Pierre-Michel FORGET
le 10.10.86.

Botanical Journal of the Linnean Society (1986), *93*: 205–223

Microcomputer-aided identification: an application to trees from French Guiana

PIERRE MICHEL FORGET

Université Pierre et Marie Curie, Laboratoire de Botanique Tropicale, 12 rue Cuvier, 75005 Paris, France

JACQUES LEBBE

Service de Médecine Nucléaire, Hôpital Broussais, 93 rue Didot, 75014 Paris, France

HENRI PUIG

Université Pierre et Marie Curie, Laboratoire de Botanique Tropicale, 12 rue Cuvier, 75005 Paris, France

REGINE VIGNES

Service de Médecine Nucléaire, Hôpital Broussais, 93 rue Didot, 75014 Paris, France

AND

MICHEL HIDEUX F.L.S.

A.I. 031254, C.N.R.S., Laboratoire de Palynologie, Muséum National d'Histoire Naturelle, 61 rue de Buffon, 75005 Paris, France

Received February 1985, accepted for publication July 1985

FORGET, P. M., LEBBE, J., PUIG, H. VIGNES, R. & HIDEUX, M., 1986. **Microcomputer-aided identification: an application to trees from French Guiana.** In order to identify more easily trees from French Guiana along the Ste Elie Track, for which vegetative descriptions have already been made, a knowledge-based system named XPER involving a data matrix made up of OTUs, characters and character states has been devised. Conveniently working on several microcomputers (including new portable ones), the program of this expert system consists of four main subunits: an editor to create, consult and modify the data; a determiner to identify an OTU; a reorganizer to modify the structure from the base and connect several bases; and a printer to describe either on the visual display unit or on paper. In this method, the interaction is constant between the user and the computer. It is an on-line type of identification with a multi-access entry.

ADDITIONAL KEY WORDS:—Data matrix – expert system – knowledge-base – morphological characters – operational taxonomic units.

0024–4074/86/060205+19 $03/00/0

CONTENTS

INTRODUCTION

Several program-aided identification systems working on minicomputers or larger ones have been created since the beginning of the seventies (Morse, 1970). Pankhurst (1970) has been a pioneer in Europe by elaborating such programs, demonstrating their usefulness and developing several applications which he reported at a symposium in Cambridge in 1973 (Pankhurst, 1975a).

The identification of tropical trees gives rise to many problems for those who are working in the forest and have to make inventories of the trees. The field worker has a choice between using a diagnostic key and asking the local people the vernacular name in order to identify specimens. However, both methods give rise to difficulties. The disadvantages of the former come from character selection: a minimum number of characters have to be present in order to use a key successfully (for example, flower characters cannot be used with a non-flowering sample; Sabatier & Puig, 1983). The latter way can result with unreliable vernacular names. Phenome transcriptions cause additional problems, especially in relation to reference files which are not sufficiently complete enough and too restricted in coverage.

These problems have suggested to some authors (Chipp, 1922; Corner, 1940; Rosayro, 1953) the use of unique morphological characters, e.g. field characters, for the identification of trees, as for the dendrological characters commonly used by foresters describing trees in the field.

Data concerning these morphological characters have been presented by several authors (Letouzey, 1982; Radford *et al.*, 1974; Rollet, 1980, 1982; Schnell, 1950; Wyatt-Smith, 1954). In the literature, these data described overall the habit of the tree, the features of the trunk and trunk base, the bark type, the particularities of exudates and the leaf characters.

MATERIAL AND METHOD

Data collection

In order to make easier the identification of trees from the primary forest along Ste Elie Track (French Guiana), we have used a previous inventory made by Puig (1979). Two local informants, Paramaka and Wayapi, participated by

providing us with a vernacular name of each tree. The final identification was obtained by comparison with herbarium specimens, flowers and fruits being collected if possible. The morphological data are the result of observations on 68 tree taxa (Forget, 1984) localized in four different parcels of land covering a total surface of 1 ha (10 km^2).

Software used

XPER (Lebbe, 1984) is the first program of aided identification entirely conceived as a comprehensive microcomputer package. It is made up of four main subunits which are described in detail in the user's manual: (1) the EDITOR is the subunit of the creator of a knowledge-based system. It is used for data input, correction, consultation, addition or any other modification, to calculate taxonomic distances and to make a multi-access research and comparisons of individuals or groups. This subunit is particularly developed in XPER by giving all creators of knowledge-based systems, even those without any affinity to computers, an easy access; (2) the DETERMINER (or *inquirer*) is the subunit for the user of the knowledge-based system to identify an object of which the description is in agreement with an individual of the base; (3) the REORGANIZER also concerns the creators of a knowledge-based system, to modify the order of all the data of the base (variables, modalities and individuals) and connect several bases together (for example, several scientists working in different localities can easily join their databases made up of different taxa, but with the same variables, or those working in different fields may also link their databases containing the same taxa but different variables); (4) the PRINTER to describe a knowledge base on the video display unit or on paper either detailed or condensed.

A knowledge-based system consists of a matrix (or *frame*) including a list of individuals (or *objects*) described by variables (or *features*), each variable may have up to 14 modalities (or *attributes*). In taxonomy, individuals, variables and modalities are more frequently known as operational taxonomic units (OTUs), characters and character states as defined by Sneath & Sokal (1973) and adopted by Pankhurst (1978b), terms which have been used here. All the terms in italics are those exclusively used in the English version of XPER now available.

DATA FORMAT

Data which are entered into the program make up a data matrix of the 'characters × OTUs' type. The codification is made by the program without any intervention on the part of the user. Listing the OTUs consists of describing the columns of the matrix; listing the characters and character states consists of describing the rows of the matrix. After that, a correspondence is made between OTUs and character states, the matrix is automatically elaborated by the editor.

Listing the OTUs (Table A2)

Sixty-eight OTUs (trees from Ste Elie Track, French Guiana) have been repertoried as species (most of them) or as genera (some of them). A certain

number of specimens, given in Table A2, have been examined in order to delimitate precisely each OTU of the base and also to detect any variability within it. Fortuitously, no variable response occurs in any 'OTU × character' cell in the present example, but another knowledge-based system of pollen grains from Northern Europe (Lebbe *et al.*, in press) has shown a lot of variation in the cells (several states and sometimes all the states from a character have been scored); some cells were without significance and were not scored.

Listing the characters (Table A1)

Only qualitative characters are directly allowed but quantitative characters may be used by dividing their range of variability into several classes and transforming them into qualitative characters. The characters describe all the OTUs. Sixteen morphological characters covering 44 character states are always easily obtainable during the year. They are based on the field characters given by Letouzey (1982).

For each character, all possible character states have to be listed (Table A1). For example, in all OTUs observed, the base of the trunk has four possible states: expansions ('empattements'), buttresses, props, and nothing remarkable.

Basic and conditioned characters also have to be carefully examined to define exclusive and dependent modalities: e.g. the character 'phyllotaxis of leaflets' is not applicable when the leaves are simple.

Elaborating the data matrix

This operation consists of making a correspondence between each OTU and states of each character. The result is a table made automatically by the system such as the one shown in Table A3 where correspondence is symbolized by three asterisks and non-correspondence by dashes. In the best conditions for taking into account potential variability, the inventoried sampling should be as large as possible for each OTU (1–59 individuals in the present study).

Data input

The data are introduced into the microcomputer by typing on the keyboard associated with the central unit and are clearly displayed on a television monitor or a visual display unit. All the descriptors, e.g. OTUs, characters and character states, are easily readable because they are always textual. The matrix content is entered by giving one or (several) number(s) corresponding to all character states for each OTU. During this operation, the data remain visible.

The user need not intervene in the codification of the data for processing because it is done automatically by the software. However, the user can easily correct existing data or add new data. He can also split the process of data input, save the file in order to consult or amend it later, or merge the initial file with other files. Saving and storage are on floppy disks ($5\frac{1}{4}$ in.).

Once a file is established, many applications exist, such as:

quick retrieval of textual data,

automatic typesetting of the data matrix (Table A3) and of taxa descriptions (Table A5),

multicriteria access,
estimation of resemblance by distance or similarity coefficients (giving to the system an indirect polythetic means of identification),
comparisons between OTUs or groups of OTUs (in order to find the characters which particularize a group of OTUs and those which differentiate it from any other group = AND and OR comparisons).
search for general rules from dissimulated characteristics or knowledge of the database (this particularity will be discussed later in the text).

Some other applications are potential (these will be introduced later in the software):

dendrograms or hierarchical agglomerative classificatory procedures,
multidimensional data analysis, and
automatic generated keys.

Most of these applications are discussed by Pankhurst (1971, 1974, 1975b, 1978a, b, 1984), Pankhurst & Aitchinson (1975), Shetler (1975) and Heywood (1984). They are more or less related to computer-aided identification discussed in more detail below.

MICROCOMPUTER-AIDED IDENTIFICATION

This is an on-line method of identification, the strategy being left free to the user (Pankhurst & Aitchinson, 1975) who makes decisions in an easy dialogue with the machine, while the computer is able to justify its bearing. Thus, an estimation of the validity of the results is attainable.

Some programs of computer-aided identification are already available (see references in Pankhurst, 1978b). One particular feature of the XPER program created by Lebbe (1984), and used in this contribution, is its ability to be adapted on several popular microcomputers (either 8 or 16 bits) such as Commodore 64, Apple 2, IBM PC, Apricot F1, or any other ones using MS DOS (Microsoft Disk Operating System). From the beginning of the identification process by the determiner, the list of available characters may be retrieved on the visual display unit (see Appendix, example of identification).

The identification is made step by step, and may be split into two phases (see Appendix): first, the choice of the character described by the user (Question 1); secondly, the choice of one or several states of this character taken by the individual to identify (Question 1). At any step of the identification, the list of variables (characters) is available on the visual display unit except: non-discriminating variables (optional); daughter variables (dependent characters) if present, and where the mother variable is still available (e.g. 'leaflets' is a daughter variable from 'compound leaves' (mother variable)).

This operation is repeated until (Questions 2–5) either: (1) there is an identification where only one OTU has the proposed characters (Question 5); (2) a discrimination is not possible (partial key), because more than one OTU is identified after the complete description of characters. In order to obtain a complete discrimination, some supplementary characters may be added to the initial file; or (3) an unforeseen combination of characters occurs due to an error in the identification or a description of an OTU which is not yet in the file. In this last case, adding that description makes the file more complete. At any time,

and particularly in the case of an error, the user can backtrack without losing the answers already expressed.

At any step, the number or the list of remaining OTUs (Questions 1–5), the list of eliminated OTUs including the causes of their elimination, and the list of eliminated characters are available. The elimination of characters may be due: (1) to the fact they have already been used before; (2) to the fact that they are no more discriminant for the remaining OTUs (this automatic elimination is optional); or (3) to the fact that a dependent character has already been taken into account which excludes a new occurrence.

It is also possible to research the unknown OTU by an optimization procedure. When the identity of the taxon is presumed, it is possible to ask the system for an ordered list of characters allowing the shortest way to the solution.

DISCUSSION

There is a plethora of advantages in this method, mainly because a microcomputer is used which makes the data format and input easier and gives rise to several strategies for identification detailed by Pankhurst (1978b): diagnostic keys, matching by similarity or probability and on-line methods.

In a way similar to word processing methods, data which are entered once may be modified at demand: deletions, insertions or corrections are possible throughout the matrix content, so the knowledge-based system need not have a complete or definitive content. Indeed, the database may be progressively filled. Thus, one difference with classical methods of identification is that as users define new specimens, the knowledge base can grow by addition of taxa.

The complete freedom in the choice of determination strategies gives the user the possibility of taking into account the characters in any order. This fundamental possibility does not exist in traditional keys. Consequently, identification is never impeded by a question for which the characters are not available. Thus, the same identification can be reached in several ways and the use of imperfectly discriminating characters combined with others may have an overall significance in the identification. For example, the character 'presence or absence of pneumatophorous roots' is not significant by itself in *Symphonia globulifera* L.fil. (Clusiaceae) which may or may not have them, but this character can be used because the distinction between OTUs which always or never have such roots has already been made.

A detailed process of identification is given in the Appendix. In this example, five questions are necessary to obtain the identification of *Dicorynia guianensis* Amsh. (Caesalpiniaceae). A remarkable feature in this process is that of a doubtful choice between several states for one character: as an example, in the Appendix, Question 1, the answer 'empattement or buttress' has been deliberately chosen for the character 'tree base' in order to simulate a hesitation (the user is also assured that the other states are impossible). This makes the use of inadequately distinguishable characters possible (their states may evolve in a continuous manner). This is often the case for characters describing colour.

At each step, or at the end of the identification process, the validity of the result may be appreciated by consulting the motives of elimination for each

OTU. For example, in the Appendix (Table A5), two reasons are responsible for not identifying *Enterolobium schomburgkii* Benth. (Mimosaceae).

The program of the determiner may also give the number of differences between eliminated and identified OTUs. In the example given (Appendix, Question 5), only *Dicorynia guianensis* (Caesalpiniaceae) is totally compatible with the description of the tree for which the identification is in progress, but Table A4 shows that only one error would have been sufficient in order to obtain one of the five OTUs given above (1d).

If one identification occurs when only some characters have been considered, the user can try to confirm it by reference to some of the remaining characters. For example, it is possible to reply at stage one to those characters describing the trunk, and then later when a result occurs, to verify the identification by using leaf characters.

Search for general rules from scattered characteristics or knowledge of the base is another very interesting particularity of XPER: in our knowledge base, a useful rule has been retrieved, that all the trees with white or yellow exudates have simple symmetrical leaves. This rule has been elaborated from the 68 OTUs and concerns 17 OTUs. This is a very important rule because from the unique observation of the colour of exudates, two characteristics of the leaf, not always available, may be deduced.

As far as the method is concerned, the most obvious prerequisite is the need for microcomputer equipment on which this program will work. The cost of the least expensive equipment comes to about 10 000 FF in France, £800 in England and $775 in the U.S.A.; this is a very reasonable price for most laboratories. However, it is still difficult to use such equipment in the field in the tropics, but the new portable microcomputers, a little more expensive, are now available.

Another important, but still obvious, restriction common to all such studies is the necessity of previous, very elaborate taxonomic research, especially on taxon delimitations and the vocabulary used in the descriptions. To use this method it is also necessary to learn the fundamental rudiments of a technical language, although one which is quickly becoming universal with the intrusion of computers in everyday life. Nevertheless, a minimum investment in time is necessary.

Lastly, it would be very useful to have some software extensions giving easy access to: (a) the DELTA taxonomic data format (Dallwitz, 1984a, b); (b) the terminology of phytography, such as presented by Radford *et al.* (1974) and Stearn (1973), or similar contributions, such as a proposal for a catalogue of stereostructures (SEM) for pollen grains (Hideux & Ferguson, 1978); (c) existing data banks or even pictorial banks; and (d) taxonomical reference indexes.

The existence of such a network of data banks is not utopic but has already been stated by Heywood (1984) in the following terms: "All taxonomic activity forms part of an international network of information and communication. Although individual pieces of research can be, and are, undertaken in apparent isolation, all taxonomy is dependent in a series of internationally agreed conventions regarding names, publications, taxonomic structure and even the basic units involved both in term of categories and actual named taxa."

The software presents a limitation in that each character is restricted to 14

states. However, this is of little significance because of the existence of dependent characters. For example, leaf shape cannot be described by just one character (more than 14 states would be necessary), so numerous characters have been used. The first one dissociates simple leaves from compound leaves and, depending on the answer to this first basic character, either characters describing simple leaves or variables describing compound ones will be considered.

Of course, if the file is too large the matrix format is limited to the capacity of the memory of the central unit. When the data become too numerous the file may be split in several parts. For example, if all flowering plants in France had to be described, first a file assessing all the families would have to be created, and then a file for each family. Again, we must take into account the fact that computer techniques are continuously progressing and that memories are become larger and larger.

Finally, another software limitation which, in theory may be a major one, is that the presence of characters in an OTU is never considered as a probability as in the program GENKEY (Payne, 1975). This could prevent the differentiation of two OTUs not distinguishable by a character presence or absence, but by the fact that presence is rare in one OTU and frequent in another. Unfortunately, the use of probability theories gives rise to important difficulties. On the practical side, the gathering of such data needs great material efforts and does not give a real probability, but rather a frequency observed in a controlled population considered as representative; thus in most cases, characters are only qualified as 'rare' or 'frequent'. On the theoretical side, the fact that the probability of the presence of one character depends on other characters makes the application of Bayes theory impossible. A solution may be found by treating the probability of each character by means of fuzzy relations but the mathematical treatment of such relations ('character A more frequent than character B') is highly complicated.

CONCLUDING REMARKS

A considerable breakthrough has occurred in systematics with the introduction of data files managed by computers (databases) and of methods of computer-aided identification (Charlwood, Morris & Grenham, 1984; White, 1984). Today, with their application on microcomputers as part of an expert system (knowledge base), the progress is even more noteworthy. The main advantages of such a system are: the strategy of determination is always chosen by the user; the step-by-step identification with possible retreat; the automatic elimination of non-discriminating characters and character states; the justification of the sort by the system at each stage of the process; the hesitation (doubt) is taken into account.

Some other advantages are more specific to XPER, such as the easiness of data input, corrections and additions and the ability to deduce general rules or relations from dissimulated ones.

These methods do not need great efforts from taxonomists but from computer professionals who have created and adapted the software to several fields outside systematics. In any case, their use does not stop the user from receiving a good

basic training in systematics, but they force the experienced specialist to make all descriptions rational.

This system gives an incomparable set of services. All data entered may be automatically retrieved and used for automatic key making or handling of taxonomic descriptions; it may also be used as a means of information exchange in a data bank network.

There is no need for the taxonomist to drop his traditional method of working; he simply has the great advantage of being assisted by the automatization of time-consuming, tedious and repetitive tasks. The appearance of microcomputers in everyday life with inexpensive software and very popular languages still reinforces these advantages.

ACKNOWLEDGEMENTS

The authors acknowledge Dr J. Challe, Dr J. van Scheepen and Dr R. J. Pankhurst for their help with English translation and for their stimulating criticism when reviewing the manuscript. Experimental observations described in this contribution are extracted from the work of Forget (1984) and the software used, entitled 'XPER' (Lebbe, 1984), is distributed by Micro Application.

REFERENCES

CHARLWOOD, B. V., MORRIS, G. S. & GRENHAM, M. J., 1984. A chemical database for the Leguminosae. In R. Allkin & F. A. Bisby (Eds), *Databases in Systematics*: 201–208. London: Academic Press.

CHIPP, T. F., 1922. Buttresses as an assistance to identification. *Kew Bulletin, 1922:* 265–268.

CLIFFORD, H. T. & STEPHENSON, W., 1977. *An Introduction to Numerical Classification.* New York: Academic Press.

CORNER, E. J. H., 1940. *Wayside Trees of Malaya,* Vol. 1. Singapore: The Government Printer.

DALLWITZ, M. J., 1984a. Automatic typesetting of computer-generated keys and descriptions. In R. Allkin & F. A. Bisby (Eds), *Databases in Systematics:* 279–290. London: Academic Press.

DALLWITZ, M. J., 1984b. *User's Guide to the DELTA System. A General System for Encoding Taxonomic Descriptions,* 2nd edition. Microfiches CSIRO. Division of Entomology, P.O. Box 1700, Canberra, ACT 2601, Australia.

FORGET, P. M., 1984. *Essai d'identification des arbres de la Guyane française d'après leurs caractères morphologiques.* Paris: D.E.A. Biologie Végétale Tropicale. Université Pierre et Marie Curie.

HEYWOOD, V. H., 1976. *Plant Taxonomy,* 2nd edition. London: Edward Arnold.

HEYWOOD, V. H., 1984. Electronic data processing in taxonomy and systematics. In R. Allkin & F. A. Bisby (Eds), *Databases in Systematics:* 1–16. London: Academic Press.

HIDEUX, M. & FERGUSON, I. K., 1978. A proposal for a catalogue of stereostructures. *Proceedings of IV International Palynological Conference,* Lucknow *(1976–1977), 1:* 207–217.

LEBBE, J., 1984. *Manuel d'utilisation du logiciel XPER.* Paris: Micro Application.

LEBBE, J., NILSSON, S., PRAGLOWSKI, J., VIGNES, S. & HIDEUX, M., in press. The morphology of airborne pollen grains and spores from Northern Europe in relation to allergenic function: a microcomputer aided identification. In S. Blackmore & I. K. Ferguson (Organizers): "Pollen and Spores: Form and Function" Symposium (Poster and exhibition, March 1985). *Grana.*

LETOUZEY, R., 1982. *Manuel de Botanique forestière,* Tome 1. Nogent sur Marne: Centre Technique Forestier Tropical.

MORSE, L. E., 1970. Time sharing computers as aids to identification of plant (Demonstration). *Abstracts of the XI International Botanical Congress:* 152.

PANKHURST, R. J., 1970. A computer program for generating diagnostic keys. *Computer Journal, 13:* 145–151.

PANKHURST, R. J., 1971. Botanical keys generated by computer. *Watsonia, 8:* 357–368.

PANKHURST, R. J., 1974. Automated identification in Systematics. *Taxon, 23:* 45–51.

PANKHURST, R. J. (Ed.), 1975a. *Biological Identification with Computers.* London: Academic Press.

PANKHURST, R. J., 1975b. Identification by matching. In R. J. Pankhurst (Ed.), *Biological Identification with Computers:* 181–196. London: Academic Press.

PANKHURST, R. J., 1978a. The printing of taxonomic descriptions by computer. *Taxon, 27:* 65–68.
PANKHURST, R. J., 1978b. *Biological Identification. The Principle and Practice of Identification Methods in Biology.* London: Edward Arnold.
PANKHURST, R. J., 1984. The construction of a floristic database. *Taxon, 32:* 193–202.
PANKHURST, R. J. & AITCHISON, R. R. 1975. An on-line identification. In R. J. Pankhurst (Ed.), *Biological Identification with Computers:* 181–196. London: Academic Press.
PAYNE, R. W., 1975. Genkey: a program for constructing diagnostic keys. In R. J. Pankhurst (Ed.), *Biological Identification with Computers:* 65–72. London: Academic Press.
PUIG, H., 1979. Production de litière en forêt guyanaise: résultats préliminaires. *Bulletin de la Société d'Histoire Naturelle de Toulouse, 115:* 338–346.
RADFORD, A. E., DICKISON, W. C., MASSEY, J. R. & BELL, C. R., 1974. *Vascular Plant Systematics.* New York: Harper & Row.
ROLLET, B., 1980. Intérêt de l'étude des écorces dans la détermination des arbres tropicaux sur pied. *Bois et Forêts des Tropiques, 194:* 3–28.
ROLLET, B., 1982. Intérêt de l'étude des écorces dans la détermination des arbres tropicaux sur pied. *Bois et Forêts des Tropiques, 195:* 31–50.
ROSAYRO, R. A. de, 1953. Field characters in the identification of tropical forest trees. *Empire forestry Review, 32:* 124–141.
SABATIER, D. & PUIG, H., 1983. Phénologie et saisonnalité de la floraison et de la fructification en forêt guyanaise. *Mémoire du Muséum National d'Histoire Naturelle* (in press).
SCHNELL, R., 1950. *La Forêt Dense.* Paris: Lechevalier.
SHETLER, G., 1975. A generalized descriptive data bank as a basis for computer assisted identification. In R. J. Pankhurst (Ed.), *Biological Identification with Computers:* 197–236. London: Academic Press.
SNEATH, P. H. A. & SOKAL, R. R., 1973. *Numerical Taxonomy.* San Francisco: Freeman.
STEARN, W. T., 1973. *Botanical Latin,* 2nd edition Newton Abbot: David & Charles.
WHITE, R. J., 1984. Implementary small database systems with specialized features. In R. Allkin and F. A. Bisby (Eds), *Databases in Systematics:* 291–308. London: Academic Press.
WYATT-SMITH, J., 1954. Suggested definitions of field characters for use in the identification of tropical forest trees in Malaya. *Malayan Forest, 14:* 170–183.

Table A1. List of characters and of character states

1 Base de l'arbre
  1 empattements
  2 contreforts
  3 racines échasses
  4 R.A.S.
2 Rhytidome
  1 non visible
  2 visible
3 Lenticelles
  1 présentes
  2 absentes
4 Exfoliations
  1 présentes
  2 absentes
5 Texture de l'écorce
  1 fibreuse
  2 granuleuse
6 Tranche de l'écorce
  1 cassante
  2 non cassante
7 Exsudats
  1 présence
  2 absence
8 Couleur des exsudats
  1 incolore
  2 blanc
  3 jaune
  4 rouge
9 Texture des exsudats
  1 liquide
  2 visqueux
  3 poisseux
10 Écoulement des exsudats
  1 en gouttelettes
  2 en flot continu
11 Débit des exsudats
  1 lent
  2 rapide
12 Phyllotaxie des feuilles
  1 alternes spiralées
  2 alternes distiques
  3 opposées décussées
  4 opposées distiques
13 Feuilles
  1 simples
  2 composées palmées
  3 composées paripennées
  4 composées imparipennées
  5 composées bipennées
14 Phyllotaxie des folioles
  1 opposées
  2 alternes
15 Nombre de folioles
  1 deux
  2 trois
  3 quatre ou cinq
  4 supérieur ou égal à 6
16 Base des feuilles (ou folioles)
  1 asymétrique
  2 symétrique

1 Tree base
  1 local expansions of the lower part of the trunk
  2 buttresses
  3 props
  4 nothing remarkable
2 Bark
  1 not visible
  2 visible
3 Lenticels
  1 present
  2 absent
4 Exfoliations
  1 present
  2 absent
5 Texture of bark
  1 fibrous
  2 granulous
6 Section of bark
  1 brittle
  2 not brittle
7 Exudates
  1 present
  2 absent
8 Colour of exudates
  1 colourless
  2 white
  3 yellow
  4 red
9 Texture of exudates
  1 liquid
  2 viscous
  3 sticky
10 Outflow of exudates
  1 droplets
  2 continuous outflow
11 Intensity of exudates
  1 slow
  2 fast
12 Phyllotaxis of leaflets
  1 alternate whorled
  2 alternate distichous
  3 opposite decussate
  4 opposite distichous
13 Division of leaves
  1 simple
  2 palmately compound
  3 paripinnately compound
  4 imparipimately compound
  5 bipinnately compound
14 Phyllotaxis of leaves
  1 opposite
  2 alternate
15 Number of leaflets
  1 two
  2 three
  3 four or five
  4 more or equal to six
16 Leaf (or leaflet base)
  1 asymmetrical
  2 symmetrical

Table A2. List of OTUs (individuals)

Family	Family abbreviation	Genus and species	Code	Number of specimens examined
Anacardiaceae	ANAC	*Tapiria guianensis* Aubl.	1	2
Annonaceae	ANNO	*Anaxagorea dolichocarpa* Sprague & Sandw.	2	27
		Duguetia calycina R. Ben.	3	12
		Guatteria chrysopetala Miq.	4	8
		Unonopsis rufescens (Baill.) R.E. Fr.	5	16
		Xylopia nitida Dun.	6	3
Apocynaceae	APOC	*Ambellania acida* Aubl.	7	6
		Aspidosperma album (Vahl) R. Ben.	8	1
		Couma guianensis Aubl.	9	1
		Geissospermum laeve (Vell) Miers	10	1
		Lacmelleae aculeata (Ducke) Monachino	11	2
		Parahancornia amapa (Huber.) Ducke	12	6
Araliaceae	ARAL	*Didymopanax morototoni* (Aubl.) Decne & Planch.	13	2
Burseraceae	BURS	*Protium* sp.	14	6
Caesalpiniaceae	CAES	*Bocoa guianensis* (Aubl.) Amsh.	15	4
		Dicorynia guianensis Amsh.	16	6
		Dimorphandra sp.	17	5
		Eperua falcata Aubl.	18	59
		Eperua grandiflora (Aubl.) Benth.	19	10
		Peltogyne pubescens Benth.	20	12
		Sclerolobium melinonii Harms	21	11
		Vouacapoua americana Aubl.	22	7
Caryocaraceae	CARY	*Caryocar glabrum* (Aubl.) Persoon.	23	3
Chrysobalanaceae	CHRY	*Licania alba* (Bern.) Cuatr.	24	47
		Licania sp.	25	12
		Parinari sp.	26	3
Clusiaceae	CLUS	*Caraipa densifolia* Mart.	27	8
		Moronobea coccinea Aubl.	28	4
		Rheedia benthamiana Pl. & Tr.	29	3
		Symphonia globulifera L.fil.	30	9
		Tovomita choysiana Pl. & Tr.	31	24
		Tovomita sp.	32	4
Ebenaceae	EBEN	*Diospyros guianensis* (Aubl.) Gürke	33	1
Euphorbiaceae	EUPH	*Hevea guianensis* Aubl.	34	1
Icacinaceae	ICAC	*Dendrobangia boliviana* Rusby	35	16
Lauraceae	LAUR	*Nectandra grandis* (Mez.) Korstem	36	3
		Ocotea sp.	37	7
Lecythidaceae	LECY	*Eschweilera amara* Aubl.	38	18
Meliaceae	MELI	*Carapa guianensis* Aubl.	39	2
Mimosaceae	MIMO	*Enterolobium schomburgkii* Benth.	40	1
		Inga sp.	41	9
		Parkia nitida Miq.	42	2
		Piptadenia suaveolens Miq.	43	2
		Pithecellobium sp.	44	3
Monimiaceae	MONN	*Siparuna* sp.	45	17
Moraceae	MORA	*Brosimum utile* Pittier	46	2
		Helicostylis pedunculata R. Ben.	47	1
		Maquira guianensis Aubl.	48	1
Myristicaceae	MYRI	*Iryanthera hostmanii* (Benth.) Warb.	49	18
		Iryanthera sagotiana (Benth.) Warb.	50	15
		Virola melinonii (R. Ben.) A. C. Smith	51	4
		Virola surinamensis (Rol.) Warb.	52	4
Myrtaceae	MYRT	*Eugenia* sp.	53	9
		Myrcia sp.	54	2
		Myrciaria sp.	55	2
Olaceae	OLAC	*Hesteria microcalix* Sagot	56	10
Papilionaceae	PAPI	*Ormosia coutinhoi* Ducke	57	3
		Poecilanthe hostmanii (Benth.) Amshoff	58	6
		Pterocarpus officinalis Jacq.	59	5

Family	Family abbreviation	Genus and species	Code	Number of specimens examined
Rubiaceae	RUBI	*Amajoua guianensis* Aubl.	60	2
		Posoqueria longiflora (Rudge) R. & S.	61	2
Sapindaceae	SAPI	*Talisia sylvatica* Radlk.	62	14
Sapotaceae	SAPO	*Chrysophyllum sericeum* Miq.	63	4
		Manilkara bidentata (DC.) Chev.	64	1
		Micropholi guianensis (DC.) Pierre	65	15
		Pouteria sp.	66	20
Simaroubaceae	SIMA	*Simarouba amara* Aubl.	67	2
Sterculiaceae	STER	*Sterculia* sp.	68	9

Table A3. Data matrix (variables × individuals). Var = variables (= characters); Mod = modalities ((= character states) (see Table A1); Individuals = OTUs (68 trees from the French Guiana) (see Table A2)

		OTUs																
Var	Mod	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—
	2	***	—	—	—	—	***	—	—	—	—	—	—	—	***	—	***	***
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	—	***	***	***	***	—	***	***	***	***	***	***	***	—	—	—	—
2	1	—	***	***	***	—	—	***	—	—	***	***	***	***	***	—	—	***
	2	***	—	—	—	***	***	—	***	***	—	—	—	—	—	***	***	—
3	1	***	***	***	—	—	—	—	***	—	—	—	—	***	—	—	***	—
	2	—	—	—	***	***	***	***	—	***	***	***	***	—	***	***	—	***
4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—
	2	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	—	***
5	1	***	***	***	***	***	***	—	***	***	***	—	—	—	—	***	—	—
	2	—	—	—	—	—	—	***	—	—	—	***	***	***	***	—	***	***
6	1	—	—	—	—	—	—	***	—	***	—	***	***	***	***	***	***	***
	2	***	***	***	***	***	***	—	***	—	***	—	—	—	—	—	—	—
7	1	—	—	—	—	—	—	***	—	***	—	***	***	—	***	—	—	***
	2	***	***	***	***	***	***	—	***	—	***	—	—	***	—	***	***	—
8	1	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	***
	2	—	—	—	—	—	—	***	—	***	—	***	***	—	—	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	***
	3	—	—	—	—	—	—	***	—	***	—	***	***	—	—	—	—	—
10	1	—	—	—	—	—	—	—	—	—	—	—	***	—	***	—	—	—
	2	—	—	—	—	—	—	***	—	***	—	***	—	—	—	—	—	***
11	1	—	—	—	—	—	—	—	—	—	—	—	***	—	***	—	—	***
	2	—	—	—	—	—	—	***	—	***	—	***	—	—	—	—	—	—
12	1	***	—	—	—	—	—	—	***	—	***	—	—	***	***	***	***	***
	2	—	***	***	***	***	***	—	—	—	—	—	—	—	—	—	—	—
	3	—	—	—	—	—	—	—	—	***	—	—	***	—	—	—	—	—
	4	—	—	—	—	—	—	***	—	—	—	***	—	—	—	—	—	—
13	1	—	***	***	***	***	***	***	***	***	***	***	***	—	—	***	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	***	—	—	—	—	—	—	—	—	—	—	—	—	***	—	***	—
	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
14	1	***	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	***
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—
15	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	***	—	—	—	—	—	—	—	—	—	—	—	—	***	—	***	***
16	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
	2	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	—

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
***	***	—	—	***	***	***	—	—	***	***	—	—	—	—	***	—
—	—	***	***	—	—	—	—	***	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	—	***	***	***	—	—
—	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—	***
—	***	***	—	***	—	***	***	—	***	***	***	—	***	***	—	***
***	—	—	***	—	***	—	—	***	—	—	—	***	—	—	***	—
***	***	***	***	—	—	***	***	***	—	—	—	***	***	***	***	—
—	—	—	—	***	***	—	—	—	***	***	***	—	—	—	—	***
***	—	***	***	—	***	—	***	***	***	***	—	***	—	—	—	—
—	***	—	—	***	—	***	—	—	—	—	***	—	***	***	***	***
***	***	***	***	—	***	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	***	—	***	***	***	***	***	***	***	***	***	***	***
—	—	—	—	***	***	***	***	***	***	***	***	***	***	***	***	***
***	***	***	***	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	***	***	***	***	—	***	—	***
***	***	***	***	***	***	***	—	***	—	—	—	—	***	—	***	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
—	—	—	—	—	—	—	—	—	***	***	***	***	—	***	—	—
—	—	—	—	—	—	—	***	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
—	—	—	—	—	—	—	—	—	***	***	***	***	—	***	—	—
—	—	—	—	—	—	—	—	—	***	***	—	***	—	***	—	—
—	—	—	—	—	—	—	***	—	—	—	***	—	—	—	—	***
—	—	—	—	—	—	—	***	—	***	***	***	***	—	***	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
***	***	***	***	***	—	***	***	***	***	—	—	—	—	—	***	***
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	***	—	—	—	—	***	—	***	***	***	—	—
—	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—	—
—	—	—	—	—	—	***	***	***	***	***	***	***	***	***	***	***
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
***	***	***	***	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	***	***	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
***	***	***	***	***	***	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	***	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
***	***	—	***	***	—	—	—	—	—	—	—	—	—	—	—	—
***	—	***	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	***	—	***	***	***	***	***	***	***	***	***	***	***	***	***	***

Table A3. continued

		OTUs																
		35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	1	***	—	***	***	—	—	***	—	—	***	—	—	***	—	—	—	***
	2	—	—	—	—	***	***	—	***	***	—	—	—	—	***	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	—	***	—	—	—	—	—	—	—	—	***	***	—	—	***	***	—
2	1	***	***	***	***	***	***	***	***	***	***	***	***	—	***	***	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	***	***
3	1	***	***	—	***	***	***	***	***	***	—	—	***	***	—	***	—	—
	2	—	—	***	—	—	—	—	—	—	***	***	—	—	***	—	***	***
4	1	***	—	***	***	—	***	—	***	—	—	—	—	***	—	***	***	—
	2	—	***	—	—	***	—	***	—	***	***	***	***	—	***	—	—	***
5	1	—	***	—	***	***	***	—	***	***	***	—	***	—	—	***	***	***
	2	***	—	***	—	—	—	***	—	—	—	***	—	***	***	—	—	—
6	1	***	***	***	—	—	***	—	—	***	—	***	—	***	***	—	—	—
	2	—	—	—	***	***	—	***	***	—	***	—	***	—	—	***	***	***
7	1	—	—	—	—	—	—	—	—	—	—	—	***	***	***	***	***	***
	2	***	***	***	***	***	***	***	***	***	***	***	—	—	—	—	—	—
8	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
	2	—	—	—	—	—	—	—	—	—	—	—	***	***	—	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	—
	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***	***	—
9	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***	***	***
	2	—	—	—	—	—	—	—	—	—	—	—	***	—	***	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—
10	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	***	***	***	***	***	***
11	1	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	***	***	—	***	***	***
12	1	***	***	***	***	***	***	***	***	***	***	—	***	***	***	***	***	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	***
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—	—	—
13	1	***	***	***	***	—	—	—	—	—	—	***	***	***	***	***	***	***
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	3	—	—	—	—	***	—	***	—	—	—	—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	5	—	—	—	—	—	***	—	***	***	***	—	—	—	—	—	—	—
14	1	—	—	—	—	***	***	***	***	***	***	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	—	—	—	—	***	***	***	***	***	***	—	—	—	—	—	—	—
16	1	—	—	—	—	—	—	***	—	—	***	—	—	—	—	—	—	—
	2	***	***	***	***	***	***	—	***	***	—	***	***	***	***	***	***	***

OTUs																
52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
—	—	—	—	—	***	***	—	—	—	—	—	***	—	***	—	—
***	—	—	—	—	—	—	***	***	—	—	***	—	***	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	***	***	***	***	—	—	—	—	***	***	—	—	—	—	***	***
***	***	***	—	***	***	—	***	—	***	***	—	***	—	—	—	***
—	—	—	***	—	—	***	—	***	—	—	***	—	***	***	***	—
—	—	***	—	—	***	***	***	—	—	—	***	—	***	—	***	***
***	***	—	***	***	—	—	—	***	***	***	—	***	—	***	—	—
—	***	—	***	—	—	—	—	—	—	***	—	***	***	***	***	—
***	—	***	—	***	***	***	***	***	***	—	***	—	—	—	—	***
***	***	—	***	***	—	***	***	***	—	—	***	—	***	***	—	***
—	—	***	—	—	***	—	—	—	***	***	—	***	—	—	***	—
***	—	***	—	—	***	***	—	—	***	***	—	—	—	—	***	—
—	***	—	***	***	—	—	***	***	—	—	***	***	***	***	—	***
***	—	—	—	—	—	—	***	—	—	—	***	***	***	***	—	—
—	***	***	***	***	***	***	—	***	***	***	—	—	—	—	***	***
***	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	***	***	***	***	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	—	—	—	—	—	—
***	—	—	—	—	—	—	—	—	—	—	—	***	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	***	—	***	***	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	***	***	***	***	—	—
***	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	***	***	***	—	—	—
***	—	—	—	—	—	—	—	—	—	—	—	—	—	***	—	—
—	—	—	—	—	***	***	***	—	—	***	***	***	***	***	***	***
***	—	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	***	***	—	—	—	—	—	—	—
—	***	***	***	—	—	—	—	—	—	—	—	—	—	—	—	—
***	***	***	***	***	—	***	—	***	***	—	***	***	***	***	—	***
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	***	—	—	—	—	***	—
—	—	—	—	—	***	—	***	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	***	—	—	—	—	***	—
—	—	—	—	—	***	—	***	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	***	—	—	—	—	—	—	—	—	—
—	—	—	—	—	***	—	—	—	—	***	—	—	—	—	***	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***

Example of identification

$N_q = 1$, $N_v = 16$, $N_i = 68$, where N_q = number of the question, N_v = number of remaining characters, N_i = number of remaining OTUs.

The list of characters and the list of modalities corresponding to the selected character are successively displayed on the TV screen (see Table A1 and A2).

Question 1: We choose the character 1 (tree base). The unknown specimen has either local expansions or buttresses so we choose the states 1 and 2 (1/2), ⇒ 38 remaining OTUs: $N_q = 2$, $N_v = 15$, $N_i = 38$.

Question 2: We choose the character 2* (lenticels). The unknown specimen has lenticels (state 1), ⇒ 22 remaining OTUs: $N_q = 3$, $N_v = 11$, $N_i = 22$.

Question 3: We choose the character 3* (texture of bark). The unknown specimen has a granulous bark (state 2), ⇒ 8 remaining OTUs: *Dicorynia guianensis* (CAES), *Licania alba.* (CHRY), *Parinari* sp. (CHRY), *Diospyros guianensis* (EBEN), *Dendrobangia boliviana* (ICAC), *Inga* sp. (MIMO), *Helicostylis pedunculata* (MORA), *Ormosia coutinhoi* (PAPI). $N_q = 4$, $N_v = 10$, $N_i = 8$.

Question 4: We choose the character 5* (leaves). The unknown specimen has imparipinnately leaves (state 4), ⇒ 2 remaining OTUs: *Dicorynia guianensis* (CAES), *Ormosia coutinhoi* (PAPI). $N_q = 5$, $N_v = 2$, $N_i = 2$.

Question 5: We choose the character 2* (exfoliations). The unknown specimen has exfoliations (state 1). Determination made: *Dicorynia guianensis* (CAES).

*The number of characters is that of the list of remaining characters.

Table A4. Comparison of the identified OTU to all other OTUs of the data matrix (example of 10 OTUs)

	Difference*	OTU	Family code
1	1d, q5	*Ormosia coutinhoi*	PAPI
2	1d, q4	*Parinari* sp.	CHRY
3	1d, q4	*Dendrobangia boliviana*	ICAC
4	1d, q4	*Helicostylis pedunculata*	MORA
5	1d, q1	*Simarouba amara*	SIMA
6	2d, q4	*Diospyros guianensis*	EBEN
7	2d, q4	*Licania alba.*	CHRY
8	2d, q3	*Peltogyne pubescens*	CAES
9	2d, q3	*Eschweilera sagotiana*	LECY
10	2d, q3	*Enterolobium schomburgkii*	MIMO

*1d, q5 = one difference in question 5.

Table A5. Justification of the choices made by computer

Pas: *Enterolobium schomburgkii* MIMO.
Si Texture de l'écorce =
1 — Fibreuse
2 — Granuleuse [in video inversion]
[This is not this OTU because its bark is fibrous instead of granulous in the identified OTU.]

Pas: *Enterolobium schomburgkii* MIMO.
Si Feuilles =
1 — Simples
2 — Composées palmées
3 — Composées paripennées
4 — Composées imparipennées [in video inversion]
5 — Composées bipennées
[This is not this species because its leaves are bipinnately compound instead of imparipinnately compound as in the identified OTU.]

Table A6. Some examples of automatic description of OTUs

Description de: *Dicorynia guianensis* CAES

Base de l'arbre: Contreforts
Rhytidome: Visible
Lenticelles: Présentes
Exfoliations: Présentes
Texture de l'écorce: Granuleuse
Tranche de l'écorce: Cassante
Exsudats: Absents
Phyllotaxie des feuilles: Alternes spiralées
Feuilles: Composées imparipennées
Phyllotaxie des folioles: Alternes
Nombre de folioles: supérieur ou égal à six
Base des feuilles (ou folioles): Symétrique

Description de: *Diospyros guianensis* EBEN

Base de l'arbre: Empattements
Rhytidome: Visible
Lenticelles: Présentes
Exfoliations: Absentes
Texture de l'écorce: Granuleuse
Tranche de l'écorce: Cassante
Exsudats: Absents
Phyllotaxie des feuilles: Alternes spiralées
Feuilles: Simples
Base des feuilles (ou folioles): Symétrique

Description de: *Ormosia coutinhoi* PAPI

Base de l'arbre: Empattements
Rhytidome: Visible
Lenticelles: Présentes
Exfoliations: Absentes
Texture de l'écorce: Granuleuse
Tranche de l'écorce: Cassante
Exsudats: Absents
Phyllotaxie des feuilles: Alternes spiralées
Feuilles: Composées imparipennées
Phyllotaxie des folioles: Alternes
Nombre de folioles: supérieur ou égal à six
Base des feuilles (ou folioles): Symétrique

XPER

The translation into English is given by the list of characters and character states given in Table A1.