	17	18	19	20	21	22	23
												−0.38**
												−0.59**
												−0.50**
−0.44*	−0.46**			−0.43*			−0.58**		−0.37*			
−0.60**	−0.55**							−0.43*				
−0.72**	−0.72**			−0.64**			−0.58**	−0.40*				
									−0.36*			−0.42*
0.98**												−0.41*
			0.85**									
				0.58**	0.61**							
						0.60**	0.50**					
	0.36*				0.39*		0.51**					
										0.58**		
							0.36*					

with a low P value, such as pod diameter and seed number per pod, show D values calculated with Tukey's method which are about 35% lower than the results obtained with Duncan's test.

Table 4 presents the significant correlation coefficients between descriptor pairs, as used in the recalculation of the D_w values. Style length, $D_w = 0.81$, was chosen as the most promising descriptor, and the D_w values of 'seed length', 'seed width', 'seed width/length' ratio, 'wet seed weight', 'dry seed weight' and 'sepal length' all had to be diminished in respect of information that they shared with 'style length'. Then the next best D_w' was chosen (= 'seed number per pod/pod weight ratio, $D_w = 0.72$) and the D_w' values of the remaining descriptors were again reduced. This procedure was repeated for successive descriptors to obtain the results presented in Table 5.

Although the D_w' values from 'seed number per pod' onwards are very small, their value was calculated in order to place all the descriptors in order. The order of preference for the descriptors would change if other selections criteria, such as minimum sample size or complication of the measurements, were taken into consideration. Due to the local situation and the subjective aspects of the initial choice of variables to be measured,these points were not considered.

Although OSTENDORF (1956) did not quantify his findings about the discriminative value, his observations agree closely with the results described here. The discriminatory values P of some bean characteristics presented by ENRÍQUEZ & SORIA (1968) are all higher than the P values of the present study. The reason is that the sampling method which was used – collection of the material in a very short period of time – resulted in a small variation within clones. However, the relative order of the clones is very similar in both studies. The smaller D values calculated from the data of Enríquez and Soria were due to the clustering of the clones into two distinct groups of 8 and

Table 5. The discriptors of quantitative cacao characteristics ordered by their respective distinctive value after the correction for correlations, as found for 32 clones.

Descriptor	D_w[1]		D_w[1]		D_w[1]
1. Style length	0.81	9. Dry seed weight/wet seed weight	0.35	17. Pod length	0.04
2. Seed number per pod/pod weight	0.72	10. Seed thickness	0.34	18. Seed width	0.03
3. Ovary length	0.69	11. Seed width/seed length	0.31	19. Pod diameter/pod lenght	0.007
4. Staminode length	0.66	12. Sepal width	0.25	20. Pod weight	0.003
5. Pod wall thickness (furrows)	0.59	13. Ovary width	0.22	21. Pod wall thickness (ridges)	0.0005
6. Sepal length	0.53	14. Petal length	0.11[2]	22. Wet seed weight	<0.0005
7. Petal width	0.50	15. Pod diameter	0.11	23. Dry seed weight	$\ll 0.0005$
8. Seed length	0.45	16. Seed number per pod	0.06		

[1]The proportion of the number of combinations of clones that can be distinguished significantly at the 5% level by using D_w or its value corrected for correlations (= D_w').

[2]Petal length is taken first because of the lower minimum sample size per clone that is needed than for pod diameter.

26 clones respectively. Findings from ENRÍQUEZ & SORIA (1964) also agree in part with the present study.

From the D_w' values of the first five discriptors shown in Table 5, it can be estimated that, theoretically, more than 99 per cent of all the possible combinations of the 32 clones examined can be separated significantly. This indicates that quantitative characteristics of cacao, which are generally of higher agronomic interest than the qualitative ones, are useful in the characterization and identification of clones or varieties in a varietal collection.

In a future publication the calculation method of the discriminative value of qualitative descriptors will be presented and the practical compatibility of the discriminative values of both types of characteristics will be shown. In a later article the relations between descriptors and between clones will be shown, using conventional numerical taxonomic methods.

ACKNOWLEDGEMENTS

The present work is part of the Plant Genetic Resources Unit at CATIE, supported by the German Agency for Technical Cooperation (GTZ) with funds from the German Ministry of Economic Cooperation at Bonn, Federal Republic of Germany.

I wish to express my thanks to Mr R. Bonilla and Mr R. Sánchez for collecting the data, to Mr J. Palmer and Mr K. Reinink for fruitful discussions and to Dr W. G. Dyson for carefully reading the manuscript.

REFERENCES

CHANG, T. T., 1976. Manual on genetic conservation of rice germ plasm for evaluation and utilization. International Rice Research Institute, Los Baños, Philippines. 77 p.

CHEESMAN, E. E., 1944. Notes on nomenclature, classification and possible relationships of cacao populations. Trop. Agric. (Trinidad) 21: 144–159.

DUNCAN, D. B., 1955. Multiple range and multiple F test. Biometrics 11: 1–42.

ENGELS, J. M. M., B. G. D. BARTLEY & G. A. ENRÍQUEZ, 1980. Cacao descriptors, their states and modus operandi. Turrialba 30: 209–218.

ENRÍQUEZ, G., 1966. Selección y estudio de los caracteres de la flor, la hoja y la mazorca, últiles para la identificación y descripción de cultivares de cacao. Tesis de Mag. Sci., Turrialba, Costa Rica. 97 p.

ENRÍQUEZ, G. & J. SORIA, 1964. Selección y estudio de los caracteres útiles de la flor para la identificación y descripción de cultivos de cacao. Am. Soc. Hort. Sci., Caribbean Region 8: 139–158.

ENRÍQUEZ, G. & J. SORIA, 1968. The variability of certain bean characteristics of cacao (*Theobroma cacao* L.). Euphytica 17: 114–120.

OSTENDORF, F. W., 1956. Identifying characters for cacao clones. In: Reuniao do Comité Técnico Interamericano de Cacau, 6th, Bahia, Brasil 1956. Salvador. Brasil, Instituto de Cacau de Bahia. pp. 89–110.

POUND, J. F., 1931. The genetic constitution of the cacao crop. In: Imperial College of Tropical Agriculture, Trinidad. A. Rep. Cacao Res. 1: 10–24.

SOKAL, R. R. & F. J. ROHLF, 1969. Biometry. The principles and practice of statistics in biological research. W. H. Freeman and Company, San Francisco. 776 p.

THOMAS, D. A. H., 1974. Error rates in multiple comparisons among means results of a simulation exercise. Applied Statistics 23: 284–294.

TUKEY, J. W., 1949. Comparing individual means in the analysis of variance. Biometrics 5: 99–114.

Euphytica 32 (1983) 387–396

A SYSTEMATIC DESCRIPTION OF CACAO CLONES. II. THE DISCRIMINATIVE VALUE OF QUALITATIVE CHARACTERISTICS AND THE PRACTICAL COMPATIBILITY OF THE DISCRIMINATIVE VALUE OF QUANTITATIVE AND QUALITATIVE DESCRIPTORS

J.M.M. ENGELS[1]
Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica

Received 15 September 1981

INDEX WORDS

Theobroma cacao L., cacao clones, discriminative value, qualitative characteristics, taxonomic identifiers.

SUMMARY

A method is presented for calculating the discriminative value of qualitative descriptors, based on the number of taxa pairs that a certain descriptor can separate and on the amount of information that this descriptor shares with other descriptors of the same study. To make this method compatible with the calculation of the discriminative power of quantitative characteristics, statistical tests are proposed to enable a comparison between both types of characteristics. The data from 32 cacao clones are used to demonstrate the methodology for 15 qualitative characteristics of leaves, flowers and fruits of the cacao tree. The combined use of qualitative and quantitative descriptors in the calculation of their respective discriminatory power is also presented.

INTRODUCTION

In plant taxonomy non-numerical characteristics such as shape, colour and texture are frequently used. This type of qualitative descriptors in general has states which are mutually exclusive, that is: discontinuous phenotypic expressions of a certain characteristic. In this study characteristics which describe a semi-continuous pattern, such as anthocyanin intensity in certain organs, are also included in the group of qualitative characteristics, although they could be considered as quantitative characteristics (MORSE et al., 1971; ROGERS & FLEMING, 1964).

The qualitative characteristics play an important role in the classification of taxonomic units (taxa), since the probability of mis-classification is smaller. Examples of classifications of the cacao crop are given by POUND (1932) using mainly pod shape, CHEESMAN (1944) in combining pod characteristics and geographical distribution, and OSTENDORF (1956) basing his classification on pod shape and pod surface structure.

In ENGELS (1983) the discriminative value of quantitative characteristics, with continuous and discrete expressions, was estimated to check their value in taxonomic

[1] Present address: PGRC/E, P.O. Box 30726, Addis Abeba, Ethiopia.

classification. In the present paper the discriminatory powers of qualitative characteristics are estimated, so far as possible, to give compatibility with results from quantitative characteristics.

ESTABROOK (1967) presents a model for analysing non-quantitative descriptors with help of the Information Theory. The interdependency of characteristics can be calculated and can give an indication for their usefulness in a classification. However, this measure of distance and similarity of descriptors is not compatible with the methods presented in ENGELS (1983) to calculate the discriminative value of quantitative descriptors. It is also more complicated. SNEATH & SOKAL (1973) list several methods used in numerical taxonomy to calculate discriminatory power, mainly of two-state characteristics (type 0, 1). PANKHURST (1978) gives two methods for calculating which characteristics are better for constructing taxonomic keys than others. The measure of the usefulness of a (two or multi-state) descriptor is called the separation coefficient.

The number of states of a qualtitative descriptor plays an important role in the discriminatory value between taxa of that characteristic, comparable to the ratio of the variation between clones to the variation within clones for quantitative characteristics. ESTABROOK & ROGERS (1966) express it as 'the more states, the greater splitting power, the less grouping power'. It is important that the states of a characteristic be contrasting and mutually exclusive (PANKHURST, 1978). This aspect is also discussed in ENGELS et al. (1981) and is elementary if edge-punched cards are used for identification purposes. The presence of environmentally-caused variation within clones for qualitative characteristics is an additional problem in the calculation of the discriminatory value of characteristics (ROGERS & FLEMING, 1973).

Besides their importance in the classification of taxa, qualitative characteristics in the classification might be associated with desirable multi-gene characteristics and such a knowledge can be useful in breeding work (WILLIAMS et al., 1980).

MATERIAL AND METHODS

In this study the same clones are used as described in Engels (1983), but without dividing them into two sets. Table 1 presents the descriptors which were used and their states. For flower characteristics, one well developed flower from each of five trees per clone was chosen at random. For each flower, each organ was allocated a score and the five scores were averaged. The fruit characteristics were observed on the same samples as described in ENGELS (1983), in such a way that if variation was present the most representative state for a descriptor was chosen.

Although all the descriptors dealing with the intensity of pigmentation were registered in an integer scale from 1 (= very slightly pigmented) to 9 (= very intensely pigmented), only two clearly distinguishable classes were used in the final calculations; class one including organs without or with only a slight pigmentation and class two comprising the intensely pigmented ones. For some other characteristics only the absence or presence of pigmentation was registered because of the high variation within clones. Drawings were made of the states of the characteristics 9, 10, 11 and 14, for the permanent record. For a more detailed definition of the descriptors, refer to ENGELS et al. (1980). The basic colour of the fruit and the colour of the new flush were

Table 1. Qualitative cacao descriptors and their states, as used in the final calculations.

Organ	Description of	Descriptor states
Flower		
1. peduncle base	pigmentation	0 = absent, 1 = present
2. sepal	pigmentation	0 = absent, 1 = present
3. petal	intensity of pigmentation	3 = none or slight, 7 = intense
4. filament	intensity of pigmentation	3 = none or slight, 7 = intense
5. superior part ovary	intensity of pigmentation	3 = none or slight, 7 = intense
6. inferior part style	intensity of pigmentation	3 = none or slight, 7 = intense
Fruit		
7. fruit surface	rugosity	1 = very little, 3 = little, 5 = intermediate, 7 = rugose
8. cotyledon	colour	1 = white and purple mixed, 2 = only purple
9. fruit base	base constriction (bottle neck)	1 = not or very slight, 5 = moderate, 7 = pronounced
10. fruit apex	position	1 = straight, 2 = slightly bent, 3 = bent
11. fruit apex	shape	1 = rounded, 5 = intermediate, 7 = pointed
12. fruit surface	basic colour	1 = greyish green, 2 = water green, 3 = lettuce green, 4 = grass green
13. fruit ridges	anthocyanin intensity	0 = absent, 3 = slight, 5 = intermediate, 7 = intense, 9 = very intense
14. secondary fruit furrow (within a pair of ridges)	depth	1 = very superficial, 3 = superficial, 5 = intermediate, 7 = deep
Leaf		
15. flush	colour	1 = yellowish brown, 2 = light brown, 3 = brown, 4 = brownish red, 5 = reddish brown, 6 = garnet brown, 7 = cardinal red

classified with the Methuen colour chart (KORNERUP & WANSCHER, 1978), and later grouped into classes.

The calculation of the discriminative value of a qualitative descriptor is based on the number of pairs of taxa that can be separated (numerator) and the total number of pairs (denominator). For a characteristic with two states the formula will be (PANKHURST, 1978):

$$D = \frac{n_I \times n_{II}}{N(N-1)/2}$$

where D is the discriminative value of the descriptor, n_I is the number of clones showing state I, n_{II} is the number of clones showing state II, and N is the total number of clones. The maximum value of D is attained if $n_I = n_{II} = \frac{1}{2}N$. Then $D_{max} = \frac{1}{2}\left(\frac{N}{N-1}\right)$

For a three states characteristic, the formula will be:

$$D = \frac{(n_I \times n_{II}) + (n_I \times n_{III}) + (n_{II} \times n_{III})}{N(N-1)/2} \quad (1)$$

where n_I, n_{II} and n_{III} represent the number of clones showing the states I, II and III respectively. D will reach its maximum value when $n_I = n_{II} = n_{III} = \frac{N}{3}$, and the formula can be rewritten as $\frac{2N}{3(N-1)}$.

A general formula for a descriptor with k states is:

$$\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j}{\binom{N}{2}} \quad (2)$$

The influence of the number of states per characteristic on the maximum value of D is obvious. D_{max} will increase from 0.50 (for 2 state characteristics) to 0.67 (for 3 states), and to 0.75 (for 4 states), etc.
The formula for

$$D_{max} = \frac{\frac{k(k-1)}{2} (\frac{1}{k} N)^2}{\frac{N(N-1)}{2}} = (1-\frac{1}{k}) (\frac{N}{N-1}) \quad (3)$$

The value of $\frac{N}{N-1}$ can be assumed to be 1 if N is big. This implies that only multistate characteristics with discrete states should be used, which can be recorded without much risk of mis-classification.

Frequently, quantitative descriptors are transformed into qualitative ones, for instance, yield per tree expressed as high, intermediate or low instead of kg/tree. Each time it has to be decided how many states should be distinguished. Therefore, and to demonstrate the effect of the number of states on the discriminative value, relations were examined between the number of states, difference between minimum and maximum values of the observations, standard deviation and the calculated D value for the quantitative characteristic and the D value after transformation into a qualitative variable. The proposed method for calculating the appropriate number of classes for a quantitative observed descriptor is to divide the difference between maximum and minimum observation of that descriptor by the standard deviation.

The discriminatory power D of qualitative descriptors may be enhanced by adjusting D by that part of the information which is shared with other descriptors. For the amount of information shared between two descriptors the correlation coefficient can be used. Since qualitative characteristics frequently do not show a normal distribution, as is the case in the present study, a nonparametric test has to be used for calculation of the association between two characteristics. In this study the rank correlation coefficient of Spearman ('r_s') was chosen, because it is simple to calculate and one can test r_s for significance as an ordinary product moment correlation coefficient if $n > 10$ (SOKAL & ROHLF, 1969).

The first step in the selection of the most powerful discriminator among the descriptors is to look for the characteristic with the highest D value. Then the D values of

the other characteristics are recalculated, to D′, by subtracting that part of their distinctive value which was shared with the largest D, as is indicated by a significant positive or negative correlation coefficient r_s. The formula to be used is the same as the one presented in ENGELS (1983): $D' = D(1-r^2_s)$. These recalculations of D and D′ values are repeated for as many cycles as the total number of descriptors, minus one.

The D′ (or D'_w, as used in ENGELS (1983 for quantitative descriptors), values of qualitative and quantitative characteristics are completely comparable. Both express the proportion of pairs of taxa that can be separated from each other, adjusted for the information that is shared with other descriptors. To allow the inclusion of both types of descriptors in the calculation of the discriminative power, the association between qualitative and quantitative characteristics must be determined. HILL & SMITH (1976) present a method whereby a discrete (qualitative) characteristic and a continuous (quantitative) characteristic are correlated. It is an indirect way of calculating the correlation by approximating the continuous descriptor as closely as possible by an additive effects model based on the observed character states of the discrete descriptor. This can be done by replacing each of the states of the discrete descriptor by the mean value of all the corresponding states of the continuous descriptor for the same group of taxa. In the example below, R stands for red fruits, G for green and Y for yellow ones. The continuous characteristic is pod wall thickness (ridges) in cm.

clone no.	1	2	3	4	5	6	7	8	9
fruit colour	R	R	R	Y	Y	Y	G	G	G
pod wall thickness	1.0	1.0	0.7	0.7	0.7	1.0	0.1	0.1	0.1

Mean value of red fruits expressed in pod wall thickness $(1.0 + 1.0 + 0.7)/3 = 0.9$; mean value Y: $(0.7 + 0.7 + 1.0)/3 = 0.8$, and the green fruits are replaced by $(0.1 + 0.1 + 0.1)/3 = 0.1$. The correlation coefficient is then calculated between:

clone no.	1	2	3	4	5	6	7	8	9
fruit colour	0.9	0.9	0.9	0.8	0.8	0.8	0.1	0.1	0.1
pod wall thickness	1.0	1.0	0.7	0.7	0.7	1.0	0.1	0.1	0.1

In Diagram 1 is summarised what kind of statistical tests are required to enable the use of both types of characteristics in a comparative study of the discriminative values of each of them.

RESULTS AND DISCUSSION

Peduncle colour, pigmentation of inferior part of ovary, pigmentation of superior part of style, and colour of ridges and furrows of ripe pods were disregarded in consequence of perfect association with other more reliable descriptors or of absence of variation between the 32 clones used.

Table 2 presents the calculated discriminatory D value with the given method for the qualitative cacao descriptors, which remain after this first selection.

There is no doubt that the number of classes of all these descriptors has also influenced the values of D. Using the method presented by Hill and Smith a correlation of 0.74** was found between number of classes and D values for each of the descriptors with $P \leq 0.01$.

In Table 3 the results are presented of the transformation of some quantitative char-

Diagram 1. Statistical tests used for a comparative study of qualitative and quantitative characteristics of a group of taxa.

	Qualitative characteristics	Quantitative characteristics
Qualitative characteristics	Rank correlation coefficient of Spearman[1]	Correlation coefficient as calculated by Hill and Smith[2]
Quantitative characteristics	Correlation coefficient as calculated by Hill and Smith[2]	Product moment correlation coefficient[2]

[1] Test for significance of the association between two variables can be done with Olmstead and Tukey's corner test for association (SOKAL & ROHLF, 1969).
[2] A test for significance is the t-test.

acteristics into qualitative ones. The results show that the standard deviation serves as an indicator for the class size. The D value can be calculated by the general formula presented on page 390 and so compared with the original calculated D value as presented in ENGELS (1983), and in Table 3 shown in the third column.

All the new calculated D values are within 9% of the original D's as were obtained with Duncan's New Multiple Range Test.

The influence of the number of classes per descriptor on the D value is clearly shown for the first two descriptors in Table 3.

Spearman's rank correlation coefficient r_s of all the descriptor pairs are presented in Table 4. The interpretation of the individual results will be given in a later paper in this series.

New flush colour is the characteristic with the highest discriminative value (D = 0.83) for the used group of 32 clones. D values for the other descriptors in Table 2 were recalculated to D′ by subtraction of the shared information with new flush colour. The next most potent descriptor is basic fruit colour (D′ ≐ 0.71). The first round D′ values are recalculated by subtraction of the shared information with basic fruit colour. The final D′ values are presented in Table 5.

Theoretically 99.79% of the 496 possible pairs of the 32 clones, can be separated by using the D′ values of the six best descriptors as given in Table 5; only 1.04 (0.21% of 496) clones cannot be distinguished. This value of 99.79% is obtained in the following way. The descriptor 'new flush colour' has the highest D′ value (0.83). This means that 83% of all the possible pairs of clones can be separated. The second highest D value belongs to 'basic fruit colour' which can separate 71% of these clonal pairs not yet separated by the first descriptor, or say 71% of the 17% of the pairs left over (= 12.07%). The third best descriptor will separate 66% of the rest of the pairs not yet separated (100–83 + 12.07). A test of this theory was carried out by using 32 edge-punched cards, each of them containing all the information of one clone. Treating each descriptor state as completely discrete, as is assumed in the calculation of the D value, all the 32 cards (= clones) could be separated by using only the five best descriptors. Another test between the findings with the cards and the calculated discriminative values used four relatively independent characteristics (basic fruit colour, fruit apex position, pigmentation intensity of the fruit ridges and fruit base constriction) was conducted, to separate the 32 clones. The calculations showed that 19 pairs

Table 2. The calculated D values for any of the qualitative characteristics and their respective number of states.

Descriptor	Number of states	D value
1. pigmentation peduncle base	2	0.06
2. pigmentation sepal	2	0.12
3. pigmentation intensity petal	2	0.18
4. pigmentation intensity filament	2	0.18
5. pigmentation intensity superior part ovary	2	0.18
6. pigmentation intensity inf. part style	2	0.18
7. fruit surface rugosity	4	0.56
8. cotyledon colour	2	0.31
9. fruit base construction	3	0.54
10. fruit apex position	3	0.66
11. fruit apex shape	5	0.74
12. basic fruit colour	4	0.72
13. pigmentation intensity fruit ridges	4	0.61
14. depth of secondary fruit furrow	4	0.67
15. new flush colour	7	0.83

Table 3. The D values of transformed quantitative characteristics in qualitative ones related with the number of distinguished states.

(1) Descriptor	(2) The found range was arbitrary divided into the following classes	(3) D values for quantitative characteristics as presented in Engels (1983)	(4) D values based on the number of classes as presented in Col. 2	(5) Difference between minimum & maximum observation	(6) Standard deviation	(7) Column 5 divided by column 6
Style length		0.81		1.4	0.3	5
	9		0.86			
	7		0.82			
	5		0.77			
Seed number/ pod weight		0.72		1.1	0.2	6
	9		0.78			
	7		0.87*	0.5*		
	6		0.70			
	4		0.76*	0.5*		
Seed width		0.78		5.9	1.5	4
	4		0.77			
Seed length		0.79		10.5	2.6	4
	4		0.74			
Wet seed weight		0.67		36.9	9.6	4
	4		0.73			

* These results are higher than expected because of disregarding two extreme observations and therefore a much smaller difference was obtained.

Table 4. Spearman's rank correlation coefficients between descriptor pairs.

Descriptors	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. pigment. peduncle base	–													
2. pigment. sepal	0.79	–												
3. pigment. intens. petal	0.82	0.81	–											
4. pigment. intens. filament	0.82	0.81	0.81	–										
5. pigment. intens. sup part ovary	0.82	0.81	0.81	0.72	–									
6. pigment. intens. inf. part style	0.82	0.81	0.81	0.72	0.91	–								
7. fruit surface rugosity	0.62	0.51	0.41	0.48	0.38	0.47	–							
8. cotyledon colour	0.62	0.72	0.59	0.59	0.59	0.59	0.48	–						
9. fruit base constriction	0.67	0.42	0.40	0.49	0.40	0.40	0.21	0.28	–					
10. fruit apex position	0.46	0.52	0.36	0.33	0.41	0.53	0.40	0.73	–0.10	–				
11. fruit apex shape	0.51	0.59	0.35	0.37	0.46	0.63	0.55	0.66	0.19	0.72	–			
12. basic fruit colour	0.43	0.48	0.52	0.52	0.44	0.35	0.16	0.35	–0.05	–0.09	0.05	–		
13. pigment. intens. fruit ridges	0.51	0.59	0.42	0.46	0.64	0.63	0.24	0.12	–0.04	0.12	0.28	0.34	–	
14. depth of sec. fruit furrow	0.57	0.39	0.38	0.46	0.47	0.47	0.65	0.39	0.32	0.24	0.34	0.10	0.03	–
15. new flush colour	0.39	0.50	0.35	0.20	0.54	0.54	0.26	0.23	0.32	0.22	0.33	0.10	0.29	0.07

$(0.05\%) = 0.34$; $(0.01\%) = 0.44$.

Table 5. The qualitative cacao descriptors arranged by the recalculated discriminatory power (D').

Rang number	Descriptor	D'value
1	new flush colour	0.83
2	basic fruit colour	0.71
3	depth of secondary fruit furrow	0.66
4	fruit apex position	0.58
5	pigmentation intensity of fruit ridges	0.48
6	fruit base constriction	0.42
7	fruit apex shape	0.25
8	fruit surface rugosity	0.16
9	pigmentation intensity petal	0.04
10	cotyledon colour	0.03
11	pigmentation intensity filament	0.007
12	pigmentation intensity sup. part ovary	0.007
13	pigmentation intensity inf. part style	0.005
14	pigmentation sepal	<0.005
15	pigmentation peduncle base	$\ll 0.005$

could not be distinguished when the final calculated D' values from Table 5 were used. With the D values from Table 2 a total of 9 pairs could not be separated and with help of the edge-punched cards 5 pairs were inseparable. The conclusion out of this is that the calculations give comparable results with the edge-punched cards, a known method in plant taxonomy, specially for key construction (PANKHURST, 1978).

With the method of Hill and Smith the correlation coefficients were calculated between the quantitative and qualitative characteristics. Table 6 presents as an example the coefficients between the three best qualitative and the three best quantitative descriptors. To calculate their new discriminative value D' the same formula is used as is presented in Material and Methods for D'. The correlation coefficients between the qualitative descriptors taken from Table 4 and those between quantitative charac-

Table 6. Correlation coefficients between the three best quantitative and qualitative characteristics calculated by the method of 'Hill and Smith' and the D′ values of each of them.

Qualitative descriptors	Quantitative descriptors	Style length	Seed number/ pod weight	Ovary length
	D′ value \ D′$_w$ value[1]	0.81	0.72	0.69
new flush colour	0.83	0.54**	0.28	0.35*
basic fruit colour	0.71	0.34*	0.42*	0.36/
depth of secondary fruit furrow	0.66	0.13	0.37*	0.20

* Significant at the 5% level; ** significant at the 1% level.
[1] The recalculated D$_w$ values as presented in ENGELS (1983) for quantitative characteristics.

Table 7. The quantitative and qualitative descriptors arranged by their new calculated D″ value; their original rank number is also given.

New rank number	Original rank number	Descriptor	D″ value
1	1	new flush colour	0.83
2	4	basic fruit colour	0.70
3	6	depth of sec. fruit furrow	0.65
4	5	ovary length	0.51
5	2	style length	0.49
6	3	seed number/pod weight	0.47

teristics are taken from ENGELS (1983). The final results of the recalculated discriminatory values (D″) are presented in Table 7.

CONCLUSIONS

The proposed method for calculating the discriminative value of qualitative (discrete) characteristics is satisfactory; the results are very similar to those obtained when separating the clones with the help of edge-punched cards. The methodology is compatible with that used for quantitative characteristics. A final discriminative power can be calculated for each of the characteristics, either quantitative or qualitative.

ACKNOWLEDGEMENT

The present work is part of the programme of the Plant Genetic Resources Unit of CATIE (Turrialba, Costa Rica), supported by the German Agency for Technical Cooperation (GTZ) with funds from the Ministry of Economic Cooperation at Bonn, Federal Republic of Germany.

I wish to express my thanks to Mr R. Bonilla and Mr R. Sánchez for collecting

the data and to Mr J. Palmer for his critical reading of the manuscript and the fruitful discussions.

REFERENCES

CHEESMAN, E. E., 1944. Notes on the nomenclature, classification and possible relationship of cacao populations. Trop. Agric. (Trinidad) 21: 144–159.

ENGELS, J. M. M., 1983. A systematic description of cacao clones. I. The discriminative value of quantitative characteristics. Euphytica 32: 377–385.

ENGELS, J. M. M., B. G. D. BARTLEY & G. A. ENRIQUEZ, 1980. Cacao descriptors, their states and *modus operandi*. Turrialba 30: 209–218.

ENGELS, J. M. M., L. T. VAN BEUNINGEN & K. REININK, 1981. A manual system for germplasm management. II. Some useful applications of edge-punched cards in handling data of systematic descriptions. Euphytica 30: 635–642.

ESTABROOK, G. F. & D. J. ROGERS, 1966. A general method of taxonomic description for a computed similarity measure. BioScience 16: 789–793.

HILL, M. O. & A. J. E. SMITH, 1976. Principal component analysis of taxonomic data with multi-state discrete characters. Taxon 25: 249–255.

KORNERUP, A. & J. H. WANSCHER, 1978. Methuen handbook of colour. Eyre Methuen, London. 252 p.

MORSE, L. E., J. A. PETERS & P. B. HAMEL, 1971. A general data format for summarizing taxonomic information. BioScience 21: 174–180.

PANKHURST, R. J., 1978. Biological identification. The principles and practice of identification methods in biology. Edward Arnold (Publishers) Ltd., London. 104 p.

POUND, F. J., 1932. The genetic constitution of the cacao crop. In: Imperial College of Tropical Agriculture, Trinidad. Ann. Rep. Cacao Res. 2: 9–25.

ROGERS, D. J. & H. FLEMING, 1964. A computer program for classifying plants. II. A numerical handling of non-numerical data. BioScience 14: 15–28.

ROGERS, D. J. & H. FLEMING, 1973. A monograph of *Manihot esculenta* with an explanation of the taximetrics methods used. Econ. Bot. 27: 1–113.

SNEATH, P. H. A. & R. R. SOKAL, 1973. Numerical taxonomy. W. H. Freeman and Co., San Francisco. 573 p.

SOKAL, R. R. & F. J. ROHLF, 1969. Biometry. The principles and practice of statistics in biological research. W. H. Freeman and Co., San Francisco. 776 p.

WILLIAMS, W. T., R. L. BURT & G. N. LANCE, 1980. A method for establishing character inter-relations in plant collections and its possible application to plant improvement programmes. Euphytica 29: 625–633.

Euphytica 32 (1983) 719–733

A SYSTEMATIC DESCRIPTION OF CACAO CLONES. III. RELATIONSHIPS BETWEEN CLONES, BETWEEN CHARACTERISTICS AND SOME CONSEQUENCES FOR THE CACAO BREEDING

J. M. M. ENGELS[1]

Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turialba, Costa Rica

Received 10 December 1982

INDEX WORDS

Theobroma cacao, cacao clones, cacao descriptors, numerical taxonomy, multivariate statistical methods, cacao breeding.

SUMMARY

The phenetic relationships of 32 cacao clones were analysed with several multivariate statistical methods and the results were compared with their genetic relationships. Also the relationships between 33 cacao characteristics were studied based on average measurements and observations from each of 294 clones. Some conclusions were drawn with respect to cacao breeding using simple correlations between characteristics.

INTRODUCTION

One of the advantages of a systematic description of a crop is the possibility of analysing relationships between taxonomic units involved. These relationships are based on phenotypic expressions and, therefore, represent phenetic similarities between the taxa and not genetic affinities (SNEATH & SOKAL, 1973). Such phenetic similarities are of interest since the computed discriminative values of characteristics as proposed in ENGELS (1983a, 1983b) will be influenced by the degree of similarity. The same data can also be used to study the relationships between characteristics or descriptors (R-technique, SNEATH & SOKAL, 1973). Such relationships between descriptors can be of great help to the breeders. Finally, the correlation between the phenetic and genetic relationships of the clones is of practical relevance for evolutionary/phylogenetic studies.

Yield components of cacao crops have been correlated in earlier studies (ATANDA, 1972; ESKES et al., 1977; GLENDINNING, 1963; RUINARD, 1961; and others). In the present study the correlation coefficients were based on the means of many characteristics from a large number of clones and multivariate statistical methods were used to analyse the relationships of the cacao descriptors.

[1] Present address: PGRC/E, P.O. Box 30726, Addis Abeba, Ethiopia.

MATERIALS AND METHODS

The 32 clones of the previous studies (ENGELS, 1983a, 1983b) were used to determine the relationships between clones. To enable comparisons of genetic affinities, all the relevant data available for each of the clones were tabulated (Table 1). Information on the taxonomic classification in racial groups is only given for the clones that were classified in the literature. For the division of the 32 clones into two sets, see ENGELS

Table 1. The 32 cacao clones, divided in two Sets, and their corresponding origin and pedigree information.

Clonal name	Taxonomic Unit Number (TUN)	Accession number	Country of origin or selection	Pedigree	Racial group
Set I					
CC-9	1	10132	Costa Rica	Matina o.p.[a]	
CC-41	2	10146	Costa Rica	UF-676 o.p.	Trinitario
CC-46	3	10151	Costa Rica	UF-654 o.p.	Trinitario
CC-74	4	10161	Costa Rica	UF-667 o.p.	Trinitario
CC-121	5	10171	Costa Rica	UF-650 o.p.	Trinitario
CC-143	6	10180	Costa Rica	UF-672 o.p.	
CC-107	7	10169	Costa Rica	UF-221 o.p.	
CC-44	8	10149	Costa Rica	UF-667 o.p.	
CC-38	9	10143	Costa Rica	Matina × UF-676	
CC-139	10	10179	Costa Rica	UF-12 o.p.	Trinitario
CC-137	11	10177	Costa Rica	UF-12 o.p.	Trinitario
UF-701	12	10569	Costa Rica	?c	
Set II					
SGU-82	13	10475	Guatemala	Matina × Criollo	
Diamantes-800	14	10251	Costa Rica	ICS-1 × SCA-6	
P-10	15	10376	Mexico	Criollo hybrid	
UF-4	16	10532	Costa Rica	?	Ven. Red[b]
UF-11	17	10534	Costa Rica	Criollo hybrid	
UF-36	18	10540	Costa Rica	?	
UF-168	19	10543	Costa Rica	Criollo hybrid	
UF-601	20	10550	Costa Rica	?	
UF-654	21	10558	Costa Rica	?	Trinitario
UF-667	22	10560	Costa Rica	?	Trinitario
UF-672	23	10564	Costa Rica	?	
UF-676	24	10565	Costa Rica	?	Trinitario
UF-677	25	10566	Costa Rica	?	Trinitario
EET-80	26	10271	Equador	?	Ven. yellow
UF-715	27	10581	Costa Rica	?	
Laranja	28	10344	Brazil	?	
EET-59	29	10266	Equador	Nacional hybrid	
EET-75	30	10270	Equador	Ven. yellow × Ven. red	
EET-164	31	10277	Equador	Nacional × Ven. yellow	
EEG-25	32	10254	Brazil	?	

[a] Open pollinated.
[b] Venezuelan red.
[c] Pedigree unknown.

(1983a). Complete information on the 294 clones used in the analysis of the relationship between descriptors can be found in ENGELS (1981).

The data on seed setting were collected from the same 32 clones as mentioned before. Since the pollination in the collection was completely natural, self-compatible clones will have been fertilized with own and strange pollen and self-incompatible clones with only strange pollen.

Details about the descriptors, the procedure of the data collection and detailed environment description of the cacao living collection site is given in ENGELS (1983a, 1983b) and ENGELS (1981) respectively. Table 2 lists all the descriptors that have been used in one or more of the analyses. Only the descriptor numbers 1, 3, 4, 8, 20 and 25 need some additional explanation.

Fruit index descriptor (No. 1) of Table 2 is defined as the number of fruits needed to obtain one kg of dry cocoa. The productivity of a clone (No. 3) is expressed as a code for the number of fruits per clone (consisting of 10 trees) that is produced within a certain period. The code is as follows: '1' = very low ($\pm$ 40 fruits in a three year period), '5' = intermediate (40 fruits obtained from 3–4 harvests over one year) and '9' = very high production ($\geqslant$ 40 fruits harvested at one time). The maximum number of seeds per fruit (No. 4) is the highest number recorded from observations on 40 fruits per clone. The ratio of mean number of seeds per fruit and maximum seed number (No. 8) is included because it gives some indication about the efficiency of a clone in producing seeds. The mean ridge pair separation (No. 20) is the degree of separation between a pair of ridges on the fruit. It is expressed as the ratio of the distances measured between ridges within a pair and the distance between two ridges of two adjacent pairs. The hardiness of the mesocarp (No. 25) is expressed in a code, where '3' represents soft mesocarp (easy to cut with a kitchen knife), '5' the interate state and '7' a hard mesocarp (not easy to cut).

The mean and standard deviation of all the states of each characteristic were computed and each state was expressed as the deviation from the mean in standard deviation units (SOKAL, 1961) and, thereafter, the value four was added to each new calculated state. The Pearson product-moment correlation coefficients between clones and between characteristics were then calculated. The factor analyses were conducted with the 79.3 version of the Statistical Analysis System (SAS) computer programme. Full symmetric product-moment correlation coefficient matrices were used to cluster the clones and the characteristics by the weighted variable-group method (SOKAL & SNEATH, 1963). The stepwise multiple linear regressions were computed using a programme developed at CATIE (Costa Rica), version HJP0708/81 (H. J. Palmer, unpubl.). The independent variables were preselected according to expected relationships with the dependent descriptor.

RESULTS AND DISCUSSION

Relationships between clones. The phenetic relationships between the 32 clones are presented in Figure 1. The method used has resulted in some 'reversals'. These reversals have been corrected in the manner suggested by SOKAL & SNEATH (1963). If an arbitrarily chosen correlation coefficient of $\geqslant$ 0.3 is considered as the (lower) limit to form subgroups, a total of 7 clear subgroups can be distinguished. The 12 clones belonging

Table 2. Cacao fruit, seed and flower descriptors on which data were collected for the different analyses to study the relationship between clones and between characteristics.

Descriptor Number	Descriptor name	Multivariate method to analyse relationships				
		Between clones		between characteristics		
		factor analysis 32 clones	cluster analysis (SOKAL & SNEATH, 1963) 32 clones	factor analysis 294 clones	factor analysis 32 clones	cluster analysis (SOKAL & SNEATH, 1963) 32 clones
1	Fruit index	+	+	+	+	−
2	Mean dry seed weight, in g	+	+	+	+	+
3a	Productivity	−	−	+	−	−
4	Maximum number of seeds per fruit	+	+	+	+	+
5	Mean wet seed weight, in g	+	+	+	+	+
6	Ratio of mean dry/mean wet seed weight	+	+	+	+	+
7	Mean seed number per fruit	+	+	+	+	+
8	Ratio mean/maximum seed number	−	−	+	−	−
9	Ratio mean seed number/mean fruit weight	+	+	+	+	−
10	Mean seed length, in mm	+	+	+	+	+
11	Ratio mean seed width/mean seed length	+	+	+	+	+
12	Mean seed width, in mm	+	+	+	+	+
13	Mean seed thicknes, in mm	−	−	+	−	−
14	Mean fruit length, in mm	+	+	+	+	+
15	Mean fruit width, in mm	+	+	+	+	+
16	Ratio mean fruit width/mean length	+	+	+	+	+
17	Mean fruit weight, in g	+	+	+	+	+
18	Mean fruit wall thickness (ridge), in mm	−	−	+	−	−
19	Mean fruit wall thickness (sec. furrow), in mm	−	−	+	−	−
20	Mean ridge pair separation	−	−	+	−	−
21a	Secondary fruit furrow depth	−	−	+	−	−
22a	Fruit surface rugosity	−	−	+	−	−
23a	Fruit apex form	−	−	+	−	−
24a	Basal fruit constriction	−	−	+	−	−
25a	Mesocarp hardiness	−	−	+	−	−
26	Mean style length, in mm	+	+	+	+	+
27	Mean ovary length, in mm	+	+	+	+	+
28	Mean ovary width, in mm	+	+	+	+	+
29	Mean staminode length, in mm	+	+	+	+	+
30	Mean sepal length, in mm	+	+	+	+	+
31	Mean sepal width, in mm	+	+	+	+	+
32	Mean petal length, in mm	+	+	+	+	+
33	Mean ligule width, in mm	+	+	+	+	+
	Total	**22**	**22**	**33**	**22**	**20**

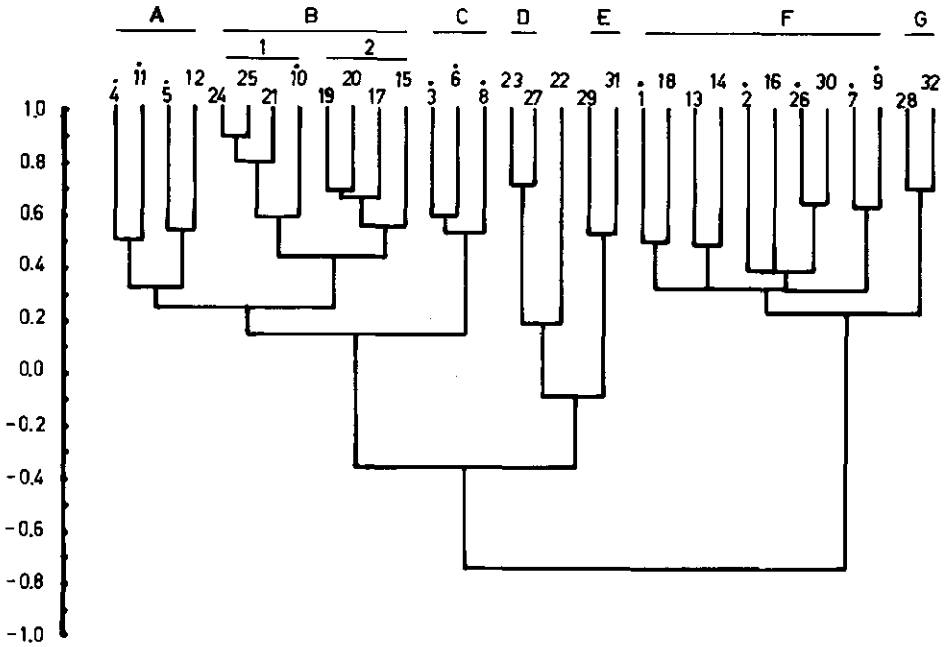


Fig. 1. Dendrogram showing relationships of 32 cacao clones obtained by the weighted variable group method. The clone numbers are given at the top (see Table 1 for identification) and the distinguished clusters are indicated by letters. The clones pertaining to Set I are marked with a dot. The ordinate is shown as the correlation coefficient scale, r.

to Set I are concentrated in or constitute the subgroups A, B1, C and F. This may suggest closer phenetic relationships between these clones than the clones of Set II. To compare the relationships of the clones within each of the two sets in a more quantitative way, all the product-moment correlation coefficients of the clones of each set are added or substracted (if negative). The average of the coefficients of Set I clones was 0.479 and the average calculated per pair of clones of Set I was 0.0435. The respective values for Set II are 0.022 and 0.0012. The percentage of negative correlations in Set I is 16.7% and in Set II 40.0%. These results show clearly the greater phenetic similarity of the Set I clones.

Comparison of the information on genetic relationships between clones (Table 1) and their calculated phenetic similarities expressed in a dendrogram (Fig. 1) allows the following observations. The clones with 'Matina' as female parent (TUN = Taxonomic Unit Numbers 1, 9 and 13 of Table 1) can be found in the subgrop F of the dendrogram (Fig. 1). The 'Criollo' related clones (TUN's 15, 17 and 19) constitute, with clone 20 (pedigree unknown), the subgroup B2. The two clones with 'Nacional' as female parent (TUN 29 and 31) form subgroup E. Furthermore, it can be seen that the two clones from Brazil ('Laranja' and 'EEG-25') constitute a separate subgroup (G), as was expected. The majority of the 'Trinitario' clones can be found in the subgroups A (TUN 4, 5 and 11) and B1 (TUN 10, 21, 24 and 25). Also TUN 3 and 22 are members of the big subgroup ABCDE. Finally, it is observed that all the clones which do not belong to the 'CC-' or 'UF'-group are found on the right

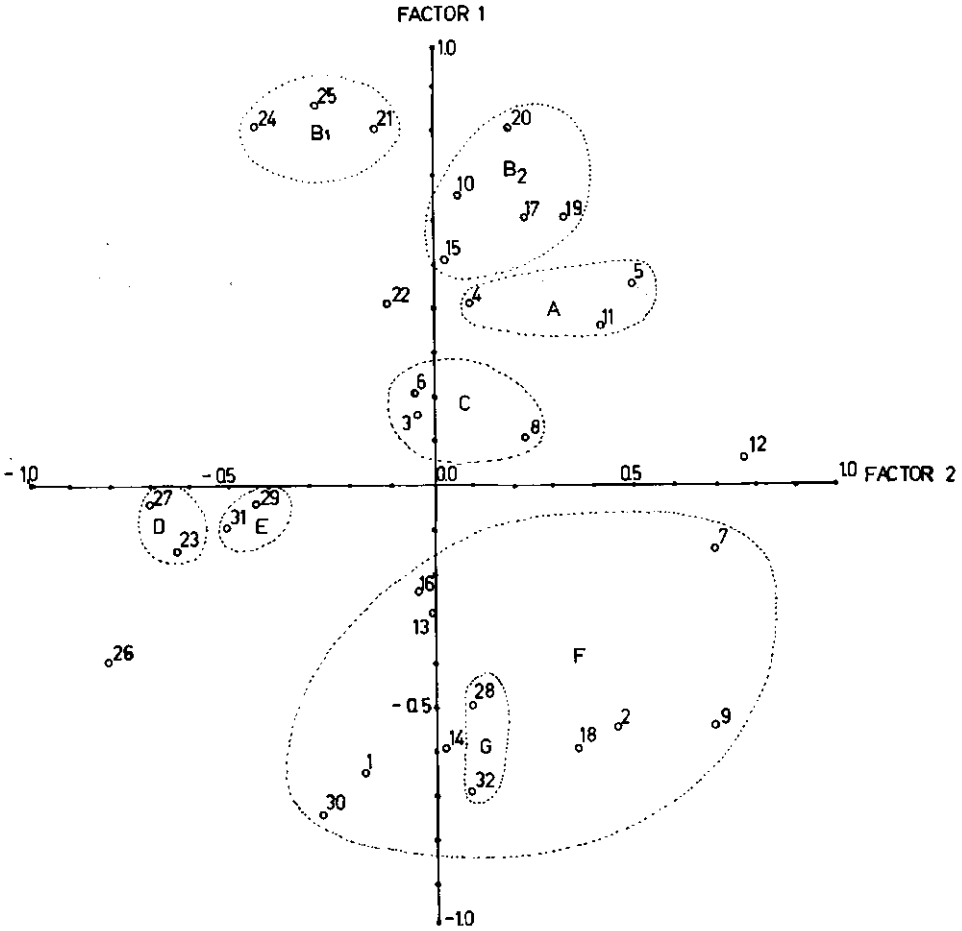


Fig. 2. Graph of the 32 cacao clones on the first two axes as recognized by factor analysis (factor 1 = vertical axis and factor 2 = horizontal axis). The clones are identified as in Table 1 and the subgroups or clusters are marked with the corresponding letters from Figure 1.

hand side of the dendrogram (subgroups E, F and G) with the exception of 'P-10'. It can be concluded that the clones with a known common parent fall into the same group. This points out that the phenetic relationship of the clones represents, to a certain extent, their genetic relationships. This conclusion confirms the expection expressed in ENGELS (1983a) that the clones of Set I are genetically more closely related to each other than the clones of Set II.

By using the first two factors of the factor analysis to plot the clones (Fig. 2), remarkable similarity with the results of the cluster analysis (Fig. 1) can be observed. Only TUN 12, 22 and 26 are less attached to their respective subgroups in the factor analysis, and the subgroup F is even more scattered than it is in the cluster analysis. The latter is in accordance, however, with the diverse origin of the clones.

The additional insertion of eight qualitative characteristics in the analysis of phenetic relationships showed, in an indirect way, the low variation of the 32 clones in these

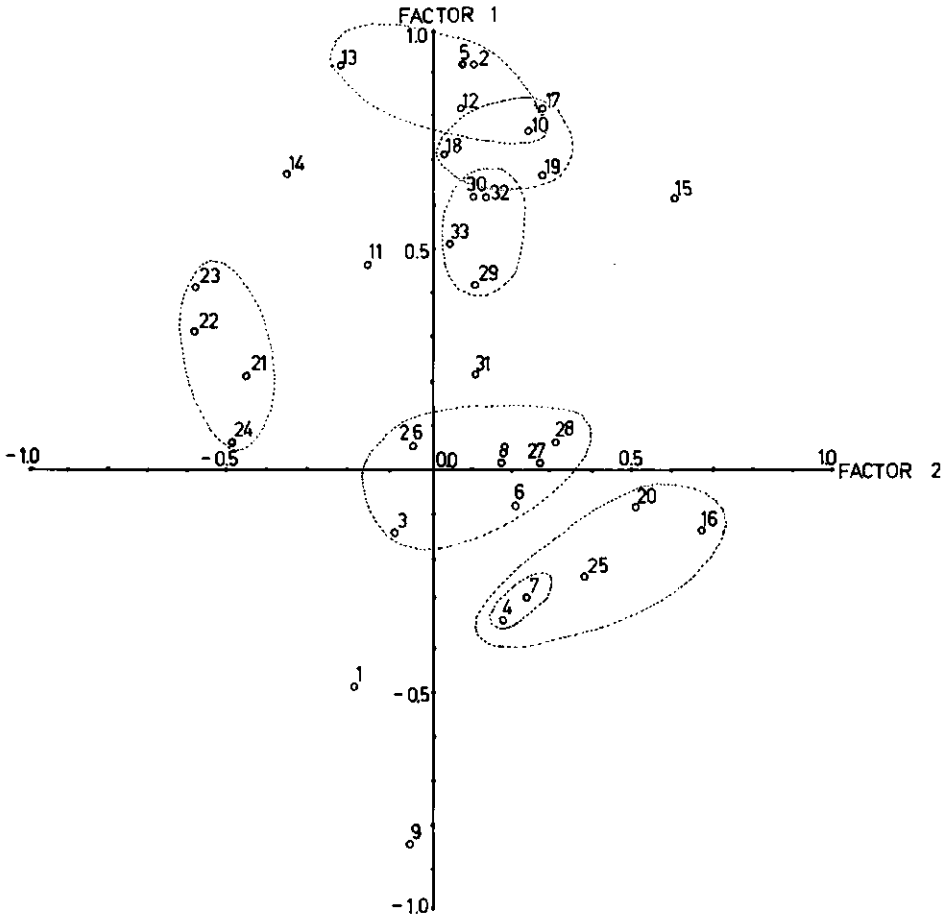


Fig. 3. The 33 cacao characteristics are plotted against the first two axes as recognized by factor analysis of data from 294 clones. Factor 1 (= vertical axis) is composed of seed size and fruit size characteristics and factor 2 (= horizontal axis) comprises mainly fruit shape and fruit appearance descriptors. The descriptors are identified by numbers as in Table 2 and related groups are encircled.

characteristics. In another dendrogram (not included) the phenetic similarity of almost all the clones is very high (>0.80) when compared with the similarity shown in Figure 1 (>0.30). This resulted in many 'reversals' in the dendrogram and biased the final result. Beside the relatively low variation in these qualitative characteristics, two other aspects should be mentioned. In general, the simple inheritance of qualitative characteristics which are recorded in a few classes can cause drastic changes in the phenotypic expression between the parent- and progeny generation. The polygenic inherited quantitative characteristics on the other hand will provide much more 'buffer' and only small(er) changes will be expected. Another aspect of this is that, owing to their relatively fewer classes, there is a greater chance of finding high correlations when qualitative characteristics are used instead of quantitative descriptors. The foregoing indicates the unreliability of using only a relatively small number of qualitative characteristics

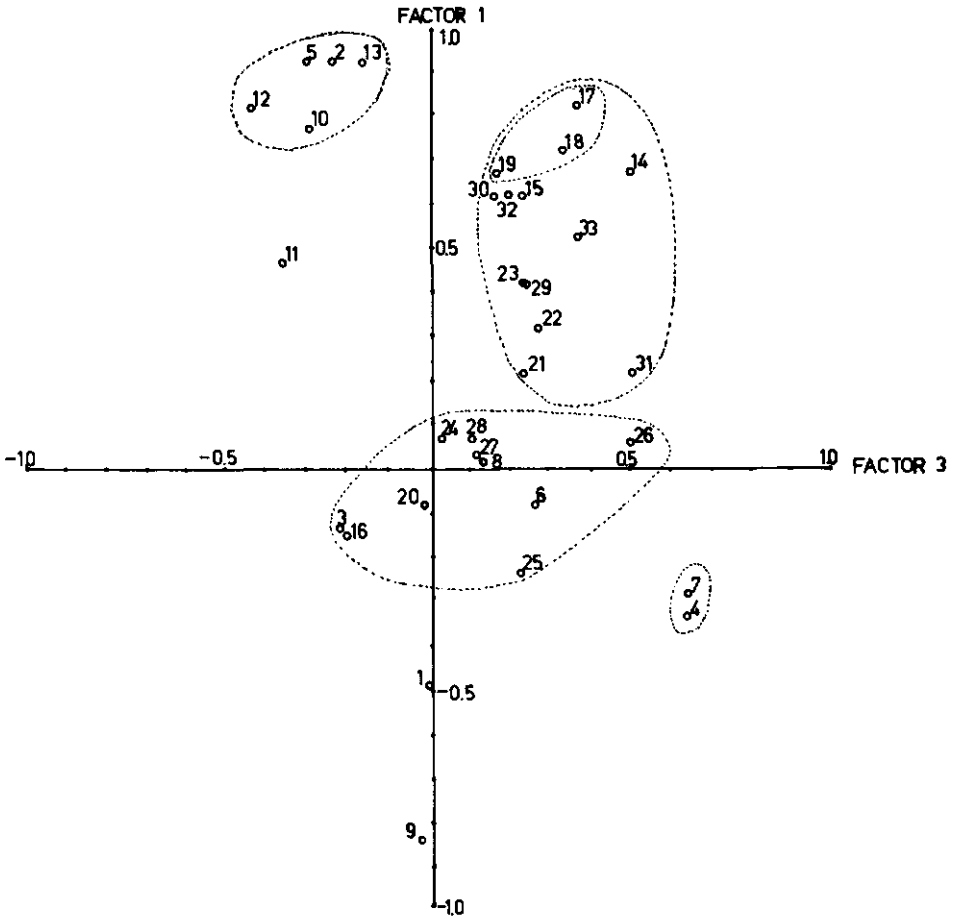


Fig. 4. Graph of 33 cacao characteristics on the factors 1 and 3, as recognized by factor analysis of the data from 294 clones. Factor 1 (= vertical axis) comprises seed and fruit size descriptors and factor 3 is made up of seed number descriptors. The descriptors are identified as in Table 2 and related groups of descriptors are encircled.

in similarity analyses of taxa which are cultivars (or clones) of one species. This would indicate that a much higher number of qualitative descriptors is desirable, as stated by several authors. HEISER et al. (1965) for example used 30 scored descriptors and 28 measured ones; CROVELLO (1969) recommended that as many characteristics as possible should be used. The result of using many qualitative descriptors is a 'buffering' effect toward large changes in quantitative characteristics. However, other authors have used only a small number of characteristics in their multivariate analyses, BURLEY & BURROWS (1972) used 12 needle characteristics to detect successfully differences between provenances of *Pinus kesiya*. GOODMAN (1968) employed 16 characteristics to measure with good results similarity in maize races. From the results obtained in this study it can be concluded that for similarity studies the number of characteristics is less important than the diversity in their respective phenotypic expression.

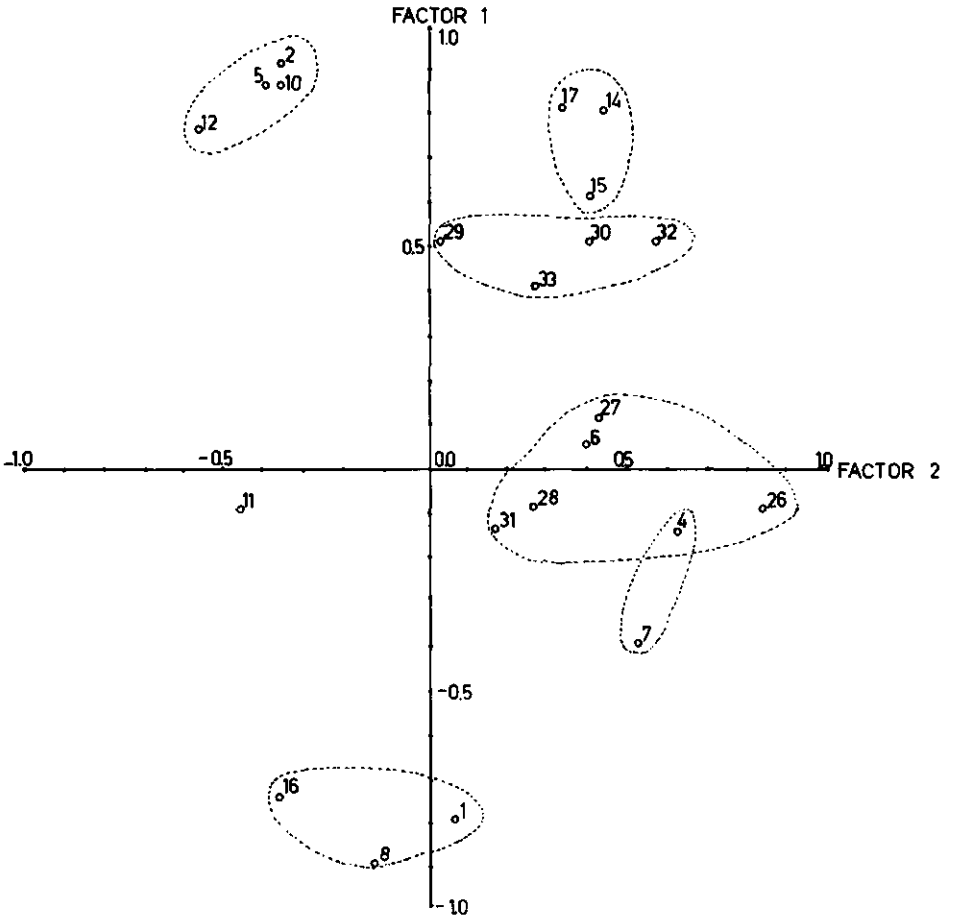


Fig. 5. Graph of 22 cacao descriptors on the first two axes as recognized by factor analysis, using data from 32 clones. Factor 1 (= vertical axis) is composed of seed and fruit size descriptors and factor 2 (= horizontal axis) consists of mainly fruit shape and fruit appearance characteristics. The descriptors are identified as in Table 2 and related descriptors are encircled.

Relationship between characteristics. The relationships between the characteristics (Table 2) are presented in Figure 3 and 4 and are obtained with a factor analysis using 294 clones. In Figure 3 the descriptors are plotted against the factors 1 and 2, and in Figure 4 the factors 1 and 3 are used as the vertical and horizontal axes, respectively. Based on the factor pattern the following factors could be distinguished: factor 1, composed of seed size (descriptor numbers 2, 5, 10, 12 and 13) and fruit size (numbers 14, 15, 17, 18 and 19); factor 2, containing mainly fruit shape and fruit appearance descriptors (16, 20, 21, 23 and 24), factor 3 includes the two seed number descriptors (4 and 7) and factor 4 is made up of flower size characteristics (27, 29, 30 and 32) and the ratio of seed number of maximum seed number (No. 8). Thus, the seed size and pod size descriptors had high final communality estimations; many of the 'ratios' (numbers 6, 8, 11 and 16) had very low communalities, the same as

the descriptors 3 (productivity), 20, 21, 24 and 25 (pod appearance characteristics), some flower (26 and 31), and ovary size descriptors (27 and 28). The group of descriptors with low final communality estimations is therefore relatively independent.

As expected, the seed size and seed weight descriptors are closely related to each other. The fruit weight and wall thickness characteristics are placed near to the seed size 'cluster' (Fig. 3). The latter descriptors are separately plotted in Figure 4, where factor 3 forms the horizontal axis.

The similarity between Figures 4 and 5 is remarkable. In the latter, 22 variables are plotted against the first two factors as were found in the analysis using 32 clones. Since in this factor analysis no qualitative fruit characteristics were used (the numbers 20–25 of Table 2), factor 2 of the first analysis using 294 clones is logically absent and is replaced by factor 3. The similarity between the two graphs indicates that the 32 clones which were used in the study to calculate the discriminative value of the characteristics, were a good representation of the phenetic (and genetic?) diversity of the cacao clones used, assuming the 294 clones represent the diversity of the cacao crop.

In Figure 6, 20 quantitative descriptors are clustered according to the weighted variable-group method of SOKAL & SNEATH (1963). The data are the means of the 32 clones as presented in Table 1. The strong clustering of the seed size and seed weight characteristics is also clearly demonstrated here. The presence of fruit size (descriptors 14, 15 and 17) and flower size descriptors (numbers 29, 30 and 32) in the same group is noticeable. This observation is similar to the positions of the same descriptors in the Figure 3 and 5. Maximum number of seeds and mean seed number per pod are closely related as are ovary length and width. The relationships between all other descriptors in the cluster analysis are comparable to the results obtained in the factor analysis.

The calculated correlation coefficients based on the means of 294 clones are presented in Table 3. These coefficients can be handled as genetic parameters of the relationship between each pair of descriptors. The correlation coefficients within clones represent, on the contrary phenetic parameters. In some earlier studies (ESKES et al., 1977; GHOSH, 1976) this difference between 'between' and 'within' clones was not always taken into consideration and this had led to erroneous conclusions. For example the mean correlation coefficient between seed number per pod and pod length 'within' clones is 0.65**, which gives rise to the conclusion that 'the bigger the pod the more seed'. However, the correlation 'between' clones for both characteristics is 0.06 which shows that there is no relationship between pod length and seed number.

Since dry cocoa yield is the main goal of planting cacao, seed weight, seed size, seed number per fruit, fruit size and weight, and of course, fruit index are relevant parameters of cacoa production. Since the productivity of a clone was not very accurately measured, the figures used are only estimates.

Seed number per fruit does not correlate with fruit size (length, width and weight), which contradicts the results of TOXOPEUS & JACOB (1970), GLENDINNING (1963) and others. Seed size (length, width, thickness and weight) is positively correlated with the fruit size descriptors, which had also been found by RUINARD (1961). Because of the existence of positive correlations between seed number and fruit size 'within'

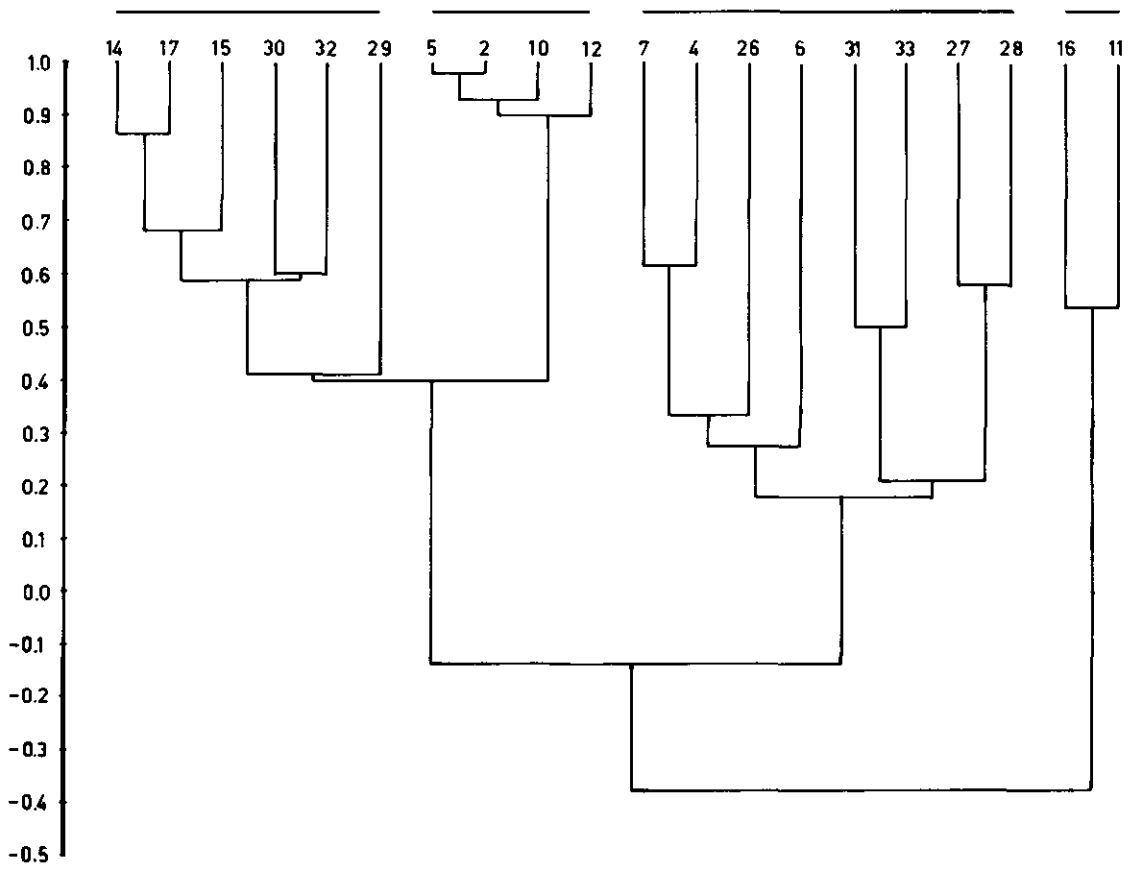


Fig. 6. Dendrogram showing relationships of 20 quantitative cacao descriptors based on the data from 32 clones and obtained by the weighted variable group method. The characteristics are identified as in Table 2 and the distinghuished clusters are marked with a horizontal line in top of the dendrogram.

clones (unpublished data) it can be concluded that seed size determines fruit size genetically. But on the other hand, the negative correlations between seed number and seed size show that breeding for higher dry cocoa production should involve higher seed number per pod and greater seed size. The relative independence of fruit index and seed number on one side, and the negative correlation between fruit index and seed size on the other, support the idea that breeding for seed size leads to better results than breeding only for seed number per fruit. At the same time bigger seeds (>1 gram per seed) have a much better market than smaller ones. KUPPERS (1953) and RUINARD (1961) came to similar conclusions with regard to seed size selection.

The maximum number of seeds per fruit showed a high positive correlation ($r = 0.77^{**}$) with the number of ovules per ovarium (Engels, unpubl.). The latter is a very constant characteristic with low variation within a clone and a high heritability.

An interesting question is how far the shape of the cacao fruit influences dry cocoa yield. The fruit shape characteristics used in this study were the ratio of fruit width to length, ridge pair separation, the secondary furrow depth, the form of the apex

Table 3. Product-moment correlation coefficients between cacao characteristics. The identification is identical to the numbers as used in Table 2. The negative correlations are presented in the top-right section of the table (above the diagonal) and the positive correlations in the bottom-left section. The number of observations is 294 and these are means of 294 clones.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	–	.53	.03		.50	.13	.13	.19		.46	.26	.47	.36	.23	.38	.03	.41	.25	.23	.08	.10	.06	.01	.02		.05	.06	.02	.19	.31	.09	.24	.11
2		–	.06	.35			.34	.03	.70							.05				.02				.00	.20	.08							
3			–	.02	.07		.11	.10		.05		.05	.07	.20	.21	.00	.25	.10	.05	.08		.06	.09	.03	.26		.01	.07	.09	.05	.11	.14	.06
4	.03			–	.41			.15		.32	.32	.41	.42		.00	.01	.03	.12	.20			.01	.13	.08					.10	.12		.19	.02
5		.98			–	.15	.37	.02	.71							.06				.03					.22	.11	.00						
6		.05	.04	.27		–		.04		.09	.12	.13	.12	.02				.04			.03	.06	.13	.03				.03	.10	.08		.13	.02
7				.71		.17	–			.30	.24	.36	.44			.02		.15	.18			.08	.15	.06						.08		.12	.09
8							.47	–		.02		.02	.10					.04			.03	.12	.05	.05	.05	.08	.04						.05
9	.35		.16	.35		.05	.41	.14	–	.62	.29	.60	.74	.59	.69		.83	.71	.67		.10	.21	.31			.09		.06	.20	.42	.17	.43	.42
10		.86			.87					–														.02	.16	.12							
11		.45	.02		.48			.02		.20	–					.09				.13					.20	.12	.04	.08		.09			
12		.88			.90					.82	.67	–								.05					.23	.16				.04			
13		.82			.83					.64	.53	.76	–			.26				.16					.30	.03	.06	.01					
14		.41		.01	.41		.06	.05		.22	.16	.23	.61	–		.48				.14					.12		.08						
15		.53			.50	.10	.03	.07		.51	.09	.42	.34	.28	–						.03	.05	.14	.22									
16						.03		.01	.07	.08		.02			.26	–	.05	.16	.02		.17	.37	.44	.20		.05						.11	.05
17		.60			.58	.04	.04	.09		.50	.20	.46	.58	.69	.81		–							.08									
18		.45			.45					.34	.22	.37	.57	.57	.55		.74	–		.13				.08	.16								
19		.46			.44	.02		.00		.41	.14	.37	.49	.38	.57		.69	.81	–	.01	.10	.04		.19	.11								
20				.09		.06	.09	.05	.04	.03					.15	.19	.02			–	.30	.28	.17	.24		.05				.08	.02		.09
21		.09	.02	.08	.09		.01			.01	.13	.08	.18	.27			.07	.26			–				.21		.02		.04				
22		.16			.16					.04	.12	.09	.37	.50			.19	.25			.52	–			.16		.07	.05					
23		.26			.28					.11	.17	.16	.49	.71			.25	.25	.10		.28	.47	–		.22		.15						
24					.01				.01		.04	.02	.14	.23							.23	.32	.31	–	.09				.03		.07	.04	.14
25	.11			.27		.12	.18		.07						.15	.20	.01			.28					–				.17	.14		.12	.10
26			.06	.28		.20	.16							.22	.15		.16	.22	.07		.18	.08	.08	.11	.01	–	.06						.02
27		.03		.10		.14	.08		.01	.02		.01			.13	.07	.02	.02	.03	.13				.04	.07		–						
28		.01		.01	.01		.01	.00		.03		.01		.04	.10	.75	.11	.03	.03	.05	.04		.05	.03	.08	.05	.05	–					
29		.27			.28		.06	.17		.26	.03	.21	.24	.26	.17		.27	.19	.21	.01		.06	.18			.02	.26	.04	–				
30		.44			.45			.02		.45	.20	.44	.39	.28	.29		.38	.38	.34		.13	.13	.16	.06		.17	.11	.02	.55	–			
31		.06		.13	.05	.05	.13	.02		.00			.07	.23	.18		.24	.19	.13		.09	.11	.16		.05	.11	.21	.14	.25	.19	–		
32		.42			.43			.09		.42	.13	.38	.39	.32	.36		.44	.39	.37	.01	.09	.12	.16			.12	.19	.05	.60	.67	.20	–	
33		.34			.33					.28	.11	.26	.40	.42	.26		.41	.46	.37		.08	.14	.29			.20		.10	.27	.34	.45	.49	–

(0.05%) = 0.11; (0.01%) = 0.15.

Table 4. The regression equations describing the relations between the dependent characteristic and the respective independent characteristics. The latters are identified as X_1, X_2, X_n, were 'n' stays for the descriptor number as given in Table 2, R^2 presents the multiple correlation coefficient (see also text). The means of each of the descriptors of 294 clones are given directly under their identification.

Dependent descriptor	Independent descriptors					R^2
Mean dry seed weight, in g[a]	$1.213 - 0.013X_4 - 8.370\,X_9 + 0.012\,X_{15} + 0.026\,X_{23}$					
mean 1.234	46.577	0.063	86.399	4.598		0.53
Mean fruit weight, in g	$-1119.28 + 2.13\,X_7 + 7.58\,X_{12} + 2.50\,X_{14} + 10.80\,X_{15} + 18.85\,X_{19}$					
mean 570.44	34.14	13.11	174.72	86.40	7.84	0.92
Fruit index	$58.253 - 13.110\,X_5 - 0.547\,X_3 - 0.801\,X_7 + 105.066\,X_9 + 0.044\,X_{14}$					
mean 26.563	1.234	4.794	34.137	0.063	174.720	0.42
Mean sepal length, in mm	$1.906 + 0.349\,X_{26} + 0.291\,X_{29} + 0.492\,X_{32}$					
mean 8.658	2.272	7.486	7.687			0.49

[a] The highly correlated seed size characteristics were not included as independent variables.

and the basal fruit constriction (or bottle neck). The apex form shows significant correlations with dry seed weight ($r = 0.26^{**}$), with seed number ($r = -0.15^{**}$), with seed width ($r = 0.16^{**}$), and with seed thickness ($r = 0.49^{**}$). The seed thickness is negatively correlated ($r = -0.26^{**}$) with fruit width/length ratio. Since the fruit index does not show significant correlations with any of the shape characteristics it can be concluded that dry cocoa production is independent of the shape of the fruit. If slender seeds are wanted the selection should be directed towards slim fruits.

The thickness of the fruit wall (both at ridges and furrows) has a positive influence on the fruit index ($r = -0.25^{**}$ and -0.23^{**} respectively) which can be explained by the positive correlations between wall thickness and seed size and weight, respectively. However, the number and maximum number of seeds are affected negatively by a thicker fruit wall. Since it can be assumed that a thicker fruit wall gives a better protection for the seeds, it seems advantageous to select for cacao fruits with thick walls. The only disadvantage could be slight negative correlation between fruit wall thickness and the hardiness of the mesocarp ($r = -0.16^{**}$ and -0.11^{*} respectively).

There is a significant positive correlation between the size of the flower (petal, sepal and staminode length), the seed size and seed weight (correlation coefficients around 0.40*). Also fruit size and weight show similar correlations with flower size. Although the correlations are not very strong, selection for big flowers at the young developmental stage of the tree could allow for an early selection for big seeds.

For some of the characteristics discussed above, an attempt has been made to determine the qualitative relationships with other descriptors using stepwise multiple linear regression. The results for some dependent descriptors are presented in Tabe 4 in the form of regression equations. The multiple correlation coefficient which is presented as R^2 shows the proportion of the variation in the dependent variable accounted for the variation in the independent variables (LITTLE & HILLS, 1978). The multiple linear regression coefficients are based on data from 294 clones and the means of the characteristics are presented to allow a better interpretation of the relative importance of their contribution to the dependent variable. It can be concluded that these results resemble those from the simple correlations as shown earlier. An exception

is the presence of average seed number (X_7) in the third equation with fruit index as dependent variable. The 'flower size' expressed in sepal length as dependent descriptor includes only flower related characteristics in the last equation of Table 4, contrary to the simple correlations, where seed size and pod size showed positive correlations.

Finally, the influence of self-incompatibility on seed setting is studied. As a parameter, the ratio of average seed number to maximum seed number per pod is chosen and the average ratio for self-compatible (s.c.) and self-incompatible (s.i.) clones is calculated. The non-significant different results (s.c. ratio is 0.74 ± 0.05 and s.i. ratio is 0.73 ± 0.07) leads one to the conclusion that under natural conditions of the cacao collection at CATIE, seed setting is non-significant different for both self-compatible and self-incompatible clones. This indicates that a mixture of s.c. and s.i. trees in a cacao plantation will be of no disadvantage in comparison to a complete self-compatible clone or population.

CONCLUSIONS

In both the analysis of the relationships between clones and between characteristics, the multivariate statistical methods employed gave very similar results.

The genetic relationships between clones reflect their phenetic relationships.

The assumption made in an earlier article in this series that the clones of Set I are more closely related to each other than the clones of Set II has been verified.

The employment of only a few qualitative characteristics in similarity measurements between genetically related taxa of a cultivated species may give inaccurate results.

Seed and fruit size, and weight characteristics formed compact groups of closely related descriptors in all the analyses.

As groups of 32 and 294 clones yielded similar relationships based on 22 characteristics, the 32 clones apparently represented the phenetic diversity of the cacao crop.

Selection for seed size will lead to higher dry cocoa production per fruit than selection solely for seed number per fruit.

Fruit shape does not seem to influence the dry cocoa production per fruit.

A thick fruit wall seems to be advantageous as protection for the seeds and it is positively correlated with seed size and seed weight. On the other hand, there is a slight negative correlation to the number of seeds per pod.

The seed set under the Turrialba cacao collection conditions is not affected by self-incompatibility of a given clone.

ACKNOWLEDGEMENTS

The present work is part of the Plant Genetic Resources Project at CATIE, supported by the German Agency for Technical Cooperation (GTZ) with funds from the German Ministry of Economic Cooperation at Bonn, Federal Republic of Germany.

I wish to express my thanks to Mrs J. Morera, R. Bonilla and R. Sánchez for collecting the data, to Mrs Heather Palmer and Dr J. Henao for the computation of the data and to Dr J. Lazier for his critical reading and correction of the manuscript.

REFERENCES

ATANDA, O. A., 1972. Correlation studies in *Theobroma cacao* L. Turrialba 22: 81–89.

BURLEY, J. & P. M. BURROWS, 1972. Multivariate analysis of variation in needles among provenance of *Pinus kesiya*. Silvae Genetica 21: 69–77.

CROVELLO, T. J., 1969. Effects of change of characters and of number of characters in numerical taxonomy. Amer. Midland Natur. 81: 68–86.

ENGELS, J. M. M., 1981. Genetic Resources of Cacao: a catalogue of the CATIE collection. Technical series. Technical bulletin/CATIE, No. 7, CATIE, Turrialba, Costa Rica.

ENGELS, J. M. M., 1983a. A systematic description of cacao clones I. The discriminative value of quantitative characteristics. Euphytica 32: 377–385.

ENGELS, J. M. M., 1983b. A systematic description of cacao clones. II. The discriminative value of qualitative descriptors and the practical compatibility of the discriminative values of quantitative and qualitative descriptors. Euphytica 32: 387–396.

ESKES, A. B., M. A. BEEK & M. TOXOPEUS, 1977. Variation and correlations of some pod and bean values in cacao (*Theobroma cacao* L.) with reference to the parental effects on fruit setting and bean number. Turrialba 27: 193–198.

GHOSH, B. N., 1976. The shape and size of cacao beans. Turrialba 26: 134–138.

GLENDINNING, D.R., 1963. The inheritance of bean size, pod size and number of beans per pod in cocoa (*Theobroma cacao* L.) with a note on bean shape. Euphytica 12: 311-322.

GOODMAN, M. M., 1968. The races of maize: II. use of multivariate analysis of variance to measure morphological similarity. Crop Sci. 8: 693–698.

HEISER, C. B. JR, J. SORIA, & D. BURTON, 1965. A numerical taxonomic study of Solanum species and hybrids. Amer. Naturalist 99: 471–488.

KUPPERS, J. R., 1953. Some biometric observations of cacao fruits. Science 117: 354–356.

LITTLE, T. M. & F. JACKSON, 1978. Agricultural experimentation. Design and analysis. John Wiley & Sons, Inc., New York.

RUINARD, Y., 1961. Variability of various pod characters as a factor in cacao selection. Euphytica 10: 134–146.

SNEATH, P. H. A., & R. R. SOKAL, 1973. Numerical taxonomy. W. H. Freeman and Co., San Francisco.

SOKAL, R. R., 1961. Distance as a measure of taxonomic similarity. Systematic Zool. 10: 70–79.

SOKAL, R. R. & P. H. A. SNEATH, 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco.

TOXOPEUS, H. & V. J. JACOB, 1970. Studies on pod and bean values of *Theobroma cacao* L. in Nigeria. II. Number of beans per pod, with special reference to the natural pollination process. Neth. J. Agric. Sci. 18: 188–194.

A SYSTEMATIC DESCRIPTION OF CACAO CLONES IV. SOME EVIDENCES OF TETRAPLOID INHERITANCE

J. M. M. ENGELS (*)
Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica

INTRODUCTION

Cacao is generally considered to be a normal diploid species with a basic chromosome number of 10 (e.g. Darlington, 1973; Cope, 1976; Dublin, 1978), although $2n = 16$ and $2n = 26$ have been reported (Zeven and Zhukovsky, 1975).

Studies on the genetics of cacao have shown little progress. Some reasons are the presence of self-incompatibility, the difficulties of artificially pollinating the small flowers under conditions that are frequently adverse and changes in personnel due to the long duration of a crossing programme. The mode of inheritance has been explained (table I) of only a few characteristics (axil spot, cotyledon colour and self-incompatibility). When results did not confirm expected segregation ratios, the presence of modifier genes has been suggested or vague explanations were given (Pound, 1933; Soria, 1978).

It has, however, not been confirmed that cacao is a diploid species and the possibility of it being tetraploid has been suggested (Opeke and Jacob, 1967; G. Enríquez, pers. communication).

Opeke and Jacob found quadrivalents in 36.6 % of the pollen mother cells of the clone « Nanay ». They concluded that this clone is primarily an autotetraploid and that it developed secondary diploidy by a process of continuous outbreeding. Because the chromosomes of cacao show little differentiation (Brücher, 1977) the formation of multivalents could be expected.

TABLE I

Genetic factors reported for qualitative characteristics in cacao

Characteristic	Allele symbol	Author
Axil spot		
- present	AB (complementary)	Harland and Prechville, 1927
- absent	Ab, aB, ab	
Bean or cotyledon colour		
-purple	dominant gene	Wellensiek, 1932 ; Pound 1933 ;
	dominant gene with one or more modifiers	Soria, 1978
Albinism		
- albino plant	recessive gene	Posnette, 1946
Fruit shape		
- bottle neck	recessive	Pound, 1932
- long point	recessive	Pound, 1932
Fruit pigmentation		
- red fruits	series of dominant genes	Pound, 1932
	single gene plus modifiers	Soria, 1978
Fruit wartiness		
- wartiness	dominant	Pound, 1932
Self-incompatibility	$S_1>S_2=S_3>S_4>S_5$	Knight and Rogers, 1955
	$S_a=S_b=S_c>S_d>S_f$ and A and B (complementary, dominant)	Cope, 1962

* Present address : Plant Genetic Resources Centre/Ethiopia. P.O. Box 30726, Addis Abeba, Ethiopia.

MATERIALS AND METHODS

The source of the collected data was a complete diallel of seven clones planted at CATIE, Turrialba (Costa Rica). Table II presents the phenotypic details of these clones. From each cross and selfing six trees were planted in four replicates in 1972. Since in several of the progenies one or more trees died or did not produce any fruit during the period of data collection, the theoretical number of forty-eight or twenty-four observed trees per progeny per cross or selfing, respectively, was never achieved.

Data on « basic fruit colour » (which is defined as the green hue of the exocarp), « ripe external fruit colour », « young flush colour », « cotyledon colour » and « fruit apex form » were collected in the period between April 1980 and May 1981. All characteristics are assumed to be qualitatively inherited. The colours of the fruits and of the young flushes were compared to Methuen Colour Chart (Kornerup and Wanscher, 1978) and were recorded according to the codes of this chart. These codes were then grouped arbitrarily into two or five classes respectively based on the hue, tone and intensity of the coded colour (table III). Some misclassification could not be avoided because of phenotypic variation within a clone or even within an individual caused by environmental variations (mainly position on the tree). The

TABLE II

List of cacao clones and their phenotypes for analysed characteristics.

Clone	Compatibility	Basic fruit colour	Ripe fruit colour	Young flush colour	Cotyledon(*) colour	Fruit apex form
SCA-6	s.i.	green	yellow	red	dark red	intermediate
Catongo	s.c.	light green	yellow	green	white	(very) blunt
P-7	s.i.	(light) green	yellow	red	dark red	blunt-intermediate
CC-42	s.c.	light green	yellow	red	dark red	blunt
UF-29	s.c.	green	yellow	light red	red	blunt
UF-613	s.i.	green	reddish	dark red	dark red	intermediate (-tapering)
UF-676	s.i.	green	yellow	light red	light red	intermediate

s.i. = self-incompatible ; s.c. = self-compatible.
(*) This phenotype is based on the analysis and not on the direct observation.

TABLE III

List of qualitative cacao characteristics, the recorded phenotypic classes and the transformed classes as used in the analysis.

Characteristic	Phenotypic expression recorded as	Transformed classes for analysis	Used code
Basic fruit colour	Methuen colour code	very light green	vlg
		light green	lg
		green	g
		dark green	dg
		very dark green	vdg
Ripe fruit colour	Methuen colour code	yellow	y
		red	r
Young flush colour	Methuen colour code	green	g
		light red	lr
		red	r
		dark red	dr
		very dark red	vdr
Cotyledon colour	white	white	w
	light red	light red	lr
	red	red	r
	dark red	dark red	dr
Fruit apex form	1 = very blunt 3 = blunt 5 = intermediate 7 = tapering 9 = very tapering	same classes from 1 to 9	

lightest cotyledon colour which was found in all the beans of each progeny was recorded, representing the most recessive genotype of that progeny. The apex form was scored visually on a scale from 1 (very blunt) to 9 (very tapering). The observed classes were used in the analysis for both characteristics.

The data were analysed in such a way that conclusions could be drawn on the number of genes involved in the inheritance of each characteristic. The next step was the determination of the individual genotypes of the seven clones for each characteristic. For the study of the ploidy level the recorded data of « basic fruit colour » and « young flush colour » were analysed in detail since they were more suitable for this than the data of the other characteristics. The observations of these two traits were compared with the expected results for diploid and tetraploid inheritance. When classes were difficult to delimit, for instance because of masking effect by pigments in case of « basic fruit colour », they were combined.

Observed and expected data of the individual progenies were not tested with Chi-square since the number of observed individuals per progeny were small, at least in so far as drawing conclusions on tetraploid inheritance. Furthermore, for several of the classes only one individual was expected, which would not have given reliable conclusions based on the Chi-square test. To increase the evidence supporting the logical hypothesis, the expected and observed results were separately summed. This was done for each class included all of the progenies of each female parent. By this means, up to 269 individuals were obtained per female parent (tables VIIa and VIIb, p. 100 and 101) and a Chi-square test was conducted for diploid and tetraploid inheritance. The test was omitted for progenies containing a class in which no individual was expected.

RESULTS AND DISCUSSION

The genotypes of the seven clones for the five characteristics are presented in table IV. The number of dominant alleles is the same for diploid and tetraploid inheritance. The gene « P », responsible for fruit pigmentation is also involved in the pigmentation of the young flushes. Also the gene « A » is a pigmentation gene controlling the pigment formation in « young flush colour » and in « cotyledon colour ».

The results for pigmentation characteristics agree in general with previous reported results (see table I). The data indicated that at least two dominant genes are involved in the pigmentation of unripe fruits and one in the pigmentation of ripe fruits. This confirms Pound's results of 1932. The observations on « fruit apex form », however, do not support Pound's (1932) results for he found that blunt fruit apex is dominant over a tapering apex whereas in the present study, a clear dominance of tapering over blunt apices is found.

Since no significant differences were found between the reciprocal crosses, the data of the corresponding reciprocals were summed for each class. Comparison of the results observed with those expected for « basic fruit colour » and « young flush colour », assuming diploid and tetraploid inheritance respectively, supports the hypothesis of tetraploid inheritance (tables V and VI, p. 98 and 99), particularly in the case of « young flush colour » where green and red flush can be distinguished without any doubt. Tetraploid inheritance also explains better the observed and expected results for both characteristics.

TABLE IV

The genotypes determined for seven cacao clones for five qualitative characteristics. The diploid genotypes are followed by the tetraploid genotypes (bracketed)

Characteristic	Clone						
	SCA-6	Catongo	P-7	CC-42	UF-29	UF-613	UF-676
Basic fruit colour	GG(gg)	Gg(gg)	GG(gg)	Gg(gg)	GG(gg)	GG(gg)	GG(gg)
Ripe fruit colour	pp(pp)	pp(pp)	pp(pp)	pp(pp)	pp(pp)	Pp(pp)	pp(pp)
Young flush colour	AA(aa) pp(pp)	aa(aa) pp(pp)	AA(aa) pp(pp)	AA(aa) pp(pp)	Aa(aa) pp(pp)	AA(aa) Pp(pp)	Aa(aa) pp(pp)
Cotyledon colour	AA(aa)	aa(aa)	AA(aa)	AA(aa)	Aa(aa)	AA(aa)	Aa(aa)
Fruit apex form	TT(tt)	tt(tt)	Tt(tt)	tt(tt)	tt(tt)	TT(tt)	TT(tt)

TABLE V

Observed (o) and expected numbers of trees (e_{II} for diploid and e_{IV} for tetraploid inheritance) for each of the distinguished classes for « basic fruit colour ». The reciprocals are summed and less delimited classes are indicated by an accolade. The meaning of the codes for the classes is given in table III.

		SCA-6			Catongo			P-7			CC-42			UF-29			UF-613			UF-676		
		o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}
SCA-6 GG(gg)	vlg	-	-	-																		
	lg	-	-	-																		
	g	9	15	15																		
	dg	6	-	-																		
	vdg	-	-	-																		
Catongo Gg(gg)	vlg	2	-	3.5	-	2.25	2.25															
	lg	11	21.5	18.0	7	4.5	4.5															
	g	27	21.5	18.0	2	2.25	2.25															
	dg	2	-	3.5	-	-	-															
	vdg	1	-	-	-	-	-															
P-7 GG(gg)	vlg	1	-	1.1	4	-	2.8	1	-	-												
	lg	12	-	9.1	23	17	14.2	9	-	-												
	g	20	41	20.6	7	17	14.2	9	21	21												
	dg	6	-	9.1	-	-	2.8	2	-	-												
	vdg	2	-	1.1	-	-	-	-	-	-												
CC-42 Gg(gg)	vlg	3	-	3.75	3	8.25	8.25	4	-	3.1	2	1.5	1.5									
	lg	7	22.5	18.75	19	16.5	16.5	18	16.5	15.4	3	3.0	3.0									
	g	30	22.5	18.75	11	8.25	8.25	14	16.5	15.4	1	1.5	1.5									
	dg	5	-	3.75	-	-	-	1	-	3.1	-	-	-									
	vdg	-	-	-	-	-	-	-	-	-	-	-	-									
UF-29 GG(gg)	vlg	2	-	1.3	2	-	1.75	1	-	1.2	4	-	2.3	-	-	0.4						
	lg	6	-	10.2	13	16.5	13.75	21	-	9.3	16	14	11.5	9	-	2.8						
	g	30	46	23.0	14	16.5	13.75	16	42	21.0	4	14	11.5	4	13	6.6						
	dg	6	-	10.2	4	-	2.75	3	-	9.3	4	-	2.3	-	-	2.8						
	vdg	2	-	1.3	-	-	-	1	-	1.2	-	-	-	-	-	0.4						
UF-613 GG(gg)	vlg	-	-	1.1	4	-	3.0	-	-	1.1	4	-	2.5	-	-	1.1	-	-	-			
	lg	8	-	8.5	5	17	14.0	11	-	8.4	7	15	12.5	20	-	8.8	4	-	-			
	g	21	38	18.8	12	17	14.0	20	38	19.0	7	15	12.5	11	40	20.2	14	23	23			
	dg	9	-	8.5	13	-	3.0	7	-	8.4	11	-	2.5	9	-	8.8	5	-	-			
	vdg	-	-	1.1	-	-	-	-	-	1.1	1	-	-	-	-	1.1	-	-	-			
UF-676 GG(gg)	vlg	-	-	0.8	1	-	1.5	1	-	1.1	4	-	1.75	3	-	0.9	-	-	0.5	-	-	-
	lg	3	-	7.2	6	9	7.5	7	-	9.1	8	10.5	8.75	7	-	7.3	5	-	4.0	2	-	-
	g	19	32	16.0	7	9	7.5	24	41	20.6	6	10.5	8.75	16	33	16.6	9	18	9.0	6	9	9
	dg	10	-	7.2	4	-	1.5	7	-	9.1	3	-	1.75	7	-	7.3	4	-	4.0	1	-	-
	vdg	-	-	0.8	-	-	-	2	-	1.1	-	-	-	-	-	0.9	-	-	0.5	-	-	-

The hypothesis of tetraploid inheritance is further strengthened by the results of the « pooled » Chi-square tests (tables VII*a* and *b*). Although the hypothesis could not be tested for all the progenies (because of observed but not expected individuals in several classes) tetraploid inheritance explained the observed results better than did diploid inheritance even after the various interaction types between alleles were considered. In case of ripe fruit colour the expected results for diploid and tetraploid inheritance are the same since one dominant allele was found for only one clone (UF-613).

The data for « cotyledon colour » and « fruit apex form » are not presented here due to lack of space. For « cotyledon colour » the genotypes of the parents could be deducted only in an indirect way because the colour of the cotyledons already represents the genetic constitution of the F_1-generation (xenia). Since the male parents were unknown in all cases no segregation ratios nor expected values could be calculated. However, all the observations made could be explained by the genotypes (table IV) if tetraploid inheritance was assumed. This particularly holds true for the presence of white cotyledons in the progenies of several crosses (e.g. CC-42 × UF-613 and CC-42 selfed).

Since the pigmentation of young flushes and of cotyledons show exactly the same inheritance within the crosses, it is concluded that the same gene « A » is responsible for the formation of pigments in both organs.

Due to large variation within clones for « fruit apex form » and the difficulties encountered in the visual recording the data was imprecise and thus conclusions on inheritance could not be drawn. It was clear, however, that tapering fruit apex (T.) was dominant over blunt fruit apex (tt).

TABLE VI

Observed (o) and expected numbers of trees (e_{II} for diploid and e_{IV} for tetraploid inheritance) for each of the distinguished classes for « young flush colour ». The reciprocals are summed and less delimited classes are indicated by an accolade. The meaning of the codes for the classes is given in Table III.

		SCA-6			Catongo			P-7			CC-42			UF-29			UF-613			UF-676		
		o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}	o	e_{II}	e_{IV}
	g	-	-	-																		
SCA-6	lr	5	-	-																		
AA(aa)	r	12	17	17																		
pp(pp)	dr	-	-	-																		
	vdr	-	-	-																		
	g	9	-	7	17	17	17															
Catongo	lr	31	42	28	-	-	-															
aa(aa)	r	2	-	7	-	-	-															
pp(pp)	dr	-	-	-	-	-	-															
	vdr	-	-	-	-	-	-															
	g	3	-	1.3	6	-	6.0	-	-	-												
P-7	lr	19	-	10.2	27	36	24.0	6	-	-												
AA(aa)	r	18	46	23.0	3	-	6.0	14	21	21												
pp(pp)	dr	6	-	10.2	-	-	-	1	-	-												
	vdr	-	-	1.3	-	-	-	-	-	-												
	g	2	-	1.3	3	-	6.5	1	-	1.2	-	-	0.3									
CC-42	lr	18	-	10.7	23	39	26.0	13	-	9.6	2	-	2.4									
AA(aa)	r	24	48	24.0	13	-	6.5	25	43	21.5	5	11	5.6									
pp(pp)	dr	4	-	10.7	-	-	-	4	-	9.6	3	-	2.4									
	vdr	-	-	1.3	-	-	-	-	-	1.2	1	-	0.3									
	g	2	-	3.7	5	18.5	18.5	6	-	3.5	1	-	2.8	3	4.7	4.7						
UF-29	lr	20	22	18.3	29	18.5	18.5	22	21	18.5	15	16.5	13.7	14	9.6	9.6						
Aa(aa)	r	18	22	18.3	3	-	-	12	21	18.5	14	16.5	13.7	2	4.7	4.7						
pp(pp)	dr	4	-	3.7	-	-	-	2	-	3.5	3	-	2.8	-	-	-						
	vdr	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-						
	g	1	-	1	3	-	3.1	1	-	0.5	1	-	0.5	1	-	1.0	-	-	-			
UF-613	lr	6	-	5	12	18.5	15.4	6	-	4.5	4	-	4.9	10	11.5	10.5	1	-	-			
AA(aa)	r	18	23	17	15	18.5	15.4	14	18.5	13.5	13	19.5	14.1	17	23.0	19.0	9	-	-			
Pp(pp)	dr	18	23	17	7	-	3.1	13	18.5	13.5	13	19.5	14.1	14	11.5	9.6	10	24	24			
	vdr	3	-	6	-	-	-	3	-	5.0	8	-	5.4	4	-	1.9	4	-	-			
	g	3	-	3.0	17	11.5	11.5	4	-	3.3	-	-	2.7	5	9.25	9.3	1	2	1.4	-	-	-
UF-676	lr	18	20.5	15.0	3	11.5	11.5	18	19.5	16.2	7	16	13.3	22	18.5	18.4	8	8.5	8.5	3	12	12
Aa(aa)	r	13	20.5	15.0	3	-	-	15	19.5	16.2	20	16	13.3	10	9.25	9.3	12	17.0	14.2	9	-	-
pp(pp)	dr	2	-	3.0	-	-	-	2	-	3.3	5	-	2.7	-	-	-	11	8.5	7.1	-	-	-
	vdr	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	2.8	-	-	-

CONCLUSIONS

It is very likely that the inheritance of the five qualitative characteristics investigated in the seven cacao clones follows tetraploid segregation. This confirms the findings of Opeke and Jacob (1967) and agrees with the observations of Brücher (1977).

In order to describe inheritance in cacao more accurately, it would be necessary to conduct comprehensive and detailed cytogenetic studies, particularly on the genomic constitution of cacao and its allies. A wider range of selected clones with larger progenies are needed to allow reliable statistical treatment.

The author is indebted to CATIE, especially to Dr J. Soria and Dr G. Enríquez for being given the opportunity to use the existing cacao diallel for the data collection. He wishes to express his thanks to J. Morera, R. Bonilla and R. Sánchez for assistance in the data collection and to Drs J. Lazier and C. Seegeler for their critical reading and correction of the manuscript.

TABLE VIIa

Pooled Chi-square values considering all the crosses with each clone for two characteristics. For each class the observed (o) numbers are summed and the expected proportions for each progeny are summed per clone and converted in expected numbers assuming diploid (e_{II}) inheritance.

Characteristic		Clone													
		SCA-6		Catongo		P-7		CC-42		UF-29		UF-613		UF-676	
	Proposed genotype	GG		Gg		GG		Gg		GG		GG		GG	
	Classes	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}
Basic fruit colour	vlg	8	0	12	12.1	12	0	20	12.1	12	0	8	0	9	0
	lg(1)	45	36.9	79	85.0	101	36.3	71	85.0	92	33.6	60	31.6	38	24.6
	g	203	221.1	79	72.9	141	217.7	79	72.9	131	201.4	153	189.4	125	147.4
χ^2		χ^2_2= np		χ^2_2= 0.94		χ^2_2= np		χ^2_2= 7.97		χ^2_2= np		χ^2_2= np		χ^2_2= np	
	Proposed genotype	AApp		aapp		AApp		AApp		Aapp		AAPp		Aapp	
		o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}	o	e_{II}
Young flush colour	g	20	0	55	55.4	21	0	8	0	18	36.9	8	0	30	22.8
	lr	107	76.9	96	124.7	111	75.4	82	70.0	103	119.8	47	37.6	79	114.1
	r	105	172.9	43	13.9	101	169.7	114	157.5	73	92.1	98	112.7	82	68.5
	dr	37	19.2	0	0	31	18.9	41	17.5	27	9.2	110	112.7	22	7.6
χ^2		χ^2_3= np		χ^2_2= 67.53		χ^2_3= np		χ^2_3= np		χ^2_3= 50.78		χ^2_3= np		χ^2_3= 43.02	

P=0.05, χ^2_2= 5.99 ; χ^2_3= 7.82

np = no calculation of Chi-square was possible since in one of the classes no observation was expected.
(1) All the recorded "dg" and "vdg" fruits were added to the "g" class, since they are not expected in case of diploid inheritance.

REFERENCES

BRUECHER (H.). — Tropische Nutzpflanzen. Ursprung, Evolution und Domestikation. Springer-Verlag (Berlin), 1977.

COPE (F. W.). — The mechanism of pollen incompatibility in *Theobroma cacao* L. *Heredity* (London), vol. 17, n° 2, May 1962, p. 157-182.

COPE (F. W.). — Cacao. *In* : Simmonds, N. W. (Ed.), Evolution of crop plants. Longman (London), 1976, p. 285-289.

DARLINGTON (C. D.). — Chromosome botany and the origins of cultivated plants. George Allen & Unwin Ltd. (London), 3rd ed., 1973.

DUBLIN (P.). — Haploïdes diploïdisés et obtention de génotypes homozygotes fertiles chez les cacaoyers cultivés (*Theobroma cacao*). *Café Cacao Thé* (Paris), vol. 22, n° 4, oct.-déc. 1978, p. 275-284.

HARLAND (S. C.), FRECHVILLE (G. E.). — Natural crossing and the genetics of axil spot in cacao. *Genetica* (The Hague), 1927, vol. 9, p. 279-289.

KNIGHT (R.), ROGERS (H. H.). — Incompatibility in *Theobroma cacao* L. *Heredity* (London), vol. 9, n° 1, April 1955, p. 69-77.

KORNERUP (A.), WANSCHER (J. H.). — Methuen handbook of colour. Eyre Methuen, London, 1978.

OPEKE (L. K.), JACOB (V. J.). — Cytological irregularities in *Theobroma cacao* L. *In* : Proceedings of Second International Cocoa Research Conference, Salvador and Itabuna (Bahia), Brazil, 19th-26th Nov. 1967, p. 114-116.

POSNETTE (A. F.). — Pollination experiments. WACRI Annual Rep. 1946, p. 57-58.

POUND (F. J.). — The genetic constitution of the cacao crop. *In* : Imperial College of Tropical Agriculture (Trinidad), Annual Report on Cacao Research, 1932, 2, p. 9-25.

POUND (J. F.). — A preliminary survey of pigment factors in cacao. Annual Report on Cacao Research (Trinidad), 1933, 3, p. 11-15.

SORIA (V. J.). — The breeding of cacao (*Theobroma cacao* L.). *In* : Tropical Agriculture Research Series (Yatabe, Japan), n° 11, March 1978, p. 161-168.

WELLENSIEK (S. J.). — Bloembiologische Waarnemingen aan cacao. *Arch. Koffiecult.*, Ned. Indie, 1932, 6, p. 87-101.

ZEVEN (A. C.), ZHUKOVSKY (P. M.). — Dictionary of cultivated plants and their centres of diversity. Centre for Agricultural Publishing and Documentation (Wageningen), 1975.

ENGELS (J. M. M.). — **Systematische Beschreibung von Kakaoklonen. IV. Einige Nachweise der vererblichen Übertragung der tetraploiden Eigenschaft.** *Café Cacao Thé* (Paris), vol. XXVIII, n° 2, avril-juin 1984, p. 95-102, 8 tabl., 15 réf.

Die Gegebenheiten betreffend fünf charakteristischen Eigenschaften wurden zusammengestellt an Hand von Abkömmlingen einer vollständigen Diallele von sieben Kakaoklonen. Diese Gegebenheiten wurden analysiert und es wurden daraus Schlüsse betreffend der Vererbung dieser Charakteristika gezogen. Die Ergebnisse erlauben daran zu denken, daß sich der Kakaobaum wie eine tetraploide Pflanze verhält und dies könnte eine Erklärung der Probleme sein, welche vorher beim Studium der Genetik des Kakaobaumes auftraten.

ENGELS (J. M. M.). — **Descripción sistemática de los clones del cacao. IV. Algunas evidencias relativas a la transmisión hereditaria del carácter tetraploide.** *Café Cacao Thé* (Paris), vol. XXVIII, n° 2, avril-juin 1984, p. 95-102, 8 tabl., 15 réf.

Los datos precisados se refieren a cinco caracteristicas cualitativas y se han reunido tomando como punto de partida descendencias de un dialelo completo de siete clones de cacao. Se han analizado estos datos y se han sacado las debidas conclusiones acerca de la transmisión hereditaria de estas características. Los resultados permiten pensar que el comportamiento del cacao es semejante al de una planta tetraploide y ello podría constituir una explicación para los problemas con que se ha tropezado en los estudios genéticos del cacao.

TABLE VIIb

Pooled Chi-square values considering all the crosses with each clone for two characteristics. For each class the observed (o) numbers are summed and the expected proportions for each progeny are summed per clone and converted in expected numbers assuming tetraploid (e_{IV}) inheritance.

Characteristic	Classes	Clone											
		SCA-6		Catongo		P-7		CC-42		UF-29		UF-613		UF-676	
	Proposed genotype	GGgg		Gggg		GGgg		Gggg		GGgg		GGgg		GGgg	
		o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}
Basic fruit colour	vlg	8	10.2 (3)	12 (1)	23.6	12	10.1 (3)	20 (2)	23.6	12	10.3 (3)	8	8.8	9	6.8
	lg	45	63.5	79	75.6	101	62.5	71	75.6	92	65.3	60	54.4 (4)	38	42.3
	g	156	141.2	68	61.4 (3)	110	139.1	66	61.4	95	111.9	94	121.0	87	94.2
	dg	44	38.9	10	9.4	26	38.3	13	9.4	33	42.9	58	33.3	36	25.9
	vdg	5	4.2	1	-	5	4.0	-	-	3	4.6	1	3.5	2	2.8
χ^2		χ^2_3= 1.55		χ^2_3= 6.64 (1)		χ^2_3= 5.00		χ^2_3= 2.55 (2)		χ^2_3= 3.66		χ^2_3= 2.47		χ^2_4= 5.87	
	Proposed genotype	pppp		pppp		pppp		pppp		pppp		Pppp		pppp	
		o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}
Ripe fruit colour	y	223	227.5	187	192.2	230	233.1	163	173.6	191	195.0	72	84.6	139	144.9
	r	22	17.5	20	14.8	21	17.9	24	13.4	19	15.0	126	113.4	17	11.1
χ^2_1		1.25		1.97		0.58		9.03		1.15		3.28		3.38	
	Proposed genotype	AAaapppp		aaaapppp		AAaapppp		AAaapppp		Aaaapppp		AAaaPppp		Aaaapppp	
		o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}	o	e_{IV}
Young flush colour	g	20	15.5 (3)	55	67.4	21	15.2 (3)	8	15.1	18	29.2	8	7.8	30	31.7 (3)
	lr	107	79.5	96	94.3 (3)	111	78.0	82	80.2	103	92.1	47	48.5	79	106.5
	r	105	129.2	36	29.6	101	126.8	114	100.1	73	79.8	98	87.7	82	58.3
	dr	34	37.4	7	2.7	28	36.7	32	41.8	23	16.9	86	97.1	20	13.9
	vdr	3	7.4	-	-	3	7.3	9	7.8	4	3.0	24	21.9	2	2.6
χ^2		χ^2_3= 4.28		χ^2_2= 5.86		χ^2_3= 7.06		χ^2_4= 7.79		χ^2_4= 8.70		χ^2_4= 2.73		χ^2_3= 2.99	

P=0.05, χ^2_1= 3.84 ; χ^2_2= 5.99 ; χ^2_3= 7.82 ; χ^2_4= 9.49

(1) The presented data do not include the results of the cross UF-613 x Catongo because of a very likely wrong reading caused by the masking effect of pigments.
(2) The presented data for CC-42 (basic fruit colour) do not include the results of the cross CC-42 x UF-613 (see also text for same reason as under 1).
(3) These classes are combined because of the difficulties encountered during the data recording to distinguish both classes properly.
(4) These two classes are combined since the intensive pigmentation of the pods has caused very likely an overestimation of the class dg.

ENGELS (J. M. M.). — **Une description systématique des clones de cacaoyer. IV. Quelques mises en évidence de la transmission héréditaire du caractère tétraploïde.** *Café Cacao Thé* (Paris), vol. XXVIII, n° 2, avril-juin 1984, p. 95-102, 8 tabl., 15 réf.

Les données concernant cinq caractéristiques qualitatives ont été rassemblées à partir de descendances d'un diallèle complet de sept clones de cacaoyer. Ces données ont été analysées et des conclusions en ont été tirées sur la transmission héréditaire de ces caractéristiques. Les résultats permettent de penser que le cacaoyer se comporte comme une plante tétraploïde et ceci pourrait être une explication aux problèmes antérieurement rencontrés dans les études génétiques du cacaoyer.

ENGELS (J. M. M.). — **A systematic description of cacao clones. IV. Some evidences of tetraploid inheritance.** *Café Cacao Thé* (Paris), vol. XXVIII, n° 2, avril-juin 1984, p. 95-102, 8 tabl., 15 réf.

The data of five qualitative characteristics were collected from the progenies of a complete diallel of seven cacao clones. They were analysed and conclusions on the inheritance of these characteristics are presented. The results suggest that cacao behaves like a tetraploid crop and this could be an explanation for the problems encountered in genetic studies of cacao in the past.

A SYSTEMATIC DESCRIPTION OF CACAO CLONES V. QUANTITATIVE GENETIC ASPECTS OF SEVERAL FRUIT CHARACTERS

J. M. M. ENGELS *

Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica

INTRODUCTION

The determination of the most adequate procedure in plant breeding, aiming to improve a given character can be significantly facilitated by information about the quantitative inheritance of the character. Various crossing programmes exist to obtain such information. The diallel cross e.g. allows a comprehensive biometrical genetic analysis to be carried out after one generation. However, the interpretation of the performance of the progenies of the diallel cross requires several assumptions which are difficult to verify. Since the genetics of cacao (*Theobroma cacao* L.) has received little research attention no concrete information exists about such aspects as ploidy level (Engels, 1984), multiple allelism and gene distribution over parents. Furthermore, cacao clones can be expected to be heterozygous for many characters (Engels, 1984). Thus it was decided not to use the comprehensive analysis (e.g. Wr/Vr analysis **) as proposed for example by Hayman (1954), but to restrict the analysis to estimating general (g.c.a.) and specific (s.c.a.) combining ability mean squares and effects, as proposed by Baker (1978) and Mayo (1980).

For a manageable interpretation of the statistical parameters in quantitative genetic terms the following assumptions for the characters under consideration have to be made (Bos, 1981; and Bos I., pers. comm.) :

(i) epistatic interactions are absent;
(ii) the parental population is in linkage equilibrium;
(iii) the HS progenies ***, and thus their parents, form a random sample;
(iv) extra chromosomal contributions to the phenotypic values are absent.

MATERIALS AND METHODS

Seven cacao clones representing a considerable part of the diversity in American cacao varieties (SCA-6, Catongo, P-7, CC-42, UF-29, UF-613 and UF-676) were selected as parents for a diallel cross to determine the genetics of disease resistance (e.g. *Phytophthora* and *Ceratocystis*) and of yield components, for which mainly fruit characters were observed. Since the reaction to the diseases was the primary selection criterium (Anonymous, 1972) it is assumed that regarding the fruit characters the seven clones are a random sample out of the Turrialba cacao collection which consists out of about three hundred clones. Detailed information on qualitative characters has been

* Present address: Plant Genetic Resources Centre, P.O. Box 30726, Addis Abeba, Ethiopia.

Note of the editor.

** Parent r and progenies rs covariance analysis (Wr) in function of the variance of progenies with the common parent r (Vr) (proposed by Hayman).

*** HS progenies : Half-Sib progenies.

given by Engels (1984). In each of the two environments (i = 1, 2), (Turrialba and La Lola, Costa Rica with altitudes of 623 and 39 m above sea level, 2 640 and 3 574 mm total annual rainfall and an annual mean temperature of 22.3 and 24.6 °C, respectively) the forty two progenies (j = 1, ..., 42) were randomly planted in four blocks in August and September 1972.

Progenies were planted six trees (k = 1, ..., 6) per plot spaced at 3 × 3 m under shade. However, in Turrialba two progenies had five trees and in La Lola thirty eight progenies had only four or five trees. As plant survival was such that only two blocks were nearly complete in each environment it was decided that only these would be analysed.

Up to forty ripe fruits (l = 1, ..., x; x_{max} = 40), with an overall mean of 12.4 fruits per tree, were harvested from each tree and observed. If no fruits were harvested from a tree the observations on the tree in the corresponding plot in one of the otherwise unused blocks within the same environment were taken. Thus blocks I and II were complemented with data from blocks III and IV, respectively. Since the non-producing trees belonged without exception to non-vigorous progenies it is assumed that variation in strength of competition between trees can be neglected. Because of self-incompatibility of some of the clones (Engels, 1984) I_1 progenies could not be taken into consideration in the analysis. Therefore, Griffing's (1956) experimental method 3 is applicable.

Because progenies were randomized across plots, the plot was the experimental unit and for each character a mean was obtained for each plant and these in turn were averaged across each plot $\left(y_{ij} = \frac{1}{6}\sum_{k=1}^{6} y_{ijk}\right)$. When appropriate a weighted plot mean was calculated for plots having one or more trees with only a few fruits.

The eight studied characters were obtained as follows :

1) maximum seed number per fruit

$$m_{ij} = \frac{1}{6}\sum_{k=1}^{6} m_{ijk}$$

where m_{ijk} is the maximum seed number observed in the x fruits obtained from tree k within progeny j in environment i.

2) number of seeds per fruit

$$n_{ij} = \frac{1}{6}\sum_{k=1}^{6}\frac{1}{x}\sum_{l=1}^{x} n_{ijkl}$$

where n_{ijkl} is the number of seeds in fruit l harvested on tree k of progeny j in environment i; n_{ij} is the weighted mean of the numbers of seed per fruit of progeny j in environment i.

3) production efficiency

$$p_{ij} = \frac{1}{6}\sum_{k=1}^{6}\left(\frac{\frac{1}{x}\sum_{l=1}^{x} n_{ijkl}}{m_{ijk}}\right)$$

4) weight per seed

$$w_{ij} = \frac{1}{6}\sum_{k=1}^{6}\left(\frac{\sum_{l=1}^{x} w_{ijkl}}{\sum_{l=1}^{x} n_{ijkl}}\right)$$

where w_{ijkl} is the mean wet seed weight in gram (including the pulp) of fruit l harvested on tree k of progeny j in environment i.

5) seed weight per fruit

$$s_{ij} = \frac{1}{6}\sum_{k=1}^{6}\frac{1}{x}\sum_{l=1}^{x} w_{ijkl}$$

where s_{ij} is the weighted seed weight per fruit in gram of progeny j in environment i.

6) fruit length

$$f_{ij} = \frac{1}{6}\sum_{k=1}^{6}\frac{1}{x}\sum_{l=1}^{x} f_{ijkl}$$

where f_{ijkl} is the length in mm of fruit l harvested on tree k of progeny j in environment i and f_{ij} is the weighted fruit length.

7) fruit diameter and

8) fruit wall thickness are calculated in the same way as fruit length. The fruit diameter is measured in mm at the widest point of the fruit and is a weighted mean. The fruit wall thickness is expressed in cm and measured at the ridges of the transversal cut of the fruit and also a weighted mean.

The statistical analysis for the two randomized complete blocks in each of two different environments consisted of two steps. In the first step the data were analysed separately for each environment.

Since the block design was complete, the corrected progeny totals are the same as the uncorrected totals per progeny. The corrected progeny totals were then entered in a table for the second step of the analysis of variance.

The « environmental error » (σ_e^2) or « pooled MS error » is estimated from the sum of the residual sum of squares for each of the two environments, calculated in step 1. Since the seven parents are considered to form a random sample of the collection all effects are random and the F-tests should be made as follows : progeny x environment interaction is tested against σ_e^2 and the main effects against this interaction.

Because the F_1's and reciprocals are included but not the parents, Griffing's experimental method 3 model 2 is applicable (Griffing, 1956). In case of significant differences among the progenies the SS * for progenies was partitioned further.

The additive genetic variance (σ^2_A) and the non-additive genetic variance (σ^2_D) can be estimated from the combining ability components. Following the (half-) sib analysis of Falconer (1981) and Mayo (1980) it holds that

$$\sigma^2_g = g_{bHS} = \frac{1}{4}\sigma^2_A \quad \text{and}$$

$$\sigma^2_s = \text{cov}_{FS} - 2\,\text{cov}_{HS} = \frac{1}{4}\sigma^2_D,$$

assuming absence of epistatic and of maternal effects (e.g. no reciprocal effect). The component σ^2_g represents the variance for general combining ability (g.c.a.), whereas σ^2_s represents the variance for the specific combining ability (s.c.a.). Furthermore, it was assumed that the population was in Hardy-Weinberg equilibrium. The standard errors of σ^2_A and σ^2_D are 4 × S.E. σ^2_g and 4 × S.E. σ^2_s, respectively.

In order to compare the additive and the non-additive variation for the different characters coefficients of variation are calculated, viz.

$$CV_A = \frac{\sqrt{\sigma^2_A}}{\bar{x}} \quad \text{and} \quad CV_D = \frac{\sqrt{\sigma^2_D}}{\bar{x}}$$

where $\bar{x}$ is the mean value of the character in question (Bos, 1981).

Another method of comparing characters for the sizes of the additive and non-additive variance has been given by Falconer (1981). If we assume absence of epistatic effects, the total genetic variance σ^2_G equals $\sigma^2_A + \sigma^2_D$. Since

$$\sigma^2_g = \frac{1}{4}\sigma^2_A \quad \text{and} \quad \sigma^2_s = \frac{1}{4}\sigma^2_D$$

one can calculate which part of the genetic variance is due to g.c.a. and which part due to s.c.a.

The correlation coefficients are Pearson's product moment correlation coefficients which were calculated on the basis of the 168 pairs of observations for each combination of two characters.

The analysis of variance for estimating the repeatability is similar to the second step analysis of variance. The expections of the mean squares are :

Environments	$\sigma^2_w + b\sigma^2_{G\times E} + bp\sigma^2_E$
Progenies	$\sigma^2_w + b\sigma^2_{G\times E} + be\sigma^2_G$
Environment × Progeny interaction	$\sigma^2_w + b\sigma^2_{G\times E}$
Pooled environmental error	σ^2_w

and the repeatability on a plot mean basis

$$R = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_{G\times E} + \sigma^2_w}$$

where e = number of environments (here e = 2), p = number of progenies (here p = 42), and b = number of blocks (here b = 2). In this study the repeatability represents the proportion of the total phenotypic variance of several measurements of each character which is due to genotype differences. A more comprehensive calculation of the repeatability could not be justified because of the kind of data and the way σ^2_w was estimated.

RESULTS AND DISCUSSION

The requirements which have to be met in order to justify the followed analysis require some comments. The presence or absence of epistatic interactions could not be verified and its absence has to be assumed in order to obtain estimates of additive and dominance genetic variance (Baker, 1978). Although no detailed information is available on the pedigrees of clones it is likely that they were selected as individuals from segregating populations. The set of clones as a whole will to some degree be therefore in linkage equilibrium. Since the selection of the seven clones was primarily based on their reaction against *Phytophthora palmivora* and *Ceratocystis fimbriata* and, in addition, contrasting levels of expression for characters of interest were considered, as was mentioned before, it was assumed that the seven parents, and thus the half sibs, form a random sample of the

Note of the editor.
* SS : Sum of squares.

Turrialba cacao collection for the studied characters. The assumption that the offspring has been derived by means of random mating from the parents is warranted because of the diallel cross. The assumed absence of extra-chromosomal contributions to the phenotypic values was verified (table III) since no significant reciprocal effects have been found for any of the characters.

Since all of the requirements have not been fully met and as there is a lack of normality in the distribution of some of the studied characters (e.g. number of seeds per fruit and production efficiency) only a restricted value can be given to the estimates and thus, the analysis was restricted to estimating g.c.a. and s.c.a. mean squares and effects (Baker, 1978; Mayo, 1980).

The influence of the two environments and of the seven cacao clones on the performance of the progenies is clearly shown in table I, where means for each character are presented, and by the highly significant F-ratios for « environment » for all the characters (table II). This environmental effect could in part be due to the smaller number of fruits measured at the lowland station La Lola in comparison to the experiment in Turrialba due to the difficulty of access.

The mean square variance due to differences between the progenies is, with the exception of the production efficiency, highly significant and allows further partitioning of the sum of squares for progenies (table II). Although the progeny mean square for production efficiency is not significant, also this mean square was further divided into components in order to compare these components for all the characters.

While all characters had a highly significant general combining ability, only three characters

TABLE I

The means for eight characters in each of two environments (e_1 = Turrialba, e_2 = La Lola) as observed among the progeny, obtained from a diallel cross, of seven clones

Character	Clones													
	SCA-6		Catongo		P-7		CC-42		UF-29		UF-613		UF-676	
	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2	e_1	e_2
Maximum seed number	49.6	44.4	47.4	46.8	49.3	46.4	43.6	39.7	45.1	42.8	43.6	42.5	40.9	39.3
Number of seeds/fruit	30.6[1]	36.1	32.4	37.3	33.8	37.8	28.0	30.9	31.8	33.4	31.2	35.6	30.0	34.3
Production efficiency	0.63	0.83	0.72	0.83	0.67	0.84	0.69	0.81	0.68	0.85	0.72	0.86	0.76	0.84
Weight per seed (g)	2.98	3.32	3.05	3.16	3.27	3.52	2.97	3.37	2.91	3.57	3.18	3.88	3.59	3.96
Seed weight/fruit (g)	89.9	111.6	97.8	121.4	108.7	131.9	81.8	99.6	92.3	116.5	98.9	141.7	103.7	133.2
Fruit length (mm)	157.3	174.2	151.8	155.0	157.1	169.0	146.3	148.3	155.2	159.6	162.1	171.8	165.2	168.7
Fruit diameter (mm)	72.6	80.6	82.3	87.1	76.0	83.2	77.1	80.5	78.8	85.0	80.1	84.7	81.3	85.0
Fruit wall thickness (cm)	1.56	1.70	1.66	1.73	1.56	1.61	1.65	1.61	1.69	1.74	1.69	1.68	1.71	1.75

1) The product of number of seeds per fruit and weight per seed is not necessarily equal to the (total) seed weight per fruit since the weight per seed is not a weighted mean, while the others are weighted means.

TABLE II

Mean squares in the analyses of variance of the observations on the forty two progenies of a diallel cross with seven cacao clones and the repeatability for each character. X_1 = maximum seed number per fruit, X_2 = number of seeds per fruit, X_3 = production efficiency, X_4 = weight (g) per seed, X_5 = seed weight (g) per fruit, X_6 = fruit length (mm), X_7 = fruit diameter (mm) and X_8 = fruit wall thickness (cm)

	DF	Mean squares of characters							
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Environment	1	264.01**	658.91**	0.7954**	6.83**	28 594.30**	2 401.20**	1 242.40**	0.0806**
Progenies	41	88.42**	46.26**	0.0100**	0.50**	1 029.50**	673.64**	93.57**	0.0450*
Environnement x progenies	41	20.79**	5.66	0.0071**	0.15	403.85**	138.59**	28.58	0.0253
Error (pooled)	82	11.04	9.53	0.0022	0.16	119.53	65.86	24.04	0.0241
Repeatability		0.52	0.49[1]	0.13	0.35[1]	0.37	0.57	0.38	0.17

* = F-ratio significant at 5 %; ** = F-ratio significant at 1 %.
1) Since the calculated interaction component would have been negative σ^2_{gxe} is assumed to be 0.

had a significant specific combining ability e.g. maximum seed number per fruit, production efficiency and total seed weight per fruit (table III).

When the proportions of the total genetic variance due to g.c.a. and s.c.a., respectively, are compared (table IV) it is apparent that s.c.a. plays only a minor role in the inheritance of the studied characters. This is also apparent when the characters' coefficients of variation for the additive and non-additive variances are compared (table IV). Thus it can be concluded that heterosis does not play an important role for the characters. These results do not confirm earlier work (Purseglove, 1968; Atanda and Toxopeus, 1971; Toxopeus, 1974; and Simmonds, 1981) where it is indicated that heterosis is important. The main reason for this discrepancy could be the use of relatively similar genetic parental material in this diallel cross, whereas the earlier work refers to trees from distinct populations.

A high g.c.a. is essential for the production of « synthetic » varieties which can be obtained by open pollination of selected clones. These should be selected from diverse populations for high yields, disease resistance, and other desirable levels of expression for the characters of interest. To assure a high degree of cross-pollination in the production of the progeny self-incompatibility can be advantageously used. High yielding trees from this synthetic variety can then be selected,

TABLE III

Estimated variance components and their respective standard errors for the characters presented in table II. F-ratio values for the general combining ability (g.c.a.), specific combining ability (s.c.a.) and the reciprocal effect

Component	Variance components of characters							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$\sigma^2_g \pm$ S.E.	45.38±28.09	24.73±14.89	0.0043±0.0027	0.210±0.137	517.33±320.60	407.45±241.16	50.17±30.44	0.0178±0.0119
$\sigma^2_s \pm$ S.E.	10.60±6.15	0.51±2.13	0.0026±0.0010	0.055±0.053	127.37±71.34	18.07±19.95	0.64±5.14	0.0018±0.0056
$\sigma^2_r \pm$ S.E.	0.63±2.08	0.0451±1.660	0.0003±0.0005	negative	27.96±28.64	negative	negative	negative
$\sigma^2_e \pm$ S.E.	11.04±1.70	9.53±1.49	0.0022±0.0003	0.160±0.025	119.53±18.67	65.86±10.29	24.04±3.75	0.0241±0.0038
	F-ratio values of characters							
g.c.a. (F^6_{14})	15.08	24.44	9.85	8.78	14.82	40.95	20.82	7.42
s.c.a. (F^{14}_{82})	2.92	1.11	2.18	1.69	3.13	1.55	1.05	1.15
rec. effect (F^{21}_{82})	1.11	1.01	1.27	0.69	1.47	0.82	0.63	0.44

P = 0.05, $F^6_{14} = 2.85$; $F^{14}_{82} = 1.83$; $F^{21}_{82} = 1.69$

P = 0.01, $F^6_{14} = 4.46$; $F^{14}_{82} = 2.32$; $F^{21}_{82} = 2.10$

TABLE IV

The contributions (in %) of the general (g.c.a.) and specific combining ability (s.c.a.) to the genetic variance and the estimated additive (σ^2_A) and non-additive (σ^2_D) genetic variances, their standard errors and the respective coefficients of variation for the characters presented in table II

Combining ability component	Contributions to the total genetic variance (%)							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
gca	81.1	98.0	62.3	79.2	80.2	95.8	98.7	90.8
sca	18.9	2.0	37.7	20.7	19.8	4.2	1.3	9.2
Genetic variance component	Genetic variances of characters							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$\sigma^2_A \pm$ S.E.	181.52±112.36	98.92±59.56	0.0172±0.0108	0.840±0.548	2 069.32±1 282.40	1 629.80±964.64	200.68±121.76	0.0712±0.0476
$\sigma^2_D \pm$ S.E.	42.40±24.60	2.04±8.52	0.0104±0.0040	0.220±0.212	509.48± 285.36	72.20± 79.80	2.56± 20.56	0.0072±0.00224
CV_A	0.30	0.30	0.17	0.27	0.42	0.25	0.17	0.16
CV_D	0.15	0.04	0.13	0.14	0.21	0.05	0.02	0.05

and progeny tested or clonally propagated which leads to a complementary breeding method similar to the one proposed by Simmonds (1981). If progeny testing (e.g. by means of a diallel cross) reveals a high s.c.a. between a pair of trees, the establishment of a bi-clonal seed garden can be proposed (Bartley and Cope, *in* Moav, 1973) using self-incompatibility in one of the parents if possible to avoid laborious hand-pollination.

Heritabilities were not estimated since the estimation of the environmental variance could not be regarded as very accurate. However, the calculated repeatabilities of the measurements indicate the impact of the environment on the phenotypic expression of each character (table II). Thus, fruit length is the character least influenced by the environment, followed by maximum seed number and number of seeds per fruit. For the latter characters an over-estimation could have occurred since the interaction component $\sigma^2_{G\times E}$ had to be assumed to be zero due to a lower mean square for the environment × progeny interaction than for the pooled environmental error. The same holds true for weight per seed which has a repeatability similar to total seed weight per fruit and fruit diameter (0.35, 0.37 and 0.38, respectively). Production efficiency (0.13) and fruit wall thickness (0.17) have the lowest repeatabilities.

The correlations between the characters (table V) show trends similar to those found in earlier studies where almost three hundred clones were used (Engels, 1983). These results support the assumption that the seven clones, and thus their offspring, represent a random sample of the Turrialba cacao collection. This will be further supported by the outcome of a cluster analysis where two hundred and ninety four clones of the Turrialba collection were included. This study shows that the seven parents of the diallel cross are well distributed over the phenogram with similarity coefficients varying from 0.220 (Catongo) to 13.156 (UF-613) (Engels, in preparation). The characters which show relatively large s.c.a. effects form a cluster due to their relatively strong correlation and this explains why the derived trait, production efficiency, is included in the group of characters which are important selection criteria in the breeding work.

TABLE V

The product-moment correlation coefficients between the characteristics studied, based on 168 observations of a complete diallel of the seven clones in four blocks

	1	2	3	4	5	6	7	8
1. Maximum seed number/fruit	-							
2. Number of seeds/fruit	0.38	-						
3. Production efficiency	-0.55	0.48	-					
4. Weight per seed	-0.43	0.04	0.46	-				
5. Seed weight/fruit	-0.07	0.60	0.61	0.75	-			
6. Fruit lenght	-0.11	0.31	0.40	0.50	0.57	-		
7. Fruit diameter	-0.32	0.37	0.65	0.50	0.63	0.45	-	
8. Fruit wall thickness	-0.34	0.00	0.36	0.36	0.29	0.39	0.79	-

(0.05 %) = 0.12 ; (0.01 %) = 0.18

CONCLUSIONS

For all characters highly significant g.c.a. effects were found; only three characters showed a significant variation for s.c.a. : maximum seed number, production efficiency and total seed weight. At the same time the percentage of the total genetic variance due to variation for s.c.a. was the highest for these characters.

On the basis of the percentage of the total genetic variance due to variation for g.c.a. and/or for s.c.a., and the coefficients of variation CV_A, and CV_D the positive association between g.c.a., s.c.a., and additive and non-additive variance, respectively, has been clearly shown for all characters.

Significant reciprocal effects were not found and, therefore, it is concluded that maternal effects are unlikely to play an important role in the inheritance of the studied characters.

Since the calculated repeatability is comparable with heritability in a broad sense it can be concluded that fruit length is the most heritable character and production efficiency the least.

In cacao breeding parents should be selected with desirable levels of expression for the characters of interest when producing open pollinated progeny. Promising trees can be selected from a synthetic variety for progeny testing or clonally propagated for a further breeding cycle. Thus, in this way clones, synthetics and/or F_1 hybrids can be established, depending on the needs of the breeder.

The present work was part of the programme of the Cacao Breeding Programme and the Plant Genetic Resources Unit of CATIE (Turrialba, Costa Rica). The latter is supported by the Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) with funds from the Ministry of Economic Cooperation at Bonn, Federal Republic of Germany.

The author expresses his thanks to Mrs. J. Morera, R. Bonilla and R. Sanchez for the data collection, to Dr I. Bos for the numerous discussions and suggestions on the statistical aspects and to Dr J. Lazier for his critical reading and correction of the manuscript.

REFERENCES

Agricultural genetics : selected topics. MOAV (R.) (ed.), Wiley (New York), 1973.

ATANDA (O. A.), TOXOPEUS (H.). — A proved case of heterosis in *Theobroma cacao* L. *In* : III[rd] International Cocoa Research Conference, Accra (Ghana), 23-29 nov. 1969, Cocoa Research Institute (Tafo, Ghana), 1971, p. 545-551.

BAKER (R. J.). — Issues in diallel analysis. *Crop Science* (Madison), 1978, vol. 18, p. 533-536.

BOS (I.). — The relative efficiency of honey comb selection and other procedures for mass selection in winter rye (*Secale cereale* L.), Agric. Univ., Dept of Plant Breeding (Wageningen), 1981, 172 p.

ENGELS (J. M. M.). — A systematic description of cacao clones. III. Relationships between clones, between characteristics and some consequences for the cacao breeding. *Euphytica* (Wageningen), vol. 32, n° 3, nov. 1983, p. 719-733.

ENGELS (J. M. M.). — A systematic description of cacao clones. IV. Some evidence of tetraploid inheritance. *Café Cacao Thé* (Paris), vol. 28, n° 2, avril-juin 1984, p. 95-101.

ENGELS (J. M. M.). — A systematic description of cacao clones. VI. Their classification and identification (in preparation).

Experimento La Lola No. 29 y Turrialba No. 14. Estudio de herencia cuantitativa de varios componentes de rendimiento en cacao : cruzamientos dialelicos entre clones con caracteres contrastados. CATIE, Turrialba, Costa Rica (mimeographed), 1972.

FALCONER (D. S.). — Introduction to quantitative genetics. 2nd ed., Longman (London), 1981, 340 p.

GRIFFING (B.). — Concept of general and specific combining ability in relation to diallel crossing systems. *Austr. J. Biol. Sciences* (Melbourne), 1956, vol. 9, p. 463-493.

HAYMAN (B. I.). — The theory and analysis of diallel crosses. *Genetics* (Austin), vol. 39, nov. 1954, p. 789-809.

MAYO (O.). — The theory of plant breeding. Clarendon Press (Oxford), 1980, 293 p.

PURSEGLOVE (J. W.). — Tropical crops, Dicotyledons, Longmans Green and Co. (London), 1968, 719 p.

SIMMONDS (N. W.). — Principles of crop improvement. Longman (London), 1981, 408 p.

TOXOPEUS (H.). — Breeding for black pod resistance in *Theobroma cacao* L. *In* : Phytophthora Disease of Cocoa (P. H. Gregory, Ed.), Longman (London), 1974, p. 202-210.

ENGELS (J. M. M.). — **Une description systématique des clones de cacaoyer. V. Aspects génétiques quantitatifs de plusieurs caractères des fruits.** *Café Cacao Thé* (Paris), vol. XXIX, n° 1, janv.-mars 1985, p. 3-10, 5 tabl., 15 réf.

Un croisement diallèle de sept clones de cacaoyer a été analysé. L'aptitude à la combinaison générale, l'aptitude à la combinaison spécifique et les effets réciproques ont été estimés et la répétabilité a été calculée pour chacun des huit caractères des fruits étudiés. La variation additive et la variation de la dominance ont été estimées et exprimées en pourcentage de la variance génétique totale et sous la forme de coefficients de variation. Les corrélations entre caractères sont données et une méthode d'amélioration basée sur la mise à l'épreuve des descendances et la multiplication clonale proposée.

ENGELS (J. M. M.). — **A systematic description of cacao clones. V. Quantitative genetic aspects of several fruit characters.** *Café Cacao Thé* (Paris), vol. XXIX, n° 1, janv.-mars 1985, p. 3-10, 5 tabl., 15 réf.

A diallel cross of seven cacao clones was analysed. General combining ability, specific combining ability and reciprocal effects were estimated and the repeatability was calculated for each of eight fruit characters. Additive and dominance variations were estimated and expressed as per cents of the total genetic variance and as coefficients of variation. The correlations between characters are given and a breeding method based on progeny testing and clonal propagation is proposed.

ENGELS (J. M. M.). — **Systematische Beschreibung von Klonen des Kakaobaumes. V. Quantitative genetische Aspekte über mehrere Eigenschaften der Früchte.** *Café Cacao Thé* (Paris), vol. XXIX, nº 1, janv.-mars 1985, p. 3-10, 5 tabl., 15 réf.

Es wurde eine diallele Kreuzung von sieben Klonen des Kakaobaumes untersucht. Die Eignung zur allgemeinen Kombination, die Eignung zur spezifischen Kombination und die reziproken Wirkungen wurden abgeschätzt und es wurde die Wiederholbarkeit für jede der acht Eigenschaften der untersuchten Früchte berechnet. Es wurden die additive Variation und die dominante Variation abgeschätzt und als Prozentsatz der gesamten genetischen Varianz und als Form von Variationskoeffizienten ausgedrückt. Die Zusammenhänge zwischen den Eigenschaften werden gegeben und eine Methode zur Verbesserung vorgestellt, welche auf der Erprobung der Nachkommenschaft und der klonalen Vermehrung beruht.

ENGELS (J. M. M.). — **Descripción sistemática de los clones del cacao. V. Aspectos genéticos cuantitativos de varios caracteres de los frutos.** *Café Cacao Thé* (Paris), vol. XXIX, nº 1, janv.-mars 1985, p. 3-10, 5 tabl., 15 réf.

Se ha procedido al análisis de un cruce dialelo de siete clones de cacao. Se ha procedido a la evaluación de la aptitud para la combinación general, la aptitud para la combinación específica y asimismo, los efectos recíprocos, habiéndose calculado también la repetabilidad para cada uno de los ocho caracteres de los frutos estudiados. La variación aditiva y la variación de la dominancia se han evaluado y expresado en porcentaje de la variación genética total y ello en forma de coeficientes de variación. Se indican las correlaciones entre caracteres y se propone un método de mejoramiento fundado en la puesta a prueba de las descendencias y la multiplicación clonal.

A SYSTEMATIC DESCRIPTION OF CACAO CLONES

VI. THEIR CLASSIFICATION AND IDENTIFICATION

J.M.M. ENGELS

INDEX WORDS

Theobroma cacao, germplasm collection, cacao characters, multivariate statistical analyses, numerical taxonomy, classification, identification.

SUMMARY

Various analyses have been conducted to search for consistent and meaningful groupings of 294 cacao cultivars (mainly clones) of which 39 characters were described when grown under similar environmental conditions. Although the classical taxonomic groupings of cacao cultivars into genetic and fruit shape classes were found to be meaningful, a new classification based on 10 characters has been proposed to facilitate identification of cacao clones, since the majority of the new cultivars are hybrids between two or more of the genetic classes and could not be classified in any of the old groupings, which were made for populations. Several methods to identify unknown, duplicate and mislabelled cultivars are presented.

INTRODUCTION

The importance of germplasm collection to a breeding programme is strongly dependent on the availability of accurate descriptions of the accessions and on the taxonomic identification of the germplasm. The classification of cultivars has received significant attention only recently, with the establishment of germplasm collections throughout the world. A detailed review of infraspecific classification is given by Baum (1981) and the methodology and results of the following classifications were studied: alfalfa (Small and Brookes, 1984), barley (Baum, 1981), Capsicum (Pickersgill et al., 1979), eggplant (Martin and Rhodes, 1979), lettuce (Tillge, 1984) and Vicia faba (Higgins et al., 1981).

In order to be able to identify a given cacao accession several classification systems have been proposed (Cuatrecasas, 1960). The most widely accepted classification of cacao clones was proposed by Cheesman (1944) using a large number of local populations:

1.	Criollo	a) Central American Criollos
		b) South American Criollos
2.	Forastero	a) Amazonian Forasteros
		b) Trinitarios (hybrid populations between 1 and 2 and of recent origin)

These groupings have been referred to as "genetic complexes" (Soria, 1970a). In another classification the Trinitario hybrids and sometimes also the Forastero cacaos are divided into different fruit shapes, e.g. 'Angoleta', 'Cundeamor', 'Amelonado' and 'Calabacillo'. These terms have no "cultivar" significance applying to nothing but the shape of the pod (Cheesman, 1944).

As in this study the cacao germplasm was almost entirely composed of clones and not of heterogenous populations, and since many of these clones were hybrids between trees belonging to different "genetic complexes" or single tree selections from segregating populations, it was decided to study the applicability of the conventional classifications in detail and to see if another classification for the identification of clonal material should be developed.

The identification of duplicate accessions and errors in identification are serious concerns in the management of genetic resources (Bryant et al., 1977; Porter and Smith, 1982, respectively). Since these problems appear particularly in clonally multiplied perennial crops these aspects were given due attention in this study too.

MATERIALS AND METHODS

Cacao cultivars evaluated

The data used here have been taken from the cacao germplasm collection maintained at the Centro Agronomico Tropical de Investigacion y Ensenanza (CATIE) at Turrialba, Costa Rica.

Detailed information on the edaphic and climatic conditions of the planting site have been presented elsewhere (Engels, 1981) along with detailed information on the origins and the pedigrees of the 294 cacao cultivars used in this study (see also Appendix 1). Of the 294 accessions a few were genetically not uniform, including two populations. Several of the cultivars are products of breeding programmes in Brazil, Costa Rica, Trinidad, Ecuador and some other countries. The collection is one of the major cacao collections in the world (Soria, 1975; IBPGR, 1981) with a wide geographical representation (Appendix 1) and so with a considerable genetic diversity.

Characters evaluated

Thirty-nine characters (descriptors) were chosen for this study (Table 1). Detailed information on these and on the methodology of data collection has been presented elsewhere (Engels, 1981 and 1983c). The data of the majority of the characters were measured and then classified on a scale from 1-9. Zero was used when a cultivar did not express a character (e.g. absence of pigmentation). The number of classes used for a given character was calculated by dividing the range of values obtained by the mean standard deviation of the 294 cultivars (Engels, 1983b). When the total number of classes calculated was higher than 9, the number of classes was reduced to 9. This ensures that when a character is highly variable 'within cultivars' there will be fewer classes, and thus the character will be less discriminative. Assuming a normal distribution of all the measurements for a given clone and character, the limits of the mean $\pm$ 2 standard deviations of that distribution contain 95.46% of all the measurements of that clone (Sokal and Rohlf, 1969). Therefore, if the frequency distribution of two clones do not overlap for a given number of observations, these clones will be significantly different for the given character.

For the classification and identification of the cacao clones in this study a selection was made within the 39 characters and the following selection criteria which have been partly studied earlier, were used:

- reliability of scoring;
- ease of observation or measurement;
- discriminative value;

- agronomic importance;
- taxonomic importance in previous classifications;
- correlations between characters.

Classification

In order to be able to investigate previous classifications (Lockwood and Gyamfi, 1977; Soderholm and Carris, 1979; Soria, 1970a, b; and Soria and Enriquez, 1981) the clones in this study were classified using the characteristics presented in the literature for each of the taxonomic classes (Cheesman, 1944; Cuatrecasas, 1960; van Hall, 1932; Mora, 1956; and Soria, 1970a). The taxonomic details of the individual clones as found in the literature are included in Appendix 1. Two computer classifications were conducted, using this information; the first one with parameters as defined in the literature, and a second one with fewer parameters and less restrictive (Table 3).

According to the procedures and recommendations of Baum (1981), Harlan and de Wet (1971), Higgins et al. (1981) and Pickersgill (1985) to obtain an informal classification of cultivars and to give - if possible - a biological basis to the recognized groupings, of the 39 characters 10 were chosen as being discontinuous (e.g. high taxonomic value), semi-continuous or continuous (e.g. low taxonomic value) (Table 2). This classification is referred to as "the ten characters classification" or "the ten character key" in the text.

Identification

The identification of clones, and the detection of identification errors and duplicate accessions was attempted using identification by comparison, identification by matching (Pankhurst, 1978), cluster and principal component analyses. The identification of clones was also attempted by totaling the descriptor states of each clone and sorting them in ascending order.

Numerical analyses

The majority of the analyses was done on a micro-computer (CPU 64 Kbytes) using the data base management programme dBASE II, version 2.3B (Ashton-Tate, 1983) and statistical package

STATPAK, version 2.1 (Anonymous, 1982). The multivariate analyses consisted of cluster analyses (nearest neighbour method), using CLUSTAN 1C package (Wishart, 1978), P1M and P2M of BMDP (Dixon et al., 1983), and principal component analysis, version 9.1 of SPSS (Nie et al., 1982). For the latter analysis the characters (marked with an asterisk in Table 1) had to be limited in number following the selection criteria given before, because of limitations in computer capacity.

RESULTS AND DISCUSSION

Classification

Relationships between characters. Since the relationships between characters will directly influence the selection of characters in any taxonomic classification, this aspect was included in this study. In an earlier study (Engels, 1983c) a correlation matrix of 33 characters was presented as well as the results of a factor analysis in which the characters were plotted against several factors. The results of the cluster analysis in this study agree with the results of the earlier factor analysis (Fig. 1). The six additional characters in this study (numbers 14, 21, 22, 37, 38 and 39 of Table 1) do not show strong relationships with other characters, although the pigmentation characters cotyledon colour (no. 14), anthocyanin in ligule (no. 37) and filament (no. 38) form a weak cluster with a highest correlation of $r = 0.31$. Anthocyanin intensity in ridges of ripe fruits (no. 22) and new flush colour (no. 39) are significantly correlated ($r = 0.41$) but they are independent of the aforementioned cluster. Basic fruit surface colour (no. 21) is the most independent character (Fig. 1) but since it is not easy to observe and the basic colour sometimes is masked by pigmentation of the fruit wall, this character was not given a high priority for classification purposes.

The characters chosen for the ten character classification (Table 2) are not strongly correlated with each other (Fig. 1). The only exception is fruit surface rugosity (no. 25) which is fairly strongly correlated with the ratio fruit width/fruit length (no. 17) and with secondary furrow depth (no. 24). It was chosen for the key because of its frequent use in the characterization of taxonomic groups (Table 3).

In the principal component analysis various discriminative and

explicable factors could be recognized (Table 4). The first two factors were used to study the grouping of the cultivars in scatter diagrams (Figures 2, 3 and 4). Factor 1 is mainly composed of "size" characters such as fruit index (no. 1), seed index (no. 2), fruit width (no. 16) and fruit wall thickness (no. 19) and these four characters together are responsible for 34.5% of the total variation between cultivars. Factor 2 consists mainly of fruit "shape" characters, e.g. ratio fruit width/fruit length (no. 17), ridge pair separation (no. 23), secondary furrow depth (no. 24), fruit surface rugosity (no. 25) and basal fruit constriction (no. 27) which are responsible for 15.9% of the total variation. Other factors are composed of flower size characters (factor 3), pigmentation characters (factors 4 and 5), seed number (factor 7) and a mixed group (factor 6). These results are very similar to the ones obtained by factor analysis (Engels, 1983c). Clones were non-selectively taken along the diagonal and main axes of the scattergram and have been presented diagrammatically in terms of longitudinal and transverse sections of their fruit (Fig. 5). There is a clear trend in the distribution of the clones in terms of fruit size (factor 1) and fruit shape (factor 2) in the scattergram.

Relationships between cultivars

Cluster analysis and principal component analysis were used to group similar cultivars as suggested by Martin and Rhodes (1979). With cluster analysis some strong clusters were found. However, about half of the cultivars were not clustered. No obvious groupings were found in principal component analysis in which all cultivars were scattered against the factors 1 and 2, 1 and 3 and 2 and 3, respectively (as is illustrated in Fig. 2).

Grouping was also attempted by calculating the totals of all the character states per cultivar and then ranking them in ascending order. No groupings were observed, the totals were evenly distributed over the whole range. The classifications as proposed by Cheesman (1944) and others were checked by plotting the cultivars with a described taxonomic grouping in a scatter diagram (Fig. 2). The cultivars previously classified as Forastero (70) were, with a few exceptions, all placed below the horizontal axis and mainly concentrated in the third quadrant. The Criollos (11) and the hybrid cultivars with a Criollo parent (25) were mainly located above the horizontal

axis in the first quadrant. The 64 Trinitario cultivars were mainly located above the horizontal axis and distributed equally in the first and fourth quadrants. The Trinitario clones in the third quadrant were concentrated near the origin and those in the second quadrant were more scattered. The Trinitario cultivars thus apparently form a hybrid group between Forastero and Criollo types with Criollo characteristics evidently dominant over Forastero characteristics. This conclusion is supported by the means of the totals of all the character states analysed in an ANOVA where $\bar{x}_{Forastero} = 136.4$, $\bar{x}_{Criollo} = 153.3$ and $\bar{x}_{Trinitario} = 157.0$ with an F value for different groupings of 41.9 ($P < 0.01$). The remaining half of the cultivars, which were not taxonomically grouped in the literature and, therefore, not plotted fall almost all within limits shown by the extremes of the Forastero and Criollo types (Fig. 3).

The classification of the cacao clones by genetic groups does not lead to a useful classification for the identification of cultivars due to a rather amorphous distribution. Cultivars previously classified on the basis of shape were plotted in a scatter diagram to identify any features held in common (Fig. 3). The 'Amelonado' cultivars were scattered throughout the diagram with a concentration in the third quadrant matching the Forastero group fairly closely. The 'Angoleta' cultivars were concentrated in the fourth quadrant and matched the distribution of Criollo and Trinitario. As only two cultivars were classified as Calabacillo no conclusions could be drawn. The 'Cundeamor' cultivars did not show similarities to any of the other groups. Fruit shape does thus not appear to be a reliable basis for a general classification but it may permit rough groupings. The coordinates of each clone of the scattergram (vertical axis is factor 1, horizontal axis is factor 2) are included in Appendix 1.

The results of the narrowly defined computer search on the data set using the genetic (Criollo and Forastero) and shape (4 classes) classifications are given in Table 5. With the exception of the Forastero group no or little correspondence was found between the characteristics for each given in the literature and the results obtained here. A second search using more broadly defined and, in general, fewer parameters showed only a low degree of concurrence between the theoretically described and the actually found taxonomic classes. The poor results produced by both classifications could be due either to

the difficulties in describing fruit shape numerically or environmental differences of the present and past experiments where upon the characteristics of the taxonomic classes have been based or to a combination of both.

In order to interpret the results of the principal component analysis some of the pedigree groups were plotted in a scatter diagram (Fig. 4) and the results are presented in Table 6. The pedigrees of the 'Nacional' cultivars resemble the Criollo group strongly, contradicting the findings of Soria (1970a). This may be due to the influence of the 'Venezuelano' cultivars which were not included in this study.

The scatter diagram also confirmed visual observations with regard to mislabeled cultivars (cultivars 139, 143, 144, 228 and 243). Thus the scattergram has shown phenotypic relationships between cultivars, allowed the determination of genetic affinities and provided information on errors.

Earlier taxonomic classifications reported in the literature are thus of little help in characterizing or typifying a cultivar as no distinct groups of cultivars occurred in cluster and principal component analysis for there was considerable overlapping due to intercrossing of the members of the different genetic classes - mainly Criollo and Trinitario -, no consistent grouping of the members of the shape classes, and little agreement between the theoretical and actual data of the taxonomic classes.

The characters of the ten characters classification were selected with the aim to obtain small cultivar groups and in so doing to facilitate identification of individual cultivars, including possible duplicates (Table 7). Ten characters gave a satisfactory classification, as 64.6% of the cultivars were keyed out singly, 19.0% in pairs, 13.3% in triplets, 1.4% in quadruplets and 1.7% in groups of five. Therefore, it can be concluded that for an appropriate classification the chosen characters are sufficient in number and do discriminate between clones in an adequate way.

Identification

The identification of unknown cultivars is a major objective of a classification. When a given cultivar has been described under the same environmental conditions as the cultivars from

which the key was constructed it will be possible to obtain a strong indication which cultivars are similar and/or the same. When two cultivars are found to be the same it will be necessary to compare the remaining characters which were not used in the key to assure that they are identical (Chang, 1976).

The ten characters classification was tested for its usefulness in the detection of potential duplicates. Accessions were chosen from the collection which had the same or similar names. Only two (clones 7 and 8) were definitively identified as duplicates and two others (clones 57 and 58) were possible duplicates. The same eight clones were also tested for duplication using other methods in order to be able to compare them. If it is assumed that subjective judgement based on all characters and their reliability gives the best results in duplicate identification, the computer search would be the next best one, followed by principal component analysis, the ten characters classification, cluster analysis, and the grand total method (Table 8).

As long as the decision whether two clones are duplicates or not is based on a phenotypic approach such a decision remains doubtful. Therefore, electrophoresis appears to hold promise as a tool to identify duplicates (Anderson, 1984).

Scatter diagrams can indicate possible mistakes in cultivar naming particularly when a given cultivar is expected to fall within a well defined cluster but actually falls outside of it. However, the number of clearly demarcated clusters in a scatter diagram is small and thus identification of errors by this method is only possible if the genetic group or the pedigree of the cultivar are known. Therefore, a complete and detailed record of the pedigree of each cultivar is an important prerequisite to fully exploit a germplasm collection (Brandenburg, et al., 1981) as well as a proper description of each accession for a set of well defined and highly heritable characters (Porter and Smith, 1982). In case less heritable characters are used in the description of accessions this has to be done under as uniform conditions as possible.

Due to space problems to include the mentioned phenogram in this paper, it was decided to use the scatter diagram (Fig. 3) to show the proper distributions of the seven clones which were used in a diallel cross of a previous study (Engels, 1985). Their distribution in the scattergram and in the ten characters

classification supported the contention in that study that they are a random sample of the collection.

ACKNOWLEDGEMENTS

The present work was part of the Plant Genetic Resources Project at CATIE, supported by the German Agency for Technical Cooperation (GTZ) with funds from the German Ministry of Economic Cooperation at Bonn, Federal Republic of Germany. The author wishes to express his thanks to the colleagues in the Project for assisting in the data collection, to Mr Solomon Zewdie (PGRC/E) and to Mr Robin Sayer (ILCA) for the computer analysis, to Dr Barbara Pickersgill for her valuable suggestions and to Dr J. Lazier for his critical reading and correction of the manuscript.

REFERENCES

ANDERSON, G., 1984. Cultivar identification. Advances in research and technology of seeds. Part 9. Pudoc, Wageningen, p. 9-24.

ANONYMOUS, 1982. The NWA STATPAK, version 2.1 (Preliminary manual): Northwest Analytical Inc. Portland, Gregon, U.S.A. 215pp.

ASHTON - TATE, 1982. dBASE II (Version 2. 3B) Assembly - Language Relational Database Management System. Ashton-Tate, Culver City, California, U.S.A. 156pp.

BAUM, B.R., 1981. Taxonomy of the infraspecific variability of cultivated plants. Kulturpflanze 29: 209-239.

BRANDENBURG, W.A., E.H. OOST and J.G. VOOREN, van de, 1981. Taxonomic aspects of the germplasm conservation of cross-pollinated cultivated plants. In: E. Porceddu and G. Jenkins (eds.). Seed regeneration in cross-pollinated species. Proc. CEC/Eucarpia seminar, Nyborg, Denmark, 15-17 June, 1981. A.A. Balkema Rotterdam. p. 33-41.

BRYANT, P., J. KALTENHAUSER, K.M. RAWAL and M.J. SNYDER, 1977. Finding duplicates in genetic resources collections. Information Sciences/Genetic Resources Program, University of Colorado, Boulder, Colorado USA. 35pp.

CHANG, T.T., 1976. Manual on genetic conservation of rice germ plasm for evaluation and utilization. International Rice Research Institute, Los Banos, Philippines. 77pp.

CHEESMAN, E.E., 1944. Notes on nomenclature, classification and possible relationships of cacao populations. Tropical Agriculture (Trinidad) 21: 144-159.

CUATRECASAS, J., 1960. Cacao and its allies. A taxonomic revision of the genus Theobroma. Contributions U.S. Nat. Herbarium (Smithsonian Institution) 35: 379-605.

DIXON, W.J., et al. (eds), 1983. BMDP Statistical Software. 1983 Revised printing. University of California Press, Berkley, USA. 73pp.

ENGELS, J.M.M., 1981. Genetic resources of cacao: a catalogue of the CATIE collection. Technical Series. Technical Bulletin CATIE, No. 7. CATIE, Turrialba, Costa Rica. 191pp.

ENGELS, J.M.M., 1983a. A systematic description of cacao clones. I. The discriminative value of quantitative characteristics. Euphytica 32: 377-385.

ENGELS, J.M.M., 1983b. A systematic description of cacao clones. II. The discriminative value of qualitative descriptors and practical compatibility of the discriminative values of quantitative and qualitative descriptors. Euphytica 32: 387-396.

ENGELS, J.M.M, 1983c. A systematic description of cacao clones. III. Relationships between clones, between characteristics and some consequences for the cacao breeding. Euphytica 32: 719-733.

ENGELS, J.M.M., 1985. A systematic description of cacao clones. V. Quantitative genetic aspects of several fruit characters. Cafe Cacao The 29: 3-10.

HALL, C.J.J. van, 1932. Cacao. MacMillan Co., 2nd ed. 514pp.

HARLAN, J.R. and J.M.J. DE WET, 1971. Toward a rational classification of cultivated plants. Taxon 20: 509-517.

HIGGINS, J., J.L. EVANS and P.J. RED, 1981. Classification of Western European Cultivars of Vicia faba L.J. natn. Inst. agric. Bot. 15: 480-487.

INTERNATIONAL BOARD FOR PLANT GENETIC RESOURCES, 1981. Genetic resources of cocoa. IBPGR secretariat, Rome. 25pp.

LOCKWOOD, G. and M.M.O. GYAMFI, 1977. The CRIG cocoa germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Cocoa Research Institute Tafo - Akim, Ghana (mimeographed). 61pp.

MARTIN, F.W. and A.M. RHODES, 1979. Subspecific grouping of eggplant cultivars. Euphytica 28: 376-383.

MORA, J., 1956. Origen y tipos de cacao. Suelo Tico (Costa Rica) 9: 196-200.

NIE, N.H. et al. (eds.), 1982. SPSS: Statistical Package for the Social Sciences (2nd ed). McGraw - Hill Book Company, New York, USA. 675pp.

PANKHURST, R.J., 1978. Biological identification. The principles and practice of identification methods in biology. Edward Arnold (Publishers) Ltd., London, England. 104pp.

PICKERSGILL, B., C.B. HEISER and J. MCNEILL, 1979. Numerical taxonomic studies on variation and domestication in some species of Capsicum. In: J.G. Hawkes, R.N. Lester and A.D. Skelding (eds.). The Biology and Taxonomy of the Solanaceae. Linnan Society Symposium Series No. 7. p. 679-700.

PICKERSGILL, B., 1985. Evolution of hierarchical variation patterns under domestication and their taxonomic treatment. (in press).

PORTER, W.M. and D.H. SMITH, JR., 1982. Detection of identification errors in germplasm collections. Crop Science 22: 701-703.

SMALL, E. and B.S. BROOKES, 1984. Taxonomic circumscription and identification in the Medicago sativa - falcata (alfalfa) continuum. Economic Botany 38:83-96.

SODERHOLM, P.K. and M.C. CARRIS, 1979. Cacao clones in the germplasm collection of the United States Department of Agriculture. Miami, Florida, and Glenn Dale, Maryland. 18pp.

SOKAL, R.R. and F.J. ROHLF, 1969. Biometry. W.H. Freeman and Co. San Francisco. 776pp.

SORIA, J., 1970a. The present status and perspectives for cacao cultivars in Latin America. In: Proceedings of the Tropical Region of the American Society for Horticultural Science. Vol. 14. Miami, Florida. p. 53-65.

SORIA, J., 1970b. The latest cocoa expeditions to the Amazon Basin. Turrialba (Costa Rica) 15: 5-15.

SORIA, J., 1975. Recent cocoa collecting expeditions. In: O.H. Frankel and J.G. Hawkes (eds.). Crop genetic resources today and tomorrow. International Biological Programme 2. Cambridge University Press, Cambridge. p. 175-179.

SORIA, J. and G.A. ENRIQUEZ (eds,), 1981. International cacao cultivar catalogue. Technical series. Technical Bulletin No. 6. CATIE, Turrialba, Costa Rica. 156pp.

TILLGE, L., 1984. Variety identification in Lactuca sativa L. Seed Sci. and Technol. 12: 919-933.

WISHART, D., 1978. Clustan user manual. Program Library Unit, Univ. Edinburgh.

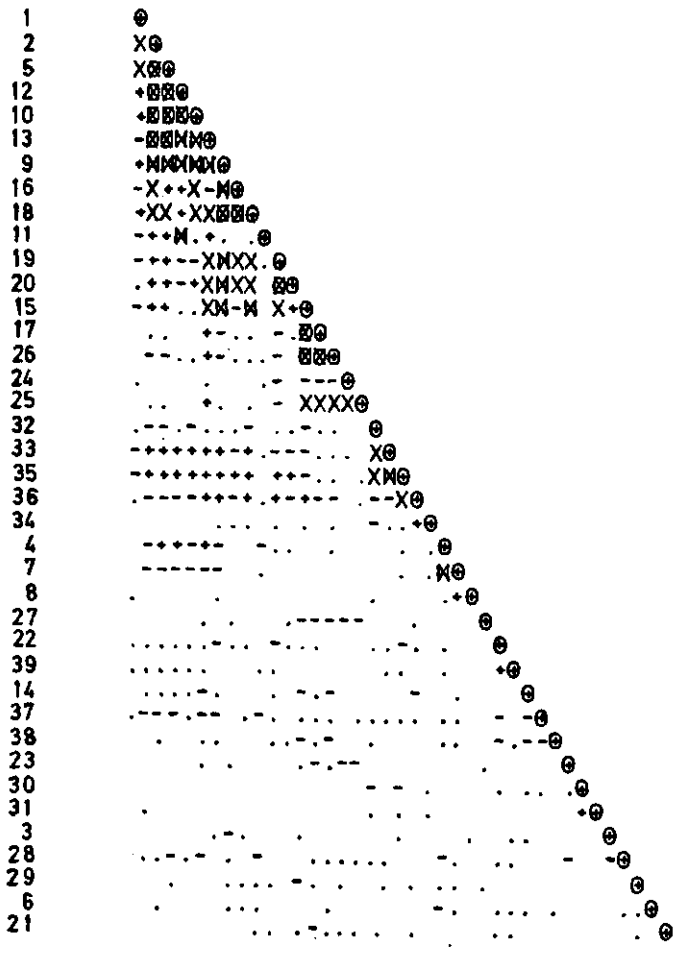


Figure 1.
Shaded correlation matrix of sorted characters based on maximum similarity. The character numbers coresspond with the numbers presented in Table 1 and the correlations used are absolute values.

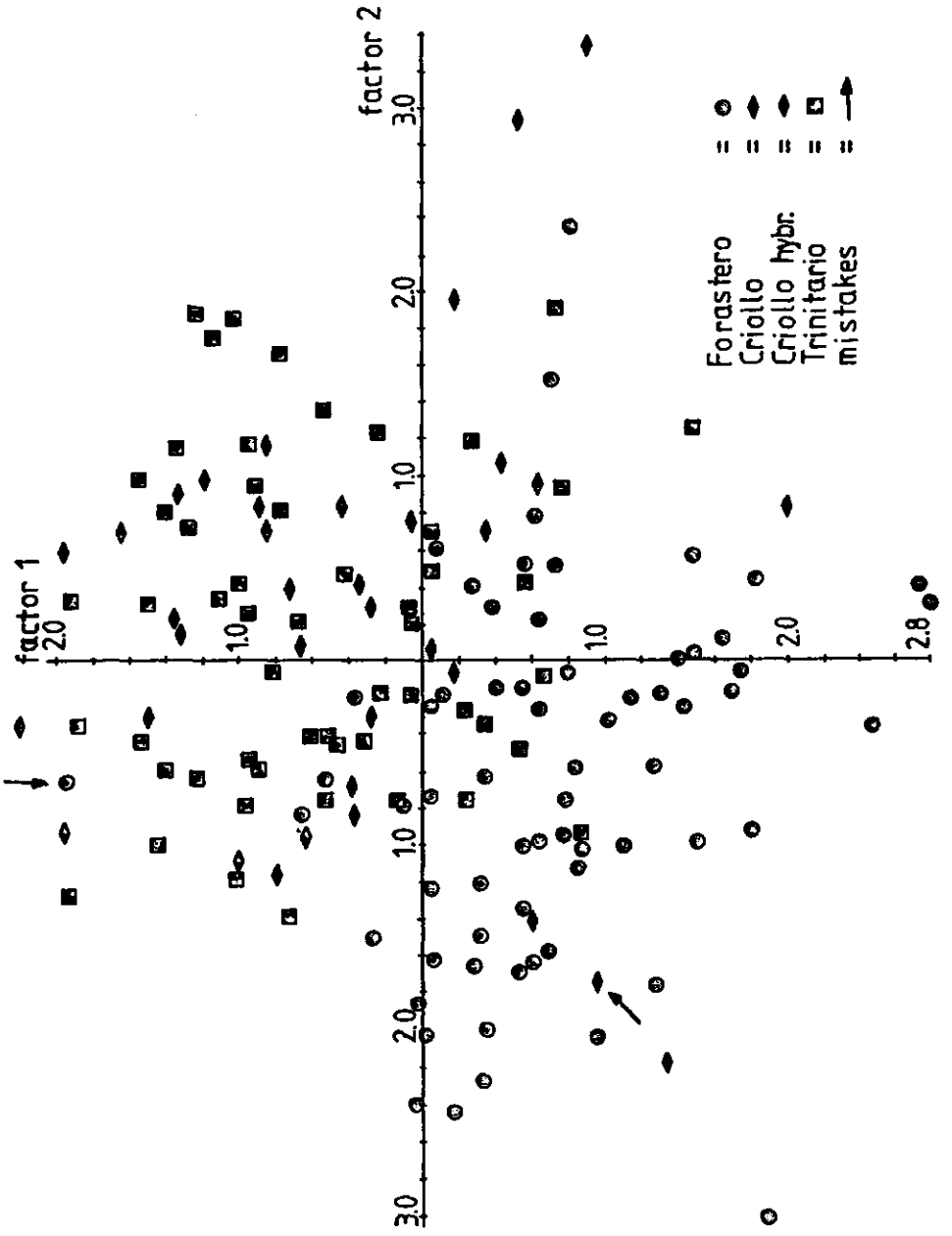


Figure 2.

Cacao clones reported to belong to any of the "genetic complexes" plotted on the first two axes as recognized by a principal component analysis.

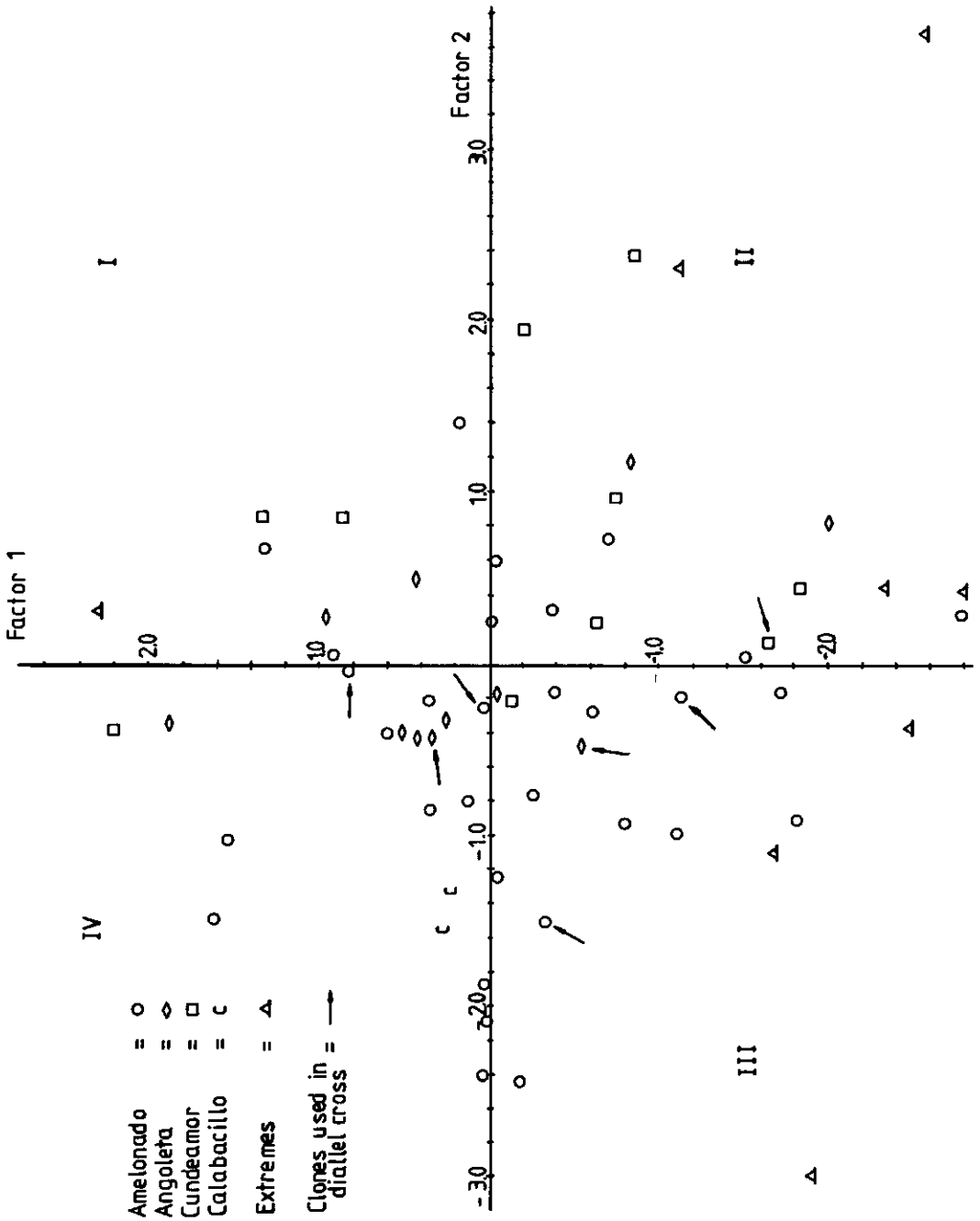


Figure 3.

Fruit shape classes and some of the extreme cacao clones plotted on the first two axes as recognised by a principal component analysis.

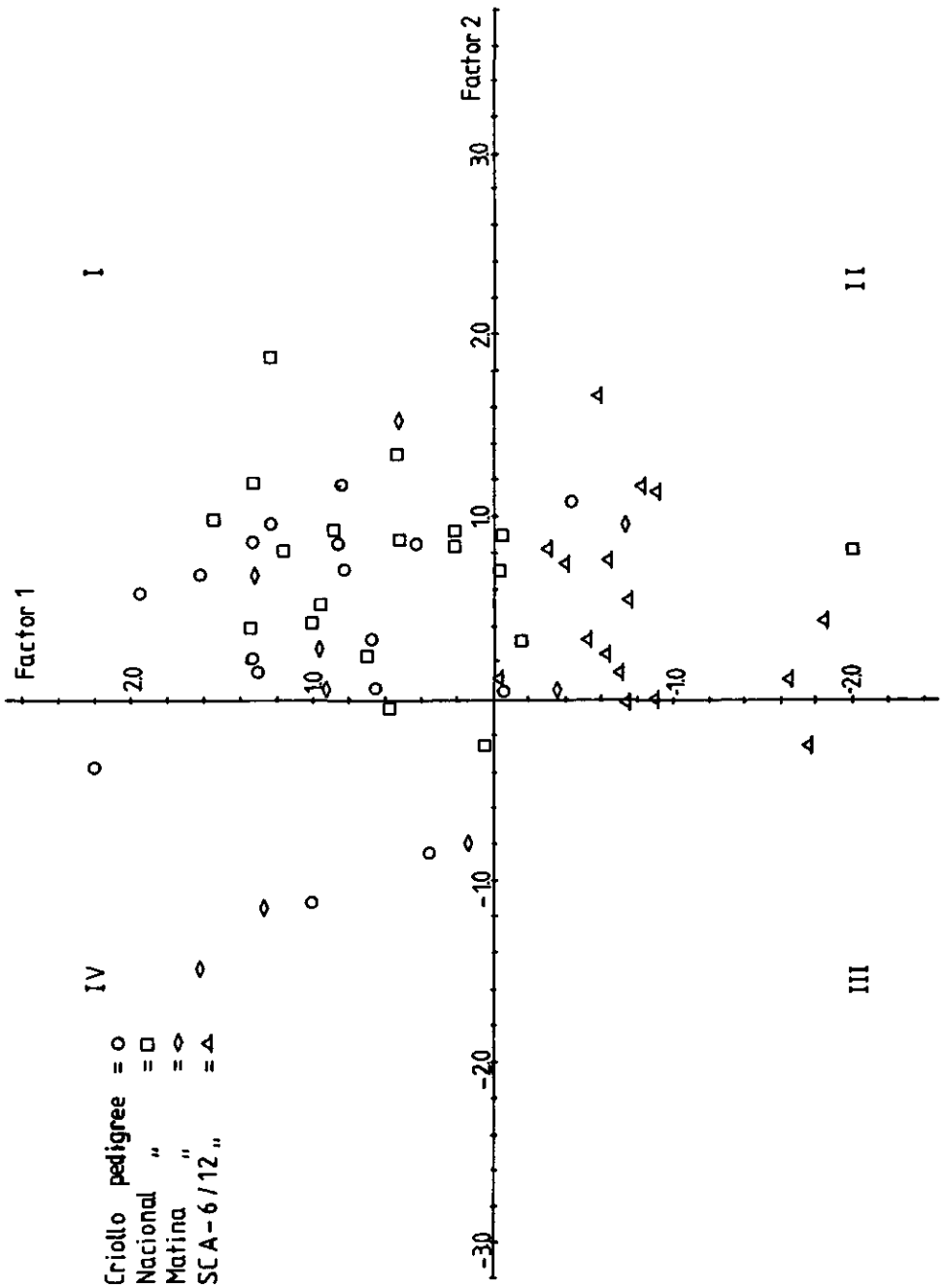


Figure 4.

Pedigrees of some important clones plotted on the first two axes as recognized by a principal component analysis.

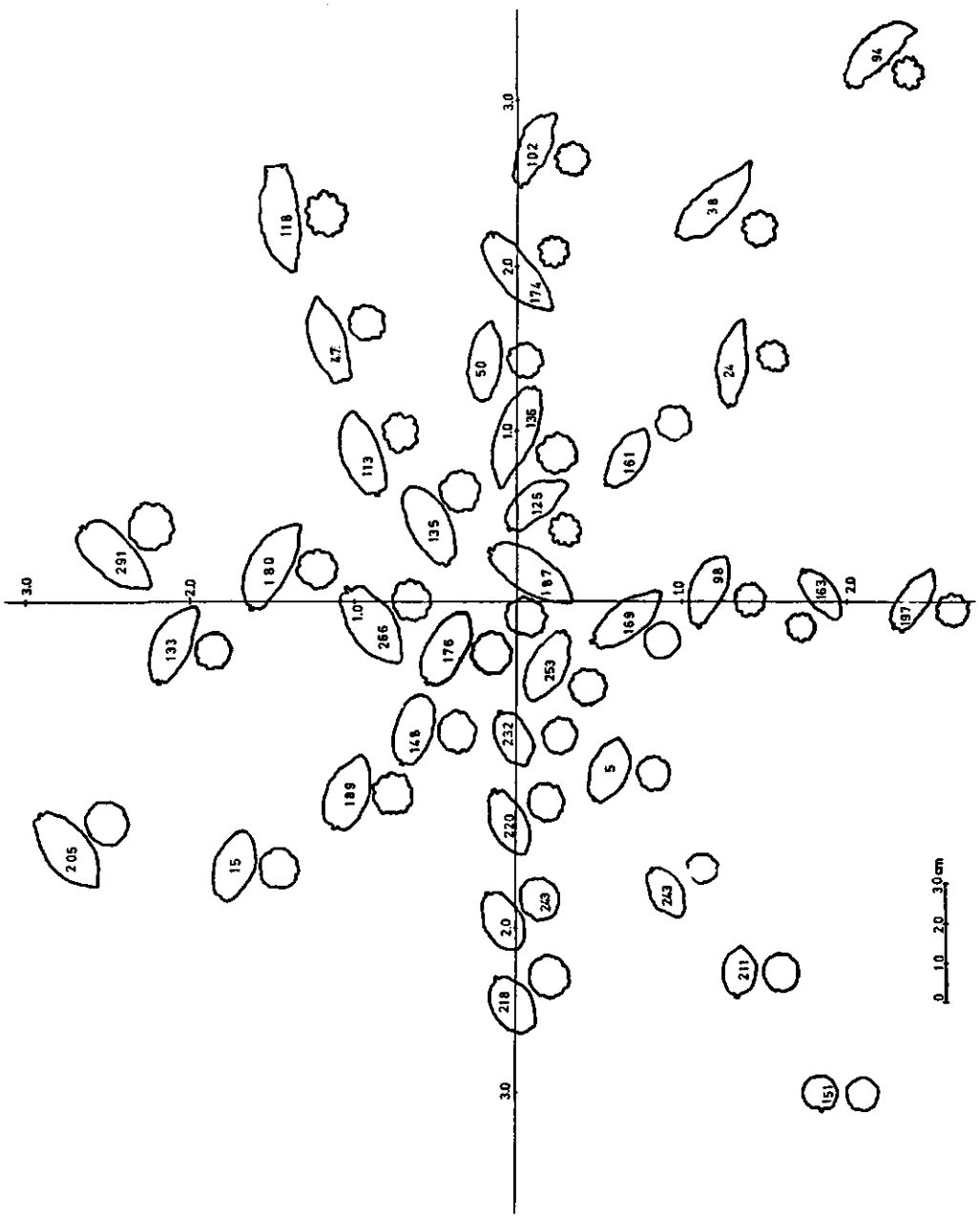


Figure 5.

Scattergram of clones randomly selected along the diagonal and main axes of Figure 2 and 3. The clones are diagramatically presented by their longitudinal and transverse fruit sections.

Table 1. Range and standard deviations of the characters evaluated in the Turrialba cacao collection.

No.	Character name	Range		Standard deviation[5]	Number of classes
		Min.	Max.		
1*	Fruit index	13.5	60.1	7.8[1)]	6
2*	Seed index (= dry seed weight)	0.55	2.23	0.34[1)]	5
3	Productivity	1[3)]	9	5.2[3)]	9
4	Maximum seed number	30	67	4.9[1)]	8
5	Wet seed weight (g)	0.92	3.39	0.24(0.28)[2)]	9
6*	Ratio dry/wet seed weight	0.50	0.77	0.04[1)]	7
7*	Seed number per fruit	10.7	45.2	8.5[2)]	4
8*	Ratio seed number/max seed number	0.27	0.88	0.08[1)]	8
9	Ratio seed number/fruit weight	0.028	0.148	0.018(0.017)[1)]	7
10	Seed length (mm)	19.6	30.0	1.4(1.5)[2)]	7
11*	Ratio seed width/length	0.36	0.64	0.04[1)]	7
12	Seed width (mm)	7.4	17.3	0.8(1.1)[2)]	9
13	Seed thickness (mm)	6.5	11.7	0.8[2)]	7
14*	Cotyledon colour	1 = white ± other colour(s)		(49)[4)]	3
		2 = light purple		(127)	
		3 = dark purple		(118)	
15	Fruit length (mm)	69	241	20.0(19.0)[2)]	9
16*	Fruit width (mm)	69	108	3.5(4.0)[2)]	9
17*	Ratio fruit width/length	0.35	1.07	0.08[1)]	9
18	Fruit weight (g)	208	1171	254.8(241.0)[2)]	4
19	Fruit wall thickness at ridges (mm)	8.2	18.6	1.3[2)]	8
20*	Fruit wall thickness at furrows (mm)	4.8	14.9	1.5[2)]	7
21	Basic fruit surface colour	1 = very light green		(58)[4)]	
		3 = light green		(79)	5
		5 = intermediate		(67)	
		7 = dark green		(86)	
		9 = very dark green		(4)	
22*	Anthocyanin intensity in ridges of ripe fruits	0	9		
23*	Ridge pair separation	0.11	0.98	0.11[1)]	8
24*	Secondary fruit furrow depth	1[3)]	9	4.4	9
25*	Fruit surface rugosity	0[3)]	9	4.3	9
26	Fruit apex form	0[3)]	9	5.1	9
27*	Basal fruit constriction	0[3)]	9	2.3	9
28	Mesocarp hardiness	3 = soft		(94)[4)]	
		5 = intermediate		(106)	
		7 = hard		(94)	3

Table 1 continued.

No.	Character name	Range		Standard deviation[5]	Number of classes
		Min.	Max.		
29*	Style length (mm)	1.35	3.14	0.17(0.20)[2]	9
30*	Ovary length (mm)	1.02	2.20	0.11(0.13)[2]	9
31	Ovary width (mm)	0.82	1.53	0.06(0.08)[2]	9
32*	Staminode length (mm)	4.00	9.13	0.40(0.57)[2]	9
33*	Sepal length (mm)	6.23	11.44	0.53(0.58)[2]	9
34*	Sepal width (mm)	1.90	2.85	0.21(0.19)[2]	5
35*	Petal length (mm)	4.93	10.46	0.54(0.61)[2]	9
36	Ligule width (mm)	1.60	3.71	0.23(0.24)[2]	9
37	Anthocyanin intensity in ligule	0[3]	9	3.3	9
38*	Anthocyanin intensity in filament	0[3]	9	2.6	9
39*	New flush colour			(53)[4]	
		1 = green		(117)	5
		3 = light brown		(77)	
		5 = brown		(40)	
		7 = dark brown		(7)	
		9 = very dark brown			

1) Calculated from the accession means.
2) Calculated as a mean of the individual standard deviations of each accession.
3) Scored on a scale from 0-9; the standard deviation value is the mean of the observations.
4) Frequencies of scored characters.
5) The values in brackets plus the underlined were calculated and the anexed listed values were actually used in the calculation of the class width.

*) Characters used in the principal component analysis.

Table 2. Characters selected for the ten characters classification, descriptor states used in the classification and the corresponding original descriptor states.

Character type	Character name	Descriptor states	
		New	Corresponding original
A. Discontinuous	1) New flush colour	1	1 (=green)
		2	2-9 (=pigmented)
	2) Anthocyanin in ripe fruit	1	0 (=unpigmented)
		2	1-9 (=pigmented)
	3) Anthocyanin in filament	1	0 (=unpigmented)
		2	1-9 (=pigmented)
B. Semi-discontinuous	4) Fruit surface rugosity	1	0,1,2,3 (=not or slightly rugose)
		2	4,5,6 (=intermediate)
		3	7,8,9 (=rugose or very rugose)
	5) Basal fruit constriction	1	0,1,2,3 (=not or slightly constricted)
		2	4,5,6 (=intermediate)
		3	7,8,9 (=strongly constricted)
	6) Secondary fruit furrow depth	1	1,2,3 (=shallow)
		2	4,5,6 (=intermedaite)
		3	7,8,9 (=deep)
C. Continuous	7) Seed number	1	1 (=low)
		2	2-3 (=intermediate)
		3	4 (=high)
	8) Ratio fruit width/length	1	1-2 (=low)
		2	3-4 (=low to intermediate)
		3	5 (=intermediate)
		4	6-7 (=intermediate to high)
		5	8-9 (=high)
	9) Seed index	1	1-2 (=low)
		2	3 (=intermediate)
		3	4-5 (=high)
	10) Style length	1	1-2 (=small)
		2	3-4 (=small to intermediate)
		3	5 (=intermediate)
		4	6-7 (=intremediate to tall)
		5	8-9 (=tall)

Table 3. The characterization of the taxonomic classes as recognized in the literature (I) and the corresponding character states as they were used in the computer search and reported here (II).

Character name and number	Taxonomic grouping											
	Criollo		Forastero		Amelonado		Angoleta		Calabacillo		Cundeamor	
	I	II	I	II	I	II	I	II	I	II	I	II[1]
Maximum seed number (4)	<3	<6	<7	-	-	-	-	-	-	-	-	-
Seed length (10)	>6	-	<4	-	<4	<5	>5	>4	-	-	-	-
Seed thickness (13)	>4	-	<4	-	<4	-	-	-	<3	<3	-	-
Cotyledon colour (14)	1or2	-	2or3	-	2or3	-	2or3	-	3	-	2or3	2or3
Ratio fruit width/length (17)	<2	<3	>3	>2	>3<6	>3<6	>2<4	-	>4	>4	<3	<3
Fruit wall thickness at ridges (19)	>3<5	-	>5	-	>6	>5	>6	>4	-	-	-	-
Anthocyanin ripe fruit (22)	0-9	-	<3	<5	-	-	-	-	-	-	-	-
Ridge pair separation (23)	>6	-	-	-	-	-	-	-	-	-	-	-
Sec. fruit furrow depth (24)	>7	>3	>3<7	<7	<3	<4	>5	>4	<3	<3	>5	>5
Fruit surface rugosity (25)	>7	>5	<3	<4	<3	<4	>7	>4	<2	<2	>3<7	>3<7
Fruit apex form (26)	>7	>5	>1<3	<4	>2<4	>2<4	-	>4	<3	<3	>7	>7
Basal fruit constriction (27)	<7	<7	<3	-	<3	<4	<2	<3	<2	<2	>3<7	>3<7
Mesocarp hardiness (28)	<3	-	>5	-	-	-	-	-	-	-	-	-
Petal length (35)	<5	-	-	-	-	-	-	-	-	-	-	-
Anthocyanin ligule (37)	<3	-	>3<7	-	-	-	-	-	-	-	-	-
New flush colour (39)	1	1,2or3	-	-	-	-	-	-	-	-	-	-

1) No changes in character states made for Cundeamor.

Table 4. Factor score coefficients and communality value for each of the characters used in the principal component analysis.

Character name and number	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Communality
Fruit index (1)	-.29116*	-.02767	.10527	.02291	-.07796	.03170	-.23902	.60485
Seed index (2)	.33109*	.00551	-.06375	-.04693	-.02032	-.04878	-.07263	.77316
Ratio dry/wet seed weight (6)	.07497	-.01539	-.08257	-.10683	.09551	.32155*	.01889	.42984
Seed number per fruit (7)	-.05007	.01067	.00531	-.01137	.01347	.13695	.48215*	.73276
Ratio seed number/max seed number (8)	.05665	-.02950	-.05164	.01328	-.03876	-.13088	.56334*	.76151
Ratio seed width/length (11)	.24973*	.05972	-.11540	-.05179	.15042	-.35254	.00796	.62871
Cotyledon colour (14)	-.05665	.01373	.01469	.24061	.44935*	.11415	.03211	.64793
Fruit width (16)	.24658*	-.13034	.00794	-.02596	-.01214	.23631	-.04583	.71676
Ratio fruit width/length (17)	.01251	-.25894*	.01808	.01051	.24728	-.06279	-.09943	.67494
Fruit wall thickness at furrows (20)	.17001*	.08530	-.00906	.09727	-.15389	.20910	-.07034	.55736
Anthocyanin intensity in ridges of ripe fruits (22)	-.02669	-.01917	-.00582	.50342*	-.03461	.08744	.02044	.68999
Ridge pair separation (23)	.00987	-.22620*	.01134	-.06321	-.16282	-.00289	.05457	.36046
Secondary fruit furrow depth (24)	.00145	.28800*	.00080	.02954	.06747	.07870	-.02818	.53365
Fruit surface rugosity (25)	.00065	.30126*	.02403	-.04990	-.08217	.03166	-.03721	.62742
Basal fruit constriction (27)	-.03580	.25082*	.02141	-.06218	.05084	-.05893	.02364	.39084
Style length (29)	-.00299	.10386	-.03137	.05923	.00655	.46174*	-.02411	.56719
Ovary length (30)	-.12188	-.03475	.44964*	.07618	.16803	-.06076	-.08033	.56089
Staminode length (32)	.02304	-.00360	.41464*	-.03314	-.08018	-.13826	.07252	.69582
Sepal length (33)	.15148	.06410	.25288*	.00323	.10814	-.05360	-.05307	.57260
Sepal width (34)	-.04897	.03157	.39069*	-.13954	-.05547	.07956	-.03021	.46698
Anthocyanin intensity in filament (38)	.09685	.05289	.03193	-.19819	.46586*	-.06499	-.03215	.69015
New flush colour (39)	-.01962	.00479	-.07119	.49821*	.11379	-.06715	-.00934	.64769

*) The characters marked with an asterisk in each column constitute the corresponding factor.

Table 5. Results of the computer searches using the taxonomic parameters for the different classes from the literature.

	Number of cultivars			
	Computer		From the literature	Correspondence between literature and search II (%)
	Search I[1)]	Search II[2)]		
Criollo (including hybrids)	0	46	35	28.6
Forastero	0	90	70	54.2
Amelonado	0	2	35	0
Angoleta	0	31	10	20.0
Calabacillo	9	17	2	0
Cundeamor	10	-	11	0

1) The average of the descriptor states as given in the literature were used (see Table 3, columns I).
2) The less restrictive descriptor states as presented in Table 3, column II, were used.

Table 6. Comparison of some pedigree groups with genetic groups of cacao cultivars in a principal component analysis.

Pedigree	Corresponding genetic group in scattergram	Remarks
Criollo hybrids	Criollo	complete matching
Comun	Forastero	complete matching
Nacional	Criollo	does not agree with Soria (1970a)
Matina	Trinitario	no strong cluster
SCA-6 and SCA-12	none	rather distinct group
ICS	Forastero	compact cluster around origin

Table 7. Results of the ten characters classification of 294 cacao clones using the characters presented in Table 2.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
Character state	1*)	1	1	1	1	1	2	1	1	2	77
										4	56
								2	1	2	10
										3	234, 243
										4	7, 8
								3	1	2	9, 233
										3	139
										4	211
							3	1	2	-**)	179
									3	-	91
				2	1	1	2	-	-	-	60
							3	-	-	-	192
						2	2	1	1	-	172
									3	-	136
									1	2	93
										4	204
					2	-	-	-	-	-	147
				3	1	2	2	1	1	-	70
									2	-	188
						3	-	-	-	-	174
					2	-	-	-	-	-	167
					3	-	-	-	-	-	94
			2	1	1	1	2	1	1	-	173
									3	-	228
									1	3	227, 242
										4	41, 225, 238
								4	-	-	100
							3	2	1	4	99, 148
						2	2	1	-	-	175
									1	2	215
										3	279
							3	1	-	-	119
								2	-	-	240
				2	1	1	2	-	-	-	59
							3	-	-	-	241
						2	2	1	1	-	109
									2	-	288
								2	-	-	135
							3	-	-	-	124
					2	-	-	-	-	-	62
				3	-	-	-	-	-	-	247

Table 7. continued.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
		2	1	1	-	-	-	-	-	-	129
				2	-	-	-	-	-	-	138
			2	1	-	-	-	-	-	-	134
				2	-	-	-	-	-	-	152
	2	1	1	1	1	1	2	1	1	3	20, 212
										4	69, 81
									3	2	189, 250, 251
								2	1	1	208
										2	40, 96, 214
										3	2
										4	170
									2	-	49
									3	2	52
										3	272
							3	2	1	3	92, 229
						2	2	1	1	-	64
									2	-	141
									3	2	249
										3	133
								2	1	3	171
										4	30
									3	2	257
										4	36
							3	1	1	3	75, 202
										4	66
									3	-	256
					2	1	2	1	2	4	207, 269
								2	1	2	155
										4	5
						2	-	-	-	-	161
				2	1	1	1	-	-	-	274
							2	1	1	-	65
									2	3	182, 190
										4	184
						2	2	1	1	1	197
										2	120, 164, 198
										3	127
										4	68, 78, 117
									2	2	50
										3	191
									3	2	291

Table 7. continued.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
										3	181
								2	1	-	287
									2	-	142
									3	-	275
							3	-	-	-	76
						3	3	1	-	-	63
								2	-	-	248
					2	1	2	-	-	-	185
							3	-	-	-	294
						2	2	1	-	-	183
								2	-	-	165
						3	-	-	-	-	44
				3	1	1	-	-	-	-	186
						2	2	1	1	-	74
									3	-	160
							3	-	-	-	24
					2	-	-	-	-	-	38
					3	-	-	-	-	-	292
			2	1	1	1	2	1	1	1	32
										2	4
										3	168
										4	203
									2	3	35
										4	293
									3	-	12
								2	1	2	199, 221, 281
										3	1, 103, 216
										4	97, 169
									2	2	15, 43, 51
									3	2	42, 48
										4	205
								3	1	-	219
									2	-	222
								5	-	-	151
							3	1	1	4	98, 149, 162, 176, 177
								2	1	2	195, 218, 220
										3	230, 231, 232
										4	239
								3	1	4	223, 224
						2	2	1	1	3	122, 163
									2	4	116

Table 7. continued.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
										5	112
									3	-	13
								2	1	1	150
										2	22, 28, 156, 166
										3	23
										4	29
									2	2	86, 101
										3	146
							3	1	1	3	67
										5	125
									2	-	206
								2	1	2	19, 194
										3	21, 253
									2	2	84, 217
						3	2	-	-	-	87
							3	-	-	-	88
					2	1	2	-	-	-	6
							3	-	-	-	154
						2	2	1	1	4	11, 82
									2	-	55
						3	-	-	-	-	16
				2	1	1	2	1	2	2	14, 53
										3	236
									3	-	180
							3	1	1	-	73
									2	-	237
						2	2	1	1	3	61, 158
									2	2	282
										3	25, 290
										4	105, 106, 113
										5	115
								2	1	-	111
									2	4	280
										5	107
							3	1	1	2	258
										3	58
										4	178
									3	-	108
								2	-	-	226
						3	2	-	-	-	83
							3	-	-	-	85

Table 7. continued.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
					2	1	2	-	-	-	210
							3	-	-	-	209
						2	2	1	2	4	260
										5	104
								2	2	-	265
									3	-	45
							3	1	-	-	3
								2	-	-	254
						3	-	-	-	-	118
					3	1	-	-	-	-	47
						3	-	-	-	-	102
				3	1	-	-	-	-	-	57
					2	1	-	-	-	-	54
						2	2	1	-	-	34
								2	-	-	39
							3	-	-	-	126
						3	2	-	-	-	17
							3	-	-	-	123
	2	1	1	1	1	2	1	1	3		79, 90, 271
									3	3	143, 267
										5	31
								2	2	2	46, 278
										4	289
							3	-	-	-	71
						2	2	1	1	-	273
									2	2	201
										4	266
									3	-	132
								2	-	-	128
							3	1	1	3	246
										4	245
						3	2	1	-	-	27
								2	-	-	252
							3	-	-	-	89
					2	1	-	-	-	-	263
						2	-	-	-	-	33
					3	-	-	-	-	-	121
				2	1	1	2	1	2	3	187, 213, 283
									3	2	159
										4	193, 235
								2	-	-	276

Table 7. continued.

Character number	1	2	3	4	5	6	7	8	9	10	Corresponding cultivars
							3	1	1	1	196
										3	72
						2	2	1	1	-	145
									2	2	18, 200
										4	140
									3	2	268
										3	270
										4	130
							3	1	1	2	244
										3	95
									2	-	285
									3	-	264
					2	1	-	-	-	-	262
						2	-	-	-	-	80
				3	2	-	-	-	-	-	153
					3	-	-	-	-	-	37
			2	1	1	1	2	1	1	2	137
										3	26
								2	-	-	255
							3	-	-	-	157
						2	-	-	-	-	259
						3	-	-	-	-	110
					2	-	-	-	-	-	284
				2	1	1	-	-	-	-	261
						2	2	1	3	3	144, 277
										4	131
					3	1	-	-	-	-	114
						2	-	-	-	-	286

*) The character states are only given when they differ from the state(s) of the previous combination.

**) This character was not anymore needed to keyout the cultivar.

Table 8. Comparison of methods to identify duplicates.

possible duplicate cultivars	subjective comparison of states[1)]	computer search (± 2 s.d.)	principal component analysis	classification in Table 7	cluster analysis	grand total of states
7 - 8	yes	yes	yes	yes	no	yes
57 - 58	yes	no	yes	yes	possible	doubtful
163 - 164	no	no	yes	possible	yes	yes
168 - 169	yes	yes	yes	yes	no	yes
171 - 172	no	no	no	no	no	yes
175 - 176	no	no	no	no	no	yes
177 - 178	yes	yes	yes	no	yes	no
195 - 196	no	no	possible	no	possible	no
Relative efficiency[2)]	-	1 1/2	2	2	3 1/2	4

1) If the difference in number of classes for each of the more reliable characters was three or less the two cultivars were considered to be duplicates.

2) The result for each cultivar pair and method is compared with the result of the subjective observation. A value of 1 was assigned if the results did not agree (e.g. yes and no) and 1/2 when a possible or doubtful observation was recorded compared to "yes" or "no". The lower the number the better the method matches with the subjective observation.

GENERAL DISCUSSION

The collection, conservation, evaluation and utilization of plant genetic resources have undergone rapid development worldwide. Besides several regional genebanks and major crop germplasm holdings at the International Agricultural Research Centres, many national programmes have been established to conserve plant genetic resources (Anonymous, 1985). Initially, the highest priority was given to collection and conservation of germplasm followed by characterization of the collections. This latter work has been supported by the publication of descriptor lists. Only recently more attention has been given to the evaluation of the accessions in order to facilitate their use in breeding programmes.

In this study the importance to plant breeders of the systematic description of a germplasm collection has been investigated. Particular attention has been paid to the use of quantitatively inherited characters. The type and number of characters which should be used in taxonomic analyses of a group of taxonomic units (e.g. accessions) will be discussed as well as those aspects which are closely related to the use of quantitative characters in taxonomic work. The discriminative power and the genetic background of the characters chosen was investigated and the implications for plant breeders will be discussed. Finally, the use of the results of the characters in taxonomic classifications of cacao clones (and populations) and the identification of unknown or mislabeled clones will be discussed and conclusions drawn.

The development of the cacao descriptor list

The rationale for having a descriptor list for a germplasm collection is presented (Chapter 2) along with background information on the development of the list. Since cacao was a crop entirely unknown to the author, and no descriptor lists had been published, it was necessary to study all available information on potentially useful characters as well as literature on taxonomy, botany, agronomy, genetics, plant breeding, diseases, etc. A first draft of the descriptor list was prepared after further information was acquired and discussions were held with plant breeders, botanists and farmers. Subsequent field observations and data recording was done to obtain information about the variation of a given

character within trees and between trees within a clone, and between clones, as well as about the ease of recording the phenotypic expression. In addition to these criteria, it was proposed to consider four further aspects in the selection of descriptors (Engels, 1980b): their heritability, taxonomic value, agronomic value, and their pure scientific value. To be included in the descriptor list a character should have low variation within trees and between trees of the same clone, high variation between clones and should be easily recorded.

Several of these criteria may be associated with each other (e.g. high taxonomic value, high heritability, and low variation within clones), and if such characters do not fulfil additional other criteria (e.g. high agronomic value or a high variation between clones) their selection could be at stake. In this study special attention was given to the agronomic importance of a character which would be of value to the potential users of a germplasm collection, plant breeders and agronomists.

Since the variation of a character within a tree and between trees within a clone will have a direct effect on the sample size to be taken, the variations were calculated separately for each non-qualitative character for a series of different clones. After analyzing the variation within trees, which was found to be non-significant, it was decided to sample the organs within trees at random and, therefore, 'within' tree variation was added to the 'between' tree variation. In case of obvious qualitative characters it was recommended to make observations on at least five trees in order to be able to assure that the given clone is not a mixture or a population. If the latter is the case this should be indicated.

Finally, a clear definition of each descriptor has to be given as well as how they have to be measured, including the growth stage, the precise location of the measurement and the sample size (Engels, 1980; Howes, 1981). Each descriptor should represent only one characteristic. In the definition of the descriptor states the discontinuously expressed characters will not cause problems. Where appropriate, drawings should be included and in case of colours the corresponding code of a colour chart should be recorded. In case of quantitatively expressed characters the state should be expressed either in metric units or in an arbitrary scale from 1 to 9 (Engels, 1980a).

The list developed in Chapter 2 can be considered as a 'maximum' list and its descriptors can be used as required, as was done in the systematic description of the CATIE cacao collection. The data from this systematic description has been published in a catalogue (Engels, 1981) and formed the basis for some analyses presented in Chapters 3, 4, 5 and 8. The catalogue also included a few computer searches of relevant combinations of desirable agronomic characters to facilitate initial selection of accessions by plant breeders.

The selection of characters for the systematic description

Mayr (1984) discerns three basic controversies regarding taxonomic characters:

1) should a worker use only one key character, several or all possible characters?
2) should one accept only morphological or also ecological, physiological and other characters?
3) should one weight characters or not - and if so, according to which criteria?

The answer to these questions will depend on the purpose for which the characters are required. In this study characterization and preliminary evaluation are the main targets. If a similarity study of taxa is required the number of characters used is of importance and recommendations vary from three to six or ten quantitative characters. Where taxonomic units are fairly closely related (Blackith and Reyment, 1971) up to 60-100 are advised (Crovello, 1969; Sokal and Sneath, 1963). In such a similarity study numerical taxonomic methods are applicable and as many different types of characters as possible should be used, all with an equal weight (Sneath and Sokal, 1973). It was for this reason that quantitative characters were used in this study. If, however, a special purpose or an artificial classification of cultivars is wanted based on fruit, seed or vegetative characters, only a few qualitative characters suffice for a useful cultivar classification (Hawkes, 1970).

The selection of characters should be done between the two extremes: general purpose classifications showing phenetic or even phylogenetic relationships between taxa, and classifications which are of direct use to the plant breeder. In Table 1 it can be observed that only in a few instances a

character possesses both agronomic and taxonomic importance. Therefore, the use of both types of characters in a classification would increase its general applicability.

To obtain more background information on the taxonomic values of characters, the discriminative values of qualitative and quantitative characters were calculated (Chapters 3 and 4) and some of the results are summarized (Table 1). Although not for all characters complete data sets could be obtained, because of the addition of such characters during the course of this study, consistency regarding the various aspects for each character can be observed from Table 1.

The assumption that a high discriminative power will be positively correlated with the heritability of a character could only partly be confirmed (Chapters 6 and 7). It was not possible to estimate the number of genes involved in the expression of the quantitative characters; only an indirect value for the heritability expressed in the repeatability, could be calculated. Although the values for h^2_w in Table 1 are only of relative reliability due to the difficulties of estimating the environmental variances, Spearman's coefficient of rank correlation ($r = 0.62$, $P < 0.01$) between h^2_w and the 'D' discriminative values show that there is a biological basis for the discriminative value, as was expected. Spearman's coefficient of rank correlation between h^2_w and the repeatability values was not significant and this indicates the relative reliability of h^2_w, although for both calculations different clones and procedures were used.

As was noted in Chapters 3, 4 and 5 the calculated value of the discriminative power of a character depends on the genetic relationships between the clones studied, the number of clones involved in the study as well as the number of classes per character. Therefore, the values for a given set of characters will be only comparable if they have been calculated from the same set of accessions.

A final set of characters for the characterization and preliminary evaluation of cacao clones can thus be recommended, based on the the discriminative values, genetic background and the taxonomic and agronomic importance. The final set should include at least the descriptors indicated in Table 1 with a mark and could form a minimum list. Further descriptors could be included, preferably from the IBPGR

descriptor list (IBPGR, 1981), according to the needs of the potential users. A combination of qualitative and quantitative characters was then selected following the suggestions of Higgins et al. (1981) to form a special classification for the identification of clones. This was then tested and found to be useful.

Some implications for cacao breeding

Systematic description leads to more efficient use of germplasm in a collection (Chang, 1976) and it will enable a breeder to make an initial selection of genotypes/phenotypes and hence, it will increase the efficiency of his activities. In the present study special attention was paid to the relationships between characters (Chapters 5 and 8) and based on the results conclusions were drawn for cacao breeding.

Although a systematic description as described here did not allow any direct conclusions on the genetic background of the characters, it can facilitate studies which are designed for such a purpose (Chapters 6 and 7). In the case of cacao it was felt to be very important to investigate the genetic background of some of the characters since such information was only partly available for only a few traits. The results of the analysis of a diallel cross of seven clones (Chapter 6) indicated that the lack of information may, at least partly be due to cacao being a tetraploid, and not a diploid species as has been previously assumed. This conclusion is supported by Allen (1985, personal communication), who, assuming tetraploid inheritance, was able to explain segregations for cotyledon colour in wild Ecuadorian populations.

Some implications for cacao clone classification and identification

As the identification of new accessions based on taxonomic criteria is a prerequisite for rational utilization of germplasm (Hawkes, 1970 and 1979; Lehmann, 1972) the applicability of existing classifications of cacao clones (and populations) was studied (Chapter 8) and it was concluded that a more flexible classification of clones was needed. Although almost 65 percent of the clones could be distinguished individually by using 10 well selected descriptors, it would be more appropriate to use fewer descriptors and to form larger

groups of clones with some characteristics in common. Such an informal classification would be of great value to breeders and other users of germplasm since it facilitates communication and selection of desired phenotypes.

The study of overall similarities between accessions (Chapters 5 and 8) indicated that phenetically similar accessions also possessed genetic similarities as was also found by Sokal and Sneath (1963) for other species. Thus numerical taxonomic studies are of direct interest to the plant breeder and can be successfully employed in genebanks for initial identification of unknown accessions as well as to identify possible duplicates, essential steps toward a rational germplasm conservation (Jain, 1979). As the number and diversity of characters employed in similarity studies of germplasm affect the outcome, quantitative characters should be included in such studies. This will augment the diversity of the employed characters as well as the total number. This is particularly important when germplasm accessions are grouped into races (Holcomb et al., 1977).

A systematic description enables identification of duplicates and facilitates the detection of errors. Two different approaches for error detection have been proposed, the first method employs only highly heritable characters and is based on the principle of matching (Chang, 1976), and the second method uses frequencies of the descriptor states (Porter and Smith, 1982).

REFERENCES

ANONYMOUS, 1981. Genetic resources of cacao. International Board for Plant Genetic Resources, Rome. 25pp.

ANONYMOUS, 1985. Annual report 1984. International Board for Plant Genetic Resources, Rome.

BLACKITH, R.E. and R.A. REYMENT, 1971. Multivariate morphometrics. Academic Press, London and New York. 412pp.

CHANG, T.T., 1976. Manual on genetic conservation of rice germ plasm for evaluation and utilization. International Rice Research Institute, Los Banos, Philippines. 77pp.

CROVELLO, T.J., 1969. Effects of change of characters and of number of characters in numerical taxonomy. The American Midland Naturalist 8: 68-86.

ENGELS, J.M.M., 1980a. Sistemas de informacion para centros de recursos geneticos. In: Proceedings of the International Course on Genetic Resources CATIE/GTZ. CATIE, Turrialba, Costa Rica (mimeographed). 46pp.

ENGELS, J.M.M., 1980b. Descripcion sistematica de colecciones de germoplasma. In: Proceedings of the International Course on Genetic Resources CATIE/GTZ. CATIE, Turrialba, Costa Rica (mimeographed). 19pp.

ENGELS, J.M.M., 1981. Genetic resources of cacao: a catalogue of the CATIE collection. Technical series. Technical bulletin /CATIE: no. 7. CATIE, Turrialba, Costa Rica. 196pp.

HAWKES, J.G., 1970. The taxonomy of cultivated plants. In: Frankel, O.H. and E. Bennett, eds. Genetic resources in plants. Oxford, Blackwell. p. 69-85.

HAWKES, J.G., 1979. Genetic poverty of the potato in Europe. In: Zeven, A.C. & A.M Harten, van, eds. Broadening the genetic base of crops. Pudoc, Wageningen. p. 19-27.

HIGGINS, J., J.L. EVANS and P.J. RED, 1981. Classification of Western European cultivars of Vicia faba L. J. natn. Inst. agric. Bot. 15: 480-487.

HOLCOMB, J., D.M. TOLBERT and S.K. JAIN, 1977. A diversity analysis of genetic resources in rice. Euphytica 26: 441-450.

HOWES, C., 1981. Guidelines for developing descriptor lists. Plant Genetic Resources Newsletter 45: 26-32.

JAIN, S.K., 1979. Biosystematic studies of populations in germ plasm collections. In: Zeven, A.C. and A.M. Harten, van, eds. Broadening the genetic base of crops. Pudoc, Wageningen. p. 103-110.

LEHMANN, C.O., 1972. Das Kulturpflanzenweltsortiment des Zentralinstitutes fuer Genetik und Kulturpflanzenforschung in Gatersleben, ein Genreservoir fuer die Pflanzenzuechtung in der DDR. Arch. Zuechtungsforschung, Band 2, Heft 2. p. 109-127.

MAYR, E., 1984. Die Entwicklung der biologischen Gedankenwelt: Vielfalt, Evolution und Vererbung. Springer-Verlag, Berlin. 766pp.

PORTER, W.M. and D.H. SMITH, 1982. Detection of identification errors in germplasm collections. Crop Science 22: 701-703.

SNEATH, P.H.A. and R.R. SOKAL, 1973. Numerical taxonomy. The principles and practice of numerical classification. W.H. Freeman and Co., San Francisco. 573pp.

SOKAL, R.R. and P.H.A. SNEATH, 1963. Principles of numerical taxonomy. W.H. Freeman and Co., San Francisco. 359pp.

Table 1. Summary of the most important attributes measured and/or calculated for each of the selected characters.

Character name	Attribute	CV between 294 clones	Type of descriptor states	Average CV of 294 clones	Average CV of 32 clones Chapter 2	Discriminative value from Chapter 2 & 3	Taxonomic importance b)	Agronomic importance b)	h^2_w c)	Number of genes involved	Repeatability (Chapter 6)	Characters for taxonomic key in order of importance	Descriptors recommended for systematic description
1	Fruit index	29	continuous	-[a]	-	-	1	9	25.8[d]		0.37		x
2	Seed index (= dry seed weight)	28	"	14	18	0.62	3	9	55.5[d]		0.34	9	x
3	Productivity	32	1-9 scale[b]	-	-	-	1	9			0.13		
4	Maximum seed number	11	continuous	-	-	-	3	9			0.52		x
5	Wet seed weight (g)	28	"	12	17	0.67	3	9	58.3[d]		0.35		
6	Ratio dry/wet seed weight	6	"	-	10	0.35	1	7	23.5[e]				
7	Seed number per fruit	12	"	25	31	0.49	7	7	20.4[d]		0.49	7	x
8	Ratio seed number/max. seed number	11	"	-	-	-	3	7					
9	Ratio seed number /fruit weight	28	"	-	30	0.72	1	5	50.1[e]				
10	Seed length (mm)	9	"	6	6	0.79	1	9	73.6[d]				x
11	Ratio seed width/seed length	8	"	-	6	0.41	5	5	34.2[e]				x
12	Seed width (mm)	13	"	6	6	0.78	1	9	70.0[d]				
13	Seed thickness (mm)	13	"	9	10	0.53	1	9					
14	Cotyledon colour	-	discrete	-	-	0.31	7	5		1			x
15	Fruit length (mm)	15	continuous	12	12	0.57	5	7	60.3[d]		0.57		x
16	Fruit width (mm)	8	"	9	9	0.46	5	7	26.5[d]		0.38		
17	Ratio fruit width/length	16	"	-	10	0.59	5	5	52.5[e]			8	x
18	Fruit weight (g)	8	"	26	30	0.48	1	7	37.2[d]				x
19	Fruit wall thickness at ridges (mm)	16	"	12	13	0.45	1	5	45.3[d]		0.17		
20	Fruit wall thickness at furrows (mm)	18	"	15	15	0.59	1	5	38.7[f]				x
21	Basic fruit surface colour	-	discrete	-	-	0.72	5	1		1			
22	Anthocyanin intensity ridges of ripe fruits	-	1-9 scale	-	-	0.61	9	1		1		2	x
23	Ridge pair separation	15	continuous	-	-	-	7	1					
24	Secondary fruit furrow depth	41	1-9 scale	-	-	0.67	9	3				6	x
25	Fruit surface rugosity	44	"	-	-	0.56	9	1				4	x
26	Fruit apex form	39	"	-	-	0.74	9	1		1			x
27	Basal fruit constriction	100	"	-	-	0.54	9	1				5	x
28	Mesocarp hardiness	33	"	-	-	-	3	3					
29	Style length (mm)	12	continuous	7	7	0.81	3	1	81.1[d]			10	x
30	Ovary length (mm)	12	"	8	9	0.69	3	1	54.6[d]				
31	Ovary width (mm)	7	"	6	7	0.33	3	1	20.3[d]				
32	Staminode length (mm)	9	"	5	6	0.66	3	1					
33	Sepal length (mm)	9	"	6	6	0.62	3	1	60.8[d]				x
34	Sepal width (mm)	9	"	9	11	0.39	3	1	43.5[d]				
35	Petal length (mm)	11	"	7	7	0.58	3	1	55.6[d]				
36	Ligule width (mm) (= petal width)	14	"	9	10	0.60	3	1	75.1[d]				
37	Anthocyanin intensity in ligule	-	1-9 scale	-	-	0.18	7	1	70.2[d]				
38	Anthocyanin intensity in filament	-	"	-	-	0.18	7	1				3	x
39	New flush colour	-	discrete	-	-	0.83	9	1		2		1	x

a) - = not calculated
b) 1 = very low; 5 = intermediate; 9 = very high
3 = low 7 = high
c) $h^2_w = \frac{\text{var p} - \text{var e}}{\text{var p}}$

d = h^2_w based on 32 clones
e = " based on 20 clones
f = " based on 12 clones

Appendix 1. Cultivars used in the study.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
1	10091	-0.42	-1.09	6	APA-4		
2	10092	0.02	-1.66	6	APA-5		
3	10102	0.08	0.61	10	BE-10	FO,AM	
4	10103	-0.43	-1.02	3	BS-2		
5	10108	-0.54	-1.00	10	CAS-1	AM	Comun
6	10109	0.03	-0.73	10	CAS-3	AM	Comun
7	10113	-0.32	-1.50	10	CATONGO BLANCO	FO,AM	Comun
8	10114	-0.03	-1.60	10	CATONGO BLANCO	FO,AM	Comun
9	10125	0.03	-2.02	10	CATONGO-O	FO,AM	Comun
10	10128	-0.32	-2.27	5	CATONGO X SELF	FO,AM	Comun
11	10132	-0.74	0.97	5	CC-9	TR,CU	Matina x TR
12	10133	0.99	0.27	5	CC-10	TR,AN	Matina x TR
13	10134	1.34	0.66	5	CC-17	AM	Matina o.p.
14	10135	0.94	0.05	5	CC-18	AM	Matina o.p.
15	10136	1.63	-1.48	5	CC-27	AM	Matina o.p.
16	10139	0.52	1.52	5	CC-34		Matina o.p.
17	10140	1.06	1.85	5	CC-35	TR	UF-650 o.p.
18	10142	0.04	0.25	5	CC-37	TR	UF-667 o.p.
19	10143	-0.39	0.06	5	CC-38		Matina x UF-676
20	10144	-0.32	-0.34	5	CC-39	TR	UF-668 o.p.
21	10145	-0.25	-0.05	5	CC-40		UF-675 o.p.
22	10146	0.01	0.25	5	CC-41	TR,AM	UF-676 o.p.
23	10147	-0.53	-0.47	5	CC-42	TR	UF-676 o.p.
24	10148	-1.49	1.34	5	CC-43	TR	UF-667 o.p.
25	10149	1.26	0.77	5	CC-44		UF-654 o.p.
26	10150	-0.90	-0.04	5	CC-45		UF-654 o.p.
27	10151	0.74	0.68	5	CC-46		UF-654 o.p.
28	10152	-0.33	0.07	5	CC-47		UF-654 o.p.
29	10153	-0.07	0.52	5	CC-48	TR	UF-670 o.p.
30	10154	-0.20	-0.20	5	CC-49	TR	UF-676 o.p.
31	10155	1.23	-0.62	5	CC-52	TR	UF-613 o.p.
32	10158	-0.18	-0.13	5	CC-67		
33	10159	1.62	0.77	5	CC-69		
34	10160	-0.10	1.34	5	CC-71		
35	10161	0.09	0.54	5	CC-74		UF-677 o.p.

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
36	10162	1.54	-0.44	5	CC-79	TR	UF-613 o.p.
37	10163	-0.71	1.94	5	CC-83	TR	UF-613 o.p.
38	10164	-1.18	2.30	5	CC-99		
39	10165	0.79	1.64	5	CC-100	TR	UF-667 o.p.
40	10166	0.71	-1.38	5	CC-102	TR	UF-667 o.p.
41	10167	0.25	-0.16	5	CC-103	TR	UF-667 o.p.
42	10168	0.92	-0.53	5	CC-106	TR	UF-221 o.p.
43	10169	0.51	-0.71	5	CC-107	TR	UF-221 o.p.
44	10170	0.95	1.20	5	CC-120	TR	UF-650 o.p.
45	10171	1.38	0.82	5	CC-121	TR	UF-650 o.p.
46	10172	0.13	-0.77	5	CC-124	TR	UF-613 o.p.
47	10175	1.16	1.77	5	CC-132	TR	UF-650 o.p.
48	10177	0.89	-0.55	5	CC-137	TR	UF-12 o.p.
49	10178	-0.09	-0.55	5	CC-138	TR	UF-12 o.p.
50	10179	0.26	1.25	5	CC-139	TR	UF-12 o.p.
51	10180	1.23	-0.37	5	CC-143		UF-672 o.p.
52	10181	0.41	-0.44	5	CC-144	TR	UF-676 o.p.
53	10182	0.78	0.82	5	CC-152	TR	UF-12 o.p.
54	10184	0.40	1.44	5	CC-169		UF-677 o.p.
55	10185	0.96	0.16	5	CC-173	TR	UF-650 o.p.
56	10186	-0.90	0.07	5	CC-210		SCA-12 o.p.
57	10188	-1.31	0.93	5	CC-211		IMC-67 o.p.
58	10189	-1.52	1.06	5	CC-211		IMC-67 o.p.
59	10190	1.19	0.85	5	CC-212		P-150 o.p.
60	10191	-1.35	0.53	5	CC-212		P-150 o.p.
61	10192	-1.83	0.16	5	CC-213		IMC-67 o.p.
62	10193	-0.20	1.31	5	CC-213		IMC-67 o.p.
63	10195	-1.51	1.00	5	CC-215		
64	10201	-1.05	-0.11	5	CC-224		
65	10202	-2.33	0.49	5	CC-225		
66	10203	-0.48	0.32	5	CC-226		
67	10204	-0.75	-0.05	5	CC-228		SCA-12x ICS-6
68	10205	-0.75	0.56	5	CC-231		SCA-? x ICS-6
69	10206	-1.18	-0.41	5	CC-232		
70	10207	-0.91	1.15	5	CC-234		SCA-6 x ICS-39

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
71	10208	-1.76	-0.23	5	CC-235		ICS-1 x SCA-6
72	10209	-0.32	0.02	5	CC-236		
73	10211	-0.92	-0.09	5	CC-240		
74	10212	-0.65	0.51	5	CC-241		
75	10214	-0.73	0.17	5	CC-244		SCA-6 X UF-667
76	10215	-0.66	1.11	5	CC-245		
77	10217	-1.30	-0.17	5	CC-249		PA-12 o.p.
78	10219	-0.37	0.73	5	CC-251		
79	10221	-1.22	0.01	5	CC-253		
80	10222	-1.39	1.05	5	CC-254		
81	10224	-1.22	-0.68	5	CC-256		
82	10225	0.18	0.49	5	CC-257		
83	10226	-0.02	0.92	8	CC-258		NAC-2 o.p.
84	10227	0.98	0.52	5	CC-259		NAC-3 o.p.
85	10228	1.19	0.81	5	CC-260		NAC-3 o.p.
86	10229	0.57	-0.04	5	CC-261		NAC-3 o.p.
87	10230	0.23	0.87	5	CC-262		NAC-3 o.p.
88	10231	-0.16	0.34	5	CC-263		NAC-2 o.p.
89	10232	0.07	0.81	5	CC-264		
90	10234	0.79	-1.16	7	CHUAO-120	CR(?)	
91	10236	1.50	-0.31	7	CNS-23	CR(?)	
92	10237	-0.83	-1.12	10	COMUN TIPICO	FO	
93	10243	-2.74	0.42	10	C.SUL-3	FO	
94	10245	-2.70	3.68	10	C.SUL-7	FO	
95	10251	-0.29	0.82	5	DIAMANTES-800		ICS-1 x SCA-6
96	10254	-1.29	-0.56	10	EEG-25	FO	
97	10255	-1.42	-0.27	10	EEG-27	FO	
98	10256	-1.40	0.04	10	EEG-29	FO	
99	10257	-0.33	-1.22	10	EEG-48	FO	
100	10258	-0.59	-1.64	10	EEG-64	FO	
101	10259	0.67	-0.85	10	EEG-65	FO	
102	10260	-0.51	2.97	8	EET-12	CR	
103	10262	-0.33	-0.71	8	EET-41		
104	10264	-0.05	0.71	8	EET-48		NACIONAL HYBRID
105	10265	1.37	0.40	8	EET-48		NACIONAL HYBRID

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
106	10266	0.54	0.87	8	EET-59		NACIONAL HYBRID
107	10267	0.70	0.24	8	EET-62		NACIONAL HYBRID
108	10268	1.56	0.99	8	EET-64		NACIONAL HYBRID
109	10269	-0.27	1.21	8	EET-67		VENEZUELAN YELLOW
110	10270	-0.55	0.47	8	EET-75		VENEZUELAN YELLOW x RED
111	10271	-0.21	0.28	8	EET-80		VENEZUELAN YELLOW
112	10272	1.53	0.31	8	EET-94		VENEZUELAN YELLOW
113	10273	0.90	0.93	8	EET-95		NACIONAL HYBRID
114	10275	1.35	1.19	8	EET-156		NACIONAL HYBRID
115	10276	0.53	1.37	8	EET-162(?)[3)]		NACIONAL HYBRID
116	10277	1.04	0.43	8	EET-164		NACIONAL HYBRID
117	10278	0.23	0.86	8	EET-183		NACIONAL HYBRID
118	10279	1.25	1.88	8	EET-228		NACIONAL HYBRID
119	10280	-0.07	0.08	8	EET-250		VENEZUELAN YELLOW
120	10281	-0.27	0.41	8	EET-333	FO	
121	10282	-0.58	1.34	8	EET-338		
122	10283	0.35	-0.27	8	EET-353		
123	10285	-0.57	1.68	8	EET-376		EET-156 x SCA-6
124	10286	-0.40	0.76	8	EET-377		SCA-6 x EET-156
125	10288	-0.05	0.13	8	EET-397(?)		SCA-6 x unknown
126	10289	-0.82	2.38	8	EET-399	FO,CU	Silecia 1 o.p.
127	10290	-0.69	0.75	8	EET-400	FO,AM	Silecia 1 o.p.
128	10291	0.38	-0.69	12	G-8		Java Criollo
129	10293	-0.55	-0.57	11	GA-11(?)		
130	10296	1.14	0.38	11	GS-29	TR,CU	
131	10297	1.96	0.34	11	GS-36	TR	
132	10309	0.61	-0.40	11	ICS-1	TR,AM	
133	10311	1.89	-0.35	11	ICS-6	TR,AN	
134	10313	1.96	-1.30	11	ICS-16	TR	
135	10315	0.38	0.41	11	ICS-39	CR	
136	10316	0.03	0.77	11	ICS-40	CR	
137	10317	-0.40	0.10	11	ICS-43		mistake
138	10318	0.44	0.01	11	ICS-44		
139	10321	-0.29	-2.17	11	ICS-46(?)		
140	10323	0.10	0.21	11	ICS-53(?)		

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
141	10325	0.31	0.32	11	ICS-60	CR	
142	10326	0.36	-0.83	11	ICS-89	TR,AM	Criollo hybrid
143	10329	1.47	-0.52	11	ICS-117(?)		ICS-6 x SCA-6
144	10333	1.34	0.65	11	ICS-13(?)		ICS-1 x SCA-12
145	10335	-2.01	0.81	11	ICS-95	TR,AN	
146	10337	0.27	-0.33	11	ICS-100	CR,AN	
147	10338	0.68	1.30	11	ICS-133		
148	10339	0.55	-0.66	9	IMC-60	FO	
149	10341	-0.10	-0.20	9	IMC-67	FO,CU	
150	10343	-1.69	-1.10	10	JACA		
151	10344	-1.89	-3.07	10	LARANJA	FO	
152	10345	1.21	-0.19	5			
153	10346	-0.90	3.37	1	LA ESMIDA		Pentagona Type
154	10351	-1.70	-0.17	10	MA-12	FO,AM	
155	10352	-1.81	-0.91	10	MA-13	FO,AM	
156	10353	0.12	-0.81	5	MATINA	FO,AM	
157	10368	-0.99	-0.71	10	MOCORONGO		
158	10372	-0.69	0.53	9	NA-34(?)	FO	
159	10374	2.20	-0.39	7	OC-77	FR,CU	Criollo hybrid
160	10376	1.25	0.96	13	P-10	FO	Criollo hybrid
161	10379	-0.61	0.79	13	P-15	FO	
162	10380	-0.88	-0.10	13	P-16	FO	
163	10381	-1.75	0.05	13	P-19	FO	
164	10382	-1.46	0.57	13	P-19	FO	
165	10383	-0.57	0.54	13	P-20	FO	
166	10385	-0.52	-0.15	13	P-23	FO	
167	10387	-0.71	1.51	9	PA-13	FO	mistake
168	10388	-0.62	-0.98	9	PA-16	FO	
169	10389	-0.79	-0.05	9	PA-16	FO	
170	10390	-0.35	-0.62	9	PA-81	FO	
171	10391	-1.08	-0.34	9	PA-121	FO	
172	10392	-1.63	0.17	9	PA-121	FO	
173	10393	-0.83	-0.58	9	PA-169	FO	
174	10397	-0.17	1.94	7	PORCELANA-3	CR,AN	mistake
175	10398	-1.14	-0.20	9	POUND-7(=P-7)	FO,AM	

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
176	10399	0.39	-0.20	9	POUND-7	FO,AM	
177	10400	-0.62	-0.26	9	POUND-12(=P-12)	FO,AM	
178	10401	-0.37	0.32	9	POUND-12	FO,AM	
179	10407	-0.19	-0.07	7	PV-4		Criollo hybrid
180	10415	1.33	0.16	1	R-2		Criollo hybrid
181	10417	1.98	0.60	1	R-8		Criollo hybrid
182	10418	0.67	0.06	1	R-9		Criollo hybrid
183	10419	1.34	0.87	1	R-10		Criollo hybrid
184	10420	0.89	0.85	1	R-13	CU	Criollo hybrid
185	10421	0.46	0.87	1	R-15		Criollo hybrid
186	10422	-0.42	1.08	1	R-19		Criollo hybrid
187	10435	-0.05	0.00	1	R-71		Criollo hybrid
188	10436	0.87	1.19	1	R-75		Criollo hybrid
189	10437	1.02	-1.12	1	R-76		Criollo hybrid
190	10438	0.16	0.05	1	R-78		Criollo hybrid
191	10441	0.85	0.72	1	R-105		Criollo hybrid
192	10443	1.64	0.68	1	R-113		Criollo hybrid
193	10444	1.37	0.20	1	R-117		Criollo hybrid
194	10447	-0.38	-0.16	10	RB-37	FO,AM	
195	10448	-0.80	-0.93	10	RB-39	FO,AM	
196	10449	-2.46	-0.36	10	(RB-39?)		
197	10450	-2.79	0.32	10	RB-41	FO,AM	
198	10452	-1.49	-1.02	10	RB-46	FO,AM	
199	10456	-1.26	-0.08	5	SANTA CLARA-3		
200	10457	0.52	-0.40	6	S,C-5	TR,AN	
201	10458	0.02	-0.18	6	S,C-6	TR,AN	
202	10464	-1.63	0.14	8	SCA-6	FO,CU	
203	10465	-0.62	0.24	8	SCA-9(?)	FO,CU	
204	10466	-1.83	0.44	8	SCA-12	FO,CU	
205	10467	2.95	-1.32	5	SCR-4		Matina hybrid (?)
206	10468	1.26	-1.16	5	SCR-5		Matina hybrid (?)
207	10469	0.72	0.39	2	SGU-60		Criollo hybrid
208	10471	-0.59	-1.40	2	SGU-3		Criollo hybrid
209	10472	-0.64	0.96	2	SGU-4		Criollo hybrid
210	10473	-0.36	0.71	2	SGU-69(?)		Criollo hybrid

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
211	10474	-1.35	-2.17	2	SGU-71		Criollo hybrid
212	10475	0.43	-0.97	2	SGU-82		Criollo hybrid
213	10477	0.53	-0.32	5	-		
214	10478	-1.50	-1.02	10	SIAL-8		
215	10479	0.26	-1.31	10	SIAL-42	CA	
216	10480	-0.24	-1.59	10	SIAL-44		
217	10481	0.62	-0.75	10	SIAL-56		
218	10482	0.02	-2.40	10	SIAL-70	FO,AM	
219	10483	-0.37	-2.01	10	SIAL-93	FO	
220	10484	-0.92	-1.23	10	SIAL-163	FO,AM	
221	10486	-1.11	-1.00	10	SIAL-325	FO,AM	
222	10487	0.29	-1.53	10	SIAL-407	FO,CA	
223	10488	-1.28	-1.76	10	SIC-1	FO	Comun
224	10489	-0.97	-2.09	10	SIC-1	FO	Comun
225	10490	-0.65	-1.60	10	SIC-2	FO	Comun
226	10491	-0.80	-0.75	10	SIC-6	FO	Comun
227	10492	-1.51	-0.97	10	SIC-7	FO	Comun
228	10493	1.97	-0.66	10	SIC-28(?)	FO	Comun
229	10494	-0.54	-1.68	10	SIC-256	FO	Comun
230	10495	-0.29	-1.65	10	SIC-329	FO	Comun
231	10496	-0.58	-1.35	10	SIC-433	FO	Comun
232	10497	-0.02	-0.72	10	SIC-802		Catongo pedigree
233	10498	-0.19	-2.43	10	SIC-806		Catongo pedigree
234	10499	0.09	-1.85	10	SIC-813		Catongo pedigree
235	10504	0.98	-0.77	12	SNK-12	TR	
236	10506	0.69	-0.46	6	SPA-5		
237	10507	1.56	-1.07	6	SPA-7	AM	
238	10508	-0.24	-0.74	6	SPA-9(?)	TR,AM	
239	10509	-0.12	-1.03	6	SPA-10(?)		
240	10510	0.03	-0.68	6	SPA-11		
241	10511	1.53	-1.09	6	SPA-12		
242	10512	-0.55	-1.02	6	SPA-17		
243	10517	-0.94	-1.74	3	T.J.-1	CR	mistake
244	10524	-0.90	0.30	11	TSAN-792		
245	10525	-0.64	0.77	11	TSH-565		SCA-6 x IMC-67

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1)]	Cultivar name	Taxonomic classification[2)]	Pedigree and/or remarks
246	10526	-0.52	0.32	11	TSH-565		SCA-6 x IMC-67
247	10527	-0.82	1.19	11	TSA-644	AN	SCA-6 x IMC-69
248	10532	-0.03	0.28	5	UF-4		
249	10533	1.15	-0.25	5	UF-10		
250	10534	1.97	-0.94	5	UF-11	TR	Criollo hybrid
251	10535	0.14	-0.80	5	UF-12		
252	10536	0.98	0.21	5	UF-20		
253	10537	0.07	-0.23	5	UF-29	FO,AM	Nacional
254	10540	-0.29	0.94	5	UF-36		
255	10541	1.16	-0.65	5	UF-93		
256	10542	0.08	-0.38	5	UF-122		
257	10543	1.45	-0.98	5	UF-168	TR	Criollo hybrid
258	10544	0.51	-0.07	5	UF-210		
259	10545	0.44	-0.42	5	UF-221	TR,AN	
260	10546	0.18	1.40	5	UF-242	AM	
261	10547	0.78	0.63	5	UF-273		
262	10548	-0.64	-0.13	5	UF-296	TR	
263	10549	-0.68	0.29	5	UF-296		
264	10550	1.60	0.08	5	UF-601		
265	10551	0.83	0.85	5	UF-602		
266	10554	0.82	-0.05	5	UF-613	TR,AM	
267	10557	1.40	-0.58	5	UF-650	TR	
268	10558	1.02	-0.10	5	UF-654		
269	10559	0.86	0.71	5	UF-666		
270	10560	0.43	0.50	5	UF-667	TR,AN	
271	10561	-0.88	-0.91	5	UF-668	TR,AN	
272	10562	1.06	-1.16	5	UF-667		
273	10564	-0.46	-0.51	5	UF-672		
274	10565	0.34	-0.42	5	UF-676	TR	
275	10566	0.83	-0.09	5	UF-677		
276	10567	-1.24	-0.93	5	UF-700		
277	10568	0.91	0.17	5	UF-700		
278	10569	0.60	-0.80	5	UF-701		
279	10570	-0.51	-0.93	5	UF-703		
280	10571	0.27	0.78	5	UF-705		

Appendix 1. continued.

Cultivar Number	CATIE accession number	Coordinates in scattergram of factors 1&2		Country code[1]	Cultivar name	Taxonomic classification[2]	Pedigree and/or remarks
281	10572	0.73	-1.26	5	UF-706		
282	10573	0.86	-0.05	5	UF-707		
283	10574	0.93	-0.91	5	UF-708		
284	10575	1.02	0.89	5	UF-709		
285	10576	1.35	0.34	5	UF-710		
286	10577	1.08	1.98	5	UF-711		
287	10578	0.12	0.48	5	UF-712		
288	10579	0.20	1.13	5	UF-713		
289	10581	0.06	-0.79	5	UF-715		
290	10583	0.14	0.71	5	UF-717		
291	10584	2.12	0.37	7	X-VERDE		
292	10585	0.38	2.04	5	189		
293	10586	1.20	-0.07	5	NO. 1		
294	111672	-0.56	0.88	11	TSH-565 (?)		

1) Country codes:

1 = Mexico	5 = Costa Rica	9 = Peru	13 = Unknown
2 = Guatemala	6 = Columbia	10 = Brazil	
3 = Honduras	7 = Venezuela	11 = Caribbean	
4 = Nicaragua	8 = Ecuador	12 = Other	

2) Explanation of abbreviations:

CR = Criollo	AM = Amelonado
FO = Forastero	AN = Angoleta
TR = Trinitario	CA = Calabacillo
CU = Cundeamor	

3) The names of these cultivars are dubious.

SUMMARY

The value of any germplasm collection depends largely on the data which accompany the individual accessions. In order to standardize cacao germplasm collection data and to facilitate genebank management and later exchange between genebanks, an extensive list of potentially useful characters or descriptors was compiled. Definitions of descriptors and their states and methodologies for their use were developed. Special attention was paid to the reliability of the data, and therefore, minimum sample size for each quantitative character was calculated. Where necessary, illustrations were included to assure clarity in the scoring of the character.

In order to facilitate the selection of the most useful characters for the systematic description of the important cacao germplasm collection at CATIE (Costa Rica) a method was developed to measure the discriminative value (D) of quantitative characters. This value D was then corrected by adjusting for correlations with other characters and it was concluded that quantitative characters, which generally have greater agronomic value than qualitative ones, are also useful in characterizing and identifying cacao clones.

Since in the past mainly qualitative characters have been used in taxonomic classifications of cacao clones it was also important to measure their discriminative power, adjusted for the amount of information shared with other characters. To enable the discriminative values of qualitative and quantitative characters to be compared, statistical tests and a formula have been devised which involve the combined use of both types of characters.

It was expected that relationships between cacao clones would influence the value of the discriminatory power of a given character for a given group of clones the relationships between clones were studied. With the help of multivariate analyses the effect of relationship between clones on the discriminative value of a character was demonstrated and phenotypic relationships proved to reflect genetic relationships. The study of the relationships between characters indicated that seed and fruit size, and weight characters formed distinct groups of closely related descriptors. It was also found that selections for seed size leads to higher dry cocoa production per fruit than selection solely for seed number per fruit.

The fruit shape does not seem to influence the total cocoa production per fruit, and a thick fruit wall does not significantly lower the production per fruit but provides a better protection of the seeds.

Since incompatibility is a well known phenomenon in cacao, an investigation was made as to whether there were differences in yields between self-compatible and self-incompatible trees or clones. Under Turrialba conditions no significant difference in seed set was found.

A study was made of the mode of inheritance of some of the qualitative characters (e.g. number of genes determining the expression of a given character) and of the discriminative value of a given character which was expected to be higher when fewer genes are involved. This relationship could not be established because too few characters were studied, however, there were strong indications that the five characters studied in a complete diallel follow tetraploid rather than diploid inheritance. This might explain the problems which have been encountered in genetic studies of cacao.

The genetic aspects of quantitative characters were studied in order to determine any relationship between the inheritance and the discriminative value of a character. Not all assumptions of the diallel cross could be verified and, therefore, it was decided not to apply the Wr-Vr regression analysis. Instead, general and specific combining abilities and reciprocal effects for quantitative fruit characters were calculated. The repeatability, additive and dominance variations were estimated. Following these results and based on progeny testing clonal propagation, a breeding method was proposed for a more efficient utilisation of cacao germplasm.

Due to an intensive exchange of cacao germplasm between Central America, the Caribbean and South America the borders between the cultivar and population groups of these regions have been ill defined or have disappeared. A study of the taxonomic classification - genetic groups based on geographical origin and fruit shape classes - was conducted and the significance of the classical groupings was proven. However, it was felt advisable to propose a new classification of the clones studied in order to enable the identification of such clones. Furthermore, multivariate analyses and simple statistical methods were employed to identify unknown, possible duplicates or mislabelled cacao clones.

The results of this study are discussed against the background of germplasm and data management. It was concluded that the quantitative characters in a crop like cacao can be successfully employed in the characterisation of germplasm, and that a systematic description of a germplasm collection can form a sound basis for further improvement of a crop. It also can greatly facilitate the taxonomic classification and identification of accessions, and this, in turn, is another important prerequisite for optimum utilisation of any germplasm collection.

SAMENVATTING

De waarde van elke verzameling van genenmateriaal hangt grotendeels af van de informatie die voor ieder afzonderlijk monster ter beschikking staat. Terwille van de gegevens standaardisatie van de cacao verzameling en om het management van de genenbank en de latere uitwisseling van materiaal tussen verschillende genenbanken te bevorderen, is een uitgebreide lijst van potentieel nuttige eigenschappen (descriptoren) opgesteld. Definities van deze descriptoren en hun klassen, evenals de procedures voor hun gebruik zijn ontwikkeld. Bijzondere aandacht is besteed aan de betrouwbaarheid van de gegevens en derhalve is de minimum-steekproefgrootte berekend. Indien nodig zijn ter illustratie tekeningen vervaardigd om de duidelijkheid van de waarnemingen te verhogen.

Om de selectie van de nuttigste eigenschappen voor de systematische beschrijving van de belangrijkste CATIE cacao collectie van voornamelijk klonen in Costa Rica te vergemakkelijken, is een methode ontwikkeld om de discriminerende waarde (D) ofwel het onderscheidingsvermogen tussen klonen van kwantitatieve eigenschappen vast te stellen. Deze D-waarde is vervolgens gecorrigeerd voor dat gedeelte, wat door correlatie met andere eigenschappen dubbel (of meervoudig) berekend wordt voor een bepaalde groep van klonen. De conclusie is dat kwantitatieve eigenschappen, die meestal een hogere landbouwkundige waarde bezitten dan de kwalitatieve, ook bruikbaar zijn voor de karakterisatie en identificatie van cacao klonen.

Aangezien in het verleden meestal kwalitatieve eigenschappen werden gebruikt in de taxonomische classificatie van cacao klonen, is het belangrijk geacht ook voor deze eigenschappen het onderscheidingsvermogen te bepalen, eveneens met de correctie voor het gedeelte van de informatie, dat met andere eigenschappen middels correlaties is gedeeld. Om een vergelijking van de discriminerende waarden van kwalitatieve en kwantitatieve eigenschappen mogelijk te maken, zijn statistische toetsen toegepast en is een formule ontwikkeld, die een gemeenschappelijk gebruik in de berekening van de discriminerende waarden toelaten.

Omdat de indruk bestond, dat de verwantschapsgraad tussen een groep van klonen de discriminerende waarde voor een bepaalde eigenschap beinvloedt, is dit onderzocht. Door middel van

"multivariate" analyses is het effect van de mate van verwantschap tussen klonen op het onderscheidingsvermogen aangetoond, evenals het feit dat de fenotypische overeenkomst tussen klonen hun genetische relatie weerspiegelt. Uit de studie van de verhoudingen tussen de verschillende eigenschappen kon geconcludeerd worden, dat de bepalende eigenschappen voor respectievelijk zaad- en vruchtgrootte onafhankelijke groepen vormen. Dit geldt ook voor de "gewicht eigenschappen". Verder is vastgesteld, dat selectie op zaadgrootte een effectievere manier is om de droge cacao productie per vrucht te verhogen, dan wanneer alleen maar op aantal zaden per vrucht wordt geselecteerd. De vruchtvorm blijkt de totale droge cacao productie per vrucht niet te beinvloeden, evenals de dikte van de vruchtwand. Hoe dikker echter de vruchtwand is, hoe beter de cacao bonen beschermd worden.

Aangezien incompatibiliteit in cacao een bekend verschijnsel is, is de invloed hiervan op de vruchtzetting in de collectie onderzocht. Er is geen significant verschil gevonden tussen de vruchtzetting op zelf-incompatibele en op compatibele bomen of klonen van de CATIE collectie in Turrialba.

Ook de vererving van enige kwalitatieve eigenschappen is onderzocht om tot een mogelijke conclusie ten aanzien van de relatie tussen het aantal genen die de expressie van een eigenschap bepalen, en hun onderscheidingsvermogen te komen. Er is verondersteld, dat hoe minder genen betrokken zijn bij de expressie van een eigenschap, des te hoger de discriminerende waarde is voor een bepaalde eigenschap. Deze relatie is echter niet vastgesteld, omdat te weinig verschillende eigenschappen onderzocht zijn. Er zijn echter sterke aanwijzingen gevonden, dat de vijf onderzochte eigenschappen in een diallele kruising tetraploide overerving vertonen en geen diploide. Dit zou een verklaring kunnen zijn voor de problemen, die bij de bestudering van de genetica in cacao tot op heden zijn ondervonden.

De genetische aspecten van de kwantitatieve eigenschappen zijn onderzocht om eveneens een relatie tussen de vererving en het onderscheidingsvermogen te kunnen vaststellen. Omdat echter niet aan alle voorwaarden, die aan een verantwoorde diallele kruisingsanalyse ten grondslag liggen, voldaan kon worden, was het niet mogelijk de Wr-Vr regressie-analyse door te voeren. In plaats daarvan zijn de algemene en specifieke combinatie-

geschiktheid en de reciproke effecten voor acht kwantitatieve vruchteigenschappen berekend. De herhaalbaarheid en de additieve en dominantie varianties werden geschat. Op grond van deze resultaten en gebaseerd op nakomelingenschapstoetsen en de klonale vermeerdering van cacao is een veredelingsmethode voorgesteld, die een beter gebruik van cacao genenmateriaal waarborgt.

Naar aanleiding van een intensieve uitwisseling van cacao genenmateriaal tussen Centraal-Amerika, de Caribische eilanden en Zuid-Amerika zijn de grenzen tussen de taxonomische groepen van cultivars en populaties uit deze gebieden verwaterd of zelfs helemaal verdwenen. Daarom is een studie van de bestaande taxonomische classificaties doorgevoerd, die zich enerzijds baseert op genetische groepen uit verschillende geografische gebieden en anderzijds op vruchtvormen. De waarde van deze classificering is aangetoond, maar om een goede identificatie van de klonen te waarborgen, was het raadzaam om een nieuwe classificatie op te stellen. Verder is het mogelijk met behulp van eenvoudige statistische methoden en met multivariate analyses onbekende, mogelijke duplicaten of fout geetiketteerde cacao klonen te identificeren.

De resultaten uit deze studie zijn in de algemene discussie belicht tegen de achtergrond van genenmateriaal en data management. Hier volgde uit dat kwantitatieve eigenschappen in een gewas als cacao met succes kunnen worden gebruikt in de karakterisering van genenmateriaal monsters en dat een systematische beschrijving van een gewas-collectie een goed uitgangspunt kan vormen voor de verdere verbetering van het gewas in kwestie. Verder kan zo'n beschrijving de taxonomische classificatie en identificatie van monsters uitermate vergemakkelijken en dit is wederom een belangrijke voorwaarde voor een optimale benutting van elke genenbank-collectie.

CURRICULUM VITAE

The author was born on November 11, 1948 in Helden (L.). He attended the "Blariacum College" (HBS) at Blerick, and started his university studies at the Agricultural University in Wageningen in 1967. He was a member of the first Executive Council of the Faculty and of various University committees and student organisations. In 1974 he passed the doctoral examination in plant breeding with genetics, and pedagogics and didactics as optional subjects.

From 1974-75 the author was employed by the Deutsche Gesellschaft fuer Technische Zusammenarbeit (GTZ) GmbH at Eschborn, Federal Republic of Germany. From 1976 - 1981 he was stationed at the genebank of CATIE, Turrialba (Costa Rica) as documentalist and administrator responsible for the GTZ project. Since 1982 he has been the project leader and documentalist of a GTZ programme in the Plant Genetic Resources Centre/Ethiopia (PGRC/E) in Addis Abeba, Ethiopia.