

Review of the genus *Helicotylenchus* Steiner, 1945. 1: A computer program for identification of the species

Renaud FORTUNER and Yinpo WONG

*Nematology Lab, Room 340, California Department of Food and Agriculture,
1220 N Street, Sacramento, Cal. 95814, USA and
University of California, Davis, Cal. 95616, USA.*

SUMMARY

A computer program is proposed to help in the identification of species in *Helicotylenchus* and other nematode genera. The identification program is believed to be more reliable and convenient than a traditional dichotomous key because: *i*) new species can easily be added to the reference list; *ii*) the intraspecific variability of measurements and morphological characters is taken into account; *iii*) the results are presented in a simple manner which allows specific identifications by non-specialists; and *iv*) the versatility of the program should make it acceptable to all scientists. Worldwide access to the program is assured through public data networks.

RÉSUMÉ

*Révision du genre Helicotylenchus Steiner, 1945.
1: Un programme d'ordinateur pour l'identification des espèces*

Un programme d'ordinateur est proposé pour aider à l'identification des espèces de *Helicotylenchus* et d'autres genres de nématodes. Le programme d'identification est supposé être plus fiable et plus facile d'emploi que les clefs dichotomiques traditionnelles parce que: *i*) les nouvelles espèces peuvent facilement être ajoutées à la liste de référence; *ii*) la variabilité intraspécifique des mensurations et des caractères morphologiques est prise en considération; *iii*) les résultats sont présentés de manière simple, ce qui rend possible les identifications faites par des non spécialistes; *iv*) la souplesse d'emploi du programme devrait le faire accepter par tous les chercheurs. L'accessibilité à l'échelle mondiale du programme est assurée par l'emploi de réseaux de banques de données.

Dichotomous keys are not a reliable tool for specific identification in genera like *Helicotylenchus* with large number of species and high intraspecific variability (Fortuner, 1983; 1984).

— Measurements vary by very small increments between species of *Helicotylenchus*. They cannot be used early in the key to define the first groupings of species.

— Morphological characters are variable in many species of *Helicotylenchus*. They cannot be used in a dichotomous manner because many species possess both states of such characters. Only a few characters are not variable, but they differentiate no more than ten percent of the species described in the genus.

For *Helicotylenchus*, and for any other genus, dichotomous keys also present other practical flaws. A key is rapidly outdated, and is difficult to update. After a number of new species have been described a new key is required. Wrong decisions at any line in the key misdirect the user and result in erroneous conclusions. Users familiar with a key, and with a preconceived idea of the identity of the population, tend to make the "right" choices at every line of the

key to arrive at what they thought was the correct name.

The tremendous increase in the number of described species; increased awareness of the intraspecific variability; and increased need to reliably identify species, require improved identification methods. The methods must be able to handle successive additions to the list of species in the genus, take into account intraspecific variability, and must be easy to use by any nematologist and help him obtain a reasonably objective identification.

The general coefficient of similarity of Gower

Once a nematode X has been determined as belonging to a particular genus G, identification of X involves comparisons with all the species in G. This nematode will be identified as the species of G with which it is the most similar. When X is dissimilar to all the species in G, it represents a new species.

Similarity between X and any species S in G can be estimated by a comparison of successive pairs of characters recorded for X and for species S. If the two values for a character are identical in X and in S,

this character will receive a score of one. If they are not quite identical, they will be scored between zero and one. A score zero is given to characters perfectly dissimilar (for example fasciculi present in X, absent in S). The overall similarity between X and S can be estimated by averaging the scores of all the successive characters.

The characters used for the diagnosis and identification of the species in *Helicotylenchus*, and in other nematode genera, are of three kinds (Table 1) :

- Measurements.
- Two-state non variable characters.
- Two - or multi - state variable characters (when several states of the character can coexist in the same species).

The scoring process is different for each kind of character. Gower (1971) proposed a general coefficient of similarity that can handle all three kinds of characters.

SCORING PROCESS FOR MEASUREMENTS

The mean value of a measurement M_i is recorded in the sample X (M_{xi}) and in a species S (M_{si}). The absolute value of the difference $M_{xi} - M_{si}$ is computed.

A measurement will be scored $S_i = 1$ when the mean values are identical in X and S. When $M_{xi} = M_{si}$, $|M_{xi} - M_{si}| = 0$. The score S_i is computed as $S_i = 1 - |M_{xi} - M_{si}|$.

When the mean values are not quite identical in X and in S they should be given lower scores. The score should be zero if $M_{xi} - M_{si}$ represents the highest possible difference in the genus for the measurement. This is equal to the difference R_i between the highest (M_{Max}) and the lowest (M_{Min}) specific mean values of M as recorded in the known species in the genus.

The scoring process must be corrected to take into account the fact that measurements vary under external factors. Food, ecological environment, and geographic origin of the sample affect the measurements. For example, in *Helicotylenchus* it has been estimated that mean body length can vary by up to 150 μm within the same species (Fortuner, 1984). If $|M_{xi} - M_{si}|$ is higher than zero, but less than the estimated intraspecific variability, the score of M should be set to one. This is achieved by subtracting a correction factor C_i from $|M_{xi} - M_{si}|$. In *Helicotylenchus*, this correction factor is equal to 150 μm for body length, 3 μm for stylet length, etc., (Tab. 1). These values are an estimate of the intraspecific variability within a genus. They are subjective values and depend on the judgement of the author and of what was known on the subject at the time they

were proposed. A user of the identification program is free to accept these values or to substitute values of his own.

The relative value of the difference $|M_{xi} - M_{si}| - C_i$ must be considered when scoring a measurement. In *Helicotylenchus*, two stylet mean values differing by 15 μm are highly dissimilar because in this genus the difference between the highest specific mean value (40.5 μm) and the lowest (18.5 μm) is only 22 μm . For body lengths, a similar difference of 15 μm is negligible because the range of specific mean body lengths is 1,250 - 420 = 830 μm . The relative value of the difference of a measurement in X and S is calculated by dividing the corrected absolute value $|M_{xi} - M_{si}| - C_i$ by the corrected range of value for this measurement : $R_i - C_i$.

The final formula for scoring measurements becomes :

$$S_i = 1 - \frac{|M_{xi} - M_{si}| - C_i}{R_i - C_i}$$

When $|M_{xi} - M_{si}|$ is smaller than C_i , the numerator becomes negative. S_i is then arbitrarily set to 1.

For example, the score of the comparison of body lengths between a sample with mean body length $M_{xi} = 561$ μm , and a species with mean body length $M_{si} = 937$ μm is calculated as :

$$S_i = 1 - \frac{|561 - 937| - 150}{830 - 150}$$

$$S_i = 0.67$$

NON-VARIABLE MORPHOLOGICAL CHARACTERS

A few morphological characters do not vary within a species. In *Helicotylenchus*, males and fasciculi (canals) are either present or absent, the labial disc is either visible or not visible in lateral view (Fortuner, 1984). Any such character appears as one of two states. The states can be coded 0 (absence, non-visible, etc.) or 1 (presence, visible, etc.). Because there is no intraspecific variability, negative matches (both characters absent) are accepted as proof of similarity.

The scoring process is very simple in this case. When the character is present in both species (1 - 1) or absent in both species (0 - 0), it will be scored 1. When it is present in one and absent in the other (1 - 0 or 0 - 1) it will be scored 0.

VARIABLE MORPHOLOGICAL CHARACTERS

Most morphological characters are intraspecifically variable in *Helicotylenchus*. For example, the labial region is not "either rounded or flattened". A third category includes the species where some individuals have a more rounded labial region while in other

specimens it is rather flattened. To accommodate this intraspecific variability, each state of the character must be coded and scored separately. The scores of the various states are then averaged to obtain the final score of the character. The negative matches (a particular state of the character is absent both in X and S) must be neutralized because here absence of a state is not proof of similarity.

For example, stylet knob shape is coded for three possible states : indented, flat/rounded, sloping (Fortuner, 1984). Let sample X be indented and species S be indented to flat/rounded. With the states coded 0 for absence, 1 for presence, sample X is coded 100, and species S is coded 110. The scores for the three states are one (1 - 1), zero (0 - 1), and neutralized (0 - 0). The final score of the character is $(1 + 0)/2 = 0.5$.

WEIGHTING THE VARIABLES

Fortuner (1984) proposed the use of 24 characters to differentiate the species in *Helicotylenchus* (Tab. 1). Some nematologists may not accept all of these characters as taxonomically significant and they have the possibility to neutralize any character. The user may omit the value of the character(s) he wants to neutralize in the description of sample X. The program affects a weight of zero to all missing values. The user may also enter the value of this character, but neutralize it during a first run of the program by giving it a weight equal to zero. He can later change this weight and run the program again.

All characters known in both sample X and species S are arbitrarily given an equal weight ($W = 1$) by the program. Some users may consider that some

Table 1
List of characters for the genus *Helicotylenchus*

Characters	Weights	Ranges	Correction Factors
MEASUREMENTS :			
Body length (μm)	0.7	830	150
Stylet length (μm)	1.0	22	3
Distance dorsal gland opening to stylet (μm)	0.3	10	5
Distance anterior end to excretory pore (μm)	0.6	50	10
Body annule width (μm)	0.1	2.5	1
Tail length (μm)	0.3	22	3
Tail annules	0.5	19	13
Position phasmids (annules from anus)	0.7	24	6
Ratio a	0.3	24	8
Ratio c	0.1	71	15
Ratio c'	0.6	1.7	0.5
Ratio m	0.8	12	4
Ratio V	1.0	40	2
Spicule length (μm)	0.7	20	3
NON-VARIABLE CHARACTERS :			
Presence of males	0.9		
Labial disc visible (transverse view)	0.8		
Presence of intestinal fasciculi	0.9		
VARIABLE CHARACTERS :			
Habitus	0.9		
Lip shape	0.9		
Incisure junction pattern	0.8		
Lip annulation	0.3		
Stylet knob shape	0.1		
Tail shape	0.6		
Posterior genital branch	1.0		

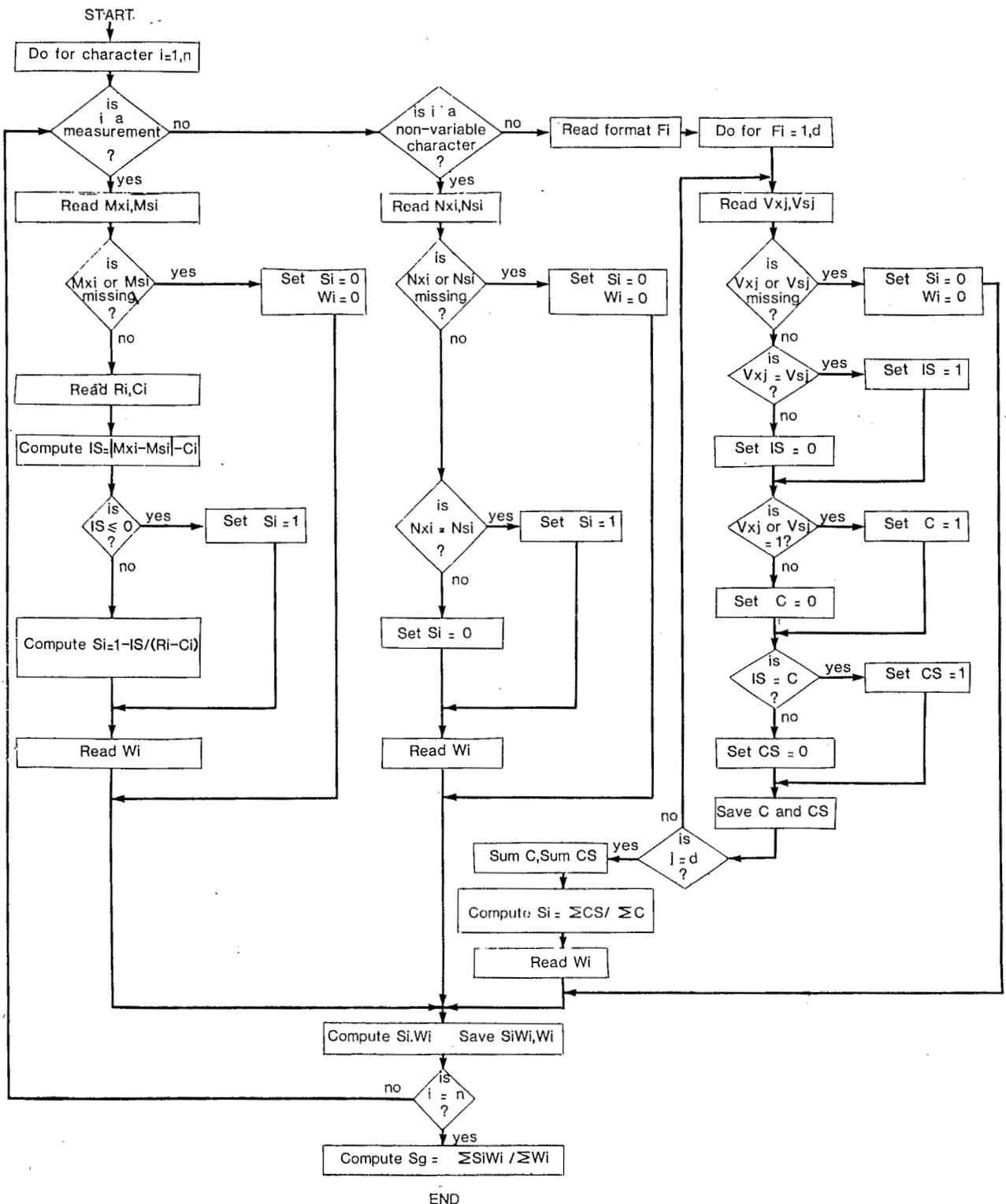


Fig. 1. Flowchart of program for computing the coefficients of similarity between a sample X and each species S in a genus data file. n : number of characters used in the genus. Mx, Ms : measurement for X, for S. Nx, Ns : non-variable morphological character for X, for S. Vx, Vs : variable morphological character for X, for S. Ri, Ci : range and correction factor for the measurement i. IS : intermediate score. Si, Wi : score and weight for the character i. d : number of digits used for a variable morphological character. C, CS : coefficient and corrected score for a variable morphological character. Sg : coefficient of similarity of Gower between X and S.

characters are taxonomically more significant than others. They have the possibility to enter the weights they want to attribute to each character they use. Weights are chosen from zero (character neutralized) to one (character taken at full value).

The computation of each successive coefficient of similarity is illustrated in Figure 1.

REFERENCE LIST OF SPECIES

The program NEMAID includes a reference datafile with the descriptions of the species in a particular genus. For *Helicotylenchus*, the present datafile includes 173 descriptions of new species or of neotypes or topotypes of old species. Some of the species included have been considered by some authors to be synonyms of other species. They are included in the datafile for the benefit of the users who do not accept the synonymizations. Users who do accept certain species to be synonyms will not consider them in the final identification decision. The datafile also includes 49 descriptions of additional populations of some known species. These additional descriptions were selected by several criteria: samples of reasonable size ($n > 10$), mean values calculated for the measurements, and intraspecific variability of the characters recorded. When a species is known from at least ten localities, a composite description has been entered in the datafile incorporating the intraspecific variability from all the successive descriptions of this species.

Users may compare their samples to all the species in the datafile (including the additional descriptions) or to only the original descriptions. They can also select the species possessing a particular characteristic. For example, a sample of *Helicotylenchus* with fasciculi present can be compared only to the species where this structure has been reported. It is planned to enter genera in the program using the widest possible generic definition. For example, *Rotylenchoides* has been synonymized to *Helicotylenchus* by Fortuner (1984). All the species in *Rotylenchoides* have been added to the datafile for *Helicotylenchus*. The character previously used to differentiate the genera (number of genital branches) has been included to participate in the specific identification. Any user who rejects the synonymization of the two genera can still use the program. If he recognizes the validity of *Rotylenchoides* and wants to identify a sample belonging to this genus, he will be able to select among the species in the datafile those with one anterior genital branch and a post uterine sac (P.U.S.). He may also select the species with the posterior genital branch either reduced to a P.U.S. or degenerated but still visible; this depends on which definition of *Rotylenchoides* is accepted by the user. The identi-

fication will be made by the program only against the selected species.

Comparisons of the sample may also be made against a list of species selected by a first run of the program; for example, using different weights or a different list of characters.

NEMAID has been designed primarily for the identification of an unknown nematode sample. It can also be used to evaluate the taxonomic validity of any one of the species in the datafile by comparing it to the other species in that file.

Output

The output is described in details in Fortuner and Wong (1983).

After coefficients of similarities have been calculated between sample X and all the species in the datafile, the user is presented with a list of the species with coefficients higher than 0.7 (Fig. 2). The user can examine in detail the comparison process between his sample and any species (Fig. 3). He can then run the program a second time after modifying some parameters (weights of the characters, characters utilized, etc.). The final decision on identification of the sample is the sole responsibility of the user. Users should be aware of the following points:

- "Garbage in, garbage out." If a sample is described from too few specimens, if its variability is not properly assessed, if errors are made on measurements, it cannot be correctly identified by NEMAID, or by any other method.
- The number of characters used, and the weight attributed to these characters, directly affect the results. When comparisons are made with only a few characters, many species will be found with a high coefficient of similarity.

Similarly, species poorly described will often be found among the species with high coefficients, because they are included in the reference file with only a few characters. New users are advised to make several test runs of the program using different parameters and lists of characters, to observe how the results are affected.

To test the effect of character weighting on the results, *H. dihysteroides* (a species very close or synonym to *H. dihyssleria*) was compared to all the descriptions in *Helicotylenchus* datafile, using three different sets of weights.

1. All weights were first set to one. Only five descriptions had coefficients of similarity S_g higher or equal to 0.97. This allows an easy discrimination of related species. However, some descriptions of

H. dihystra had Sg as low as 0.93. The accuracy of the results is not perfect.

2. The weights proposed in Table 1 were used. The accuracy of the results is definitely improved. All the descriptions of *H. dihystra* had Sg at least

equal to 0.96. However, the discriminatory powers were not as good with Sg of fourteen descriptions at least equal to 0.97.

3. When the program uses only six characters (stylet, V, presence of males, position of phasmids, shape

THE SPECIES WITH THE HIGHEST COEFFICIENTS OF SIMILARITY ARE LISTED BELOW WITH:

- THEIR COEFFICIENT OF SIMILARITY
- THE NUMBER OF CHARACTERS USED (WEIGHTS NOT EQUAL TO 0.0)
- THE NUMBER OF CHARACTERS IN AGREEMENT (SCORE AT LEAST EQUAL TO 0.7)

SPECIES: BROO DIHYSTEROIDES

.COMPARED TO: ALL SPECIES

CODE NAMES	SPECIES NAMES	COEFFICIENTS OF SIMILARITY	NUMBER OF CHARACTERS USED	NUMBER OF AGREEMENTS
BROO	H. DIHYSTEROIDES	1.00	21	21
BQ11	H. DIHYSTERA	1.00	16	16
FFOO	H. OLAAE	1.00	10	10
BQ13	H. DIHYSTERA	0.99	9	9
BQ00	H. DIHYSTERA	0.98	18	16
BQ03	H. DIHYSTERA	0.98	18	16
CS02	H. LEOCEPHALUS	0.98	20	19
DG00	H. NANNUS	0.98	16	14
EA00	H. PUNICAE	0.98	19	17
DZ01	H. PTERACERCUS	0.98	20	19
BP00	H. DIGONICUS	0.97	16	14
BQ99	H. DIHYSTERA	0.97	17	15
BQ08	H. DIHYSTERA	0.97	21	18
CU00	H. LONGICAUDATUS	0.97	17	15
AA00	H. ABUNAAMAI	0.96	21	19
BH00	H. CRENAUDA	0.96	19	16
BL00	H. DENSIBULLATUS	0.96	20	18
BU00	H. ELEGANS	0.96	13	12
BQ05	H. DIHYSTERA	0.96	20	17
GF00	H. LEUCERNIS	0.96	11	10
CT00	H. LOBUS	0.96	18	16
DP00	H. PARAPLATYURUS	0.96	20	19
GG00	H. PARAPTERACERCUS	0.96	20	18
EF00	H. ROTUNDICAUDA	0.96	21	19
GK00	H. SHAKILI	0.96	20	17
EP00	H. TALONUS	0.96	21	19
EI02	H. SERENUS	0.96	19	18
AD00	H. BORINQUENSIS	0.95	14	12

Fig. 2. Comparison of *H. dihystreroides* with all the descriptions in *Helicotylenchus* datafile, using character weights given in Table 1. List of descriptions with highest coefficient of similarity (Sg).

of tail, and development of posterior genital branch) seventeen descriptions had Sg equal to one, and 25 had Sg at least equal to 0.97. At the same time, some descriptions of *H. dihystrera* had Sg as low as 0.92.

As with the traditional identification methods, no final decision should be taken without first checking the original printed description of the probable species.

Discussion

NEMAID is available for the moment with only one genus : *Helicotylenchus*. It is hoped that taxono-

mists will soon propose other genera to be added to NEMAID datafiles.

The program NEMAID makes it very easy to add newly described species to the list of species in a genus. For each new species or for additional descriptions of known species, a new line can be added to the datafile as long as the characters previously defined for the genus are sufficient. If new characters are involved, they must be described for all the species in the genus datafile.

The program takes into consideration the intraspecific variability of the taxonomic characters. Difference of measurements are considered to be significantly different between two samples only if they

CHARACTERS	SPECIES H. DIHYSTERA		SCORE	WEIGHT	SW
	BROO	BQ99			
BODY LENGTH	0740	0670	1.00	0.7	.70
STYLET LENGTH	25.5	25.3	1.00	1.0	1.00
POSITION DGO		13			
POSITION EXCRET. PORE	112				
BODY ANNULE WIDTH	1.5				
TAIL LENGTH		18			
TAIL ANNULES	09	10	1.00	0.5	.50
POSITION PHASMIDS	30	27	1.00	0.7	.70
RAID A	1	26	1.00	0.3	.30
RAID C	45				
RATIO C'	1.2				
RATIO M	0	48	1.00	0.8	.80
RATIO V	63.0	63.8	1.00	1.0	1.00
SPICULE LENGTH		22			
PRESENCE MALE	0	0	1.00	0.9	.90
LAB. DISC VISIBLE	0	0	1.00	0.8	.80
PRESENCE CANALS	0	0	1.00	0.9	.90
HABITUS	01	01	1.00	0.9	.90
LIP SHAPE	10	10	1.00	0.9	.90
INCISURE SHAPE	10	10	1.00	0.8	.80
LIP ANNULES	011	010	0.50	0.3	0.15
KNOBS SHAPE	110	110	1.00	0.1	.10
TAIL SHAPE	0011	0111	0.67	0.6	0.40
GENITAL BRANCH	100	100	1.00	1.0	1.00
TOTAL				12.20	11.85
			SG = 11.85/12.20		
			= 0.97		

Fig. 3. Comparison of *H. dihystreroides* with all the descriptions in *Helicotylenchus* datafile, using character weights given in Table 1. Detailed comparison between *H. dihystreroides* and a composite description of *H. dihystrera*.

exceed a defined limit. Variability in morphological characters is incorporated in the sample description by the artifice of scoring the various states of the character separately.

The details of the comparisons between the sample and any species can be examined. This emphasizes the differences that may exist between sample and species. Nematologists should be able to decide on a correct identification, even if they are not specialized in the genus involved.

Particular care has been taken to ensure that anyone can use the program regardless of one's taxonomic philosophy. Splitters and lumpers, cladists, pheneticists, and tenants of evolutionary classification can all use NEMAID with the particular combination of specific and generic characters, weights, and correction factors they favor. Naturally the same sample will be identified differently by users using different parameters. This is unavoidable. Identification is the subjective appreciation of resemblance, and identifiers are expected to make their own choices about what they think constitutes similarity. The program NEMAID is offered as a tool to permit easier comparison, but in no case can it be considered as an electronic nematologist.

This freedom may frighten nematologists not familiar with the taxonomy of *Helicotylenchus* or of the genera that will be later included into the program. It is planned to ask the author describing a new genus for NEMAID to propose weights for the characters he includes. Users will be free to accept the weights provided by the taxonomic authority for the genus, or to use their own weights. As an example, weights are proposed for the characters in *Helicotylenchus* (Tab. 1).

Availability of NEMAID

NEMAID is deposited in the computers of University of California at Berkeley, It can be accessed

from many countries in the world through TELENET and many other public data networks. Most nematologists have computer services available from their university or research facility. Prospective users without any computer service need only limited equipment: a terminal, a phone connection module (modem), and eventually a printer. The purchase price of this material (about \$ 2,500) is small when compared to the price of a light microscope (about \$ 25,000), another tool indispensable for identification.

ACKNOWLEDGEMENTS

Nematode identification by computer was first investigated in collaboration with Dr. Merny. Dr. Collette Roux (INRA, Jouy-en-Josas) made the preliminary analyses. Mr. Déjardin (ORSTOM, Bondy) wrote a first version of the program that demonstrated its feasibility. Drs. Merny and H. Ferris have reviewed the manuscript. Their help is here gratefully acknowledged.

REFERENCES

- FORTUNER, R. (1983). Computer assisted semi-automatic identification of *Helicotylenchus* species — The program NEMAID. *California Plant Pest and Disease Report*, 2 : 45-48.
- FORTUNER, R. (1984). Morphometrical variability in *Helicotylenchus* Steiner, 1945. 6 : Value of the specific characters used for identification. *Revue Nématol.*, 7 : 245-264.
- FORTUNER, R. & WONG, Y. (1983). *NEMAID, computer program for identification of nematodes — User's manual*. Publ. 640, California Department of Food and Agriculture, 44 pp.
- GOWER, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27 : 857-871.