

DITYL : computerized key for species identification of *Ditylenchus* (Nematoda : Anguinidae)

Tomasz VISCARDI and Michał W. BRZESKI

Instytut Warzywnictwa, 96-100 Skierniewice, Poland.

Accepted for publication 1st December 1992.

Summary – A computer program to assist the identification of species of *Ditylenchus* Filipjev, 1936 is described. The program has user-friendly interface, is simple to use and can easily be extended. The general coefficient of similarity, defined by Gower, was taken as a measure of similarity. It allows for simultaneous comparison of all features used for identification. The computer output lists five species most similar to the unknown specimens being identified.

Résumé – *DITYL : une clé informatique pour la détermination des espèces du genre Ditylenchus (Nematoda : Anguinidae)* – Description est donnée d'un programme informatique pour l'identification des espèces du genre *Ditylenchus* Filipjev, 1936. Ce programme comporte une interface facile et agréable à utiliser, et peut aisément être étendu à d'autres espèces. Le « coefficient général de similarité » défini par Gower a été utilisé pour mesurer la similarité. Il permet en effet la comparaison simultanée des différents caractères utilisés pour l'identification. La liste produite par l'ordinateur comporte les cinq espèces les plus semblables à une espèce inconnue en voie d'identification.

Key-words : Nematodes, *Ditylenchus*, identification, computerized key.

Computers have been used for identification of species of some nematode genera (Fortuner & Wong, 1984; Rey & Mahajan, 1988; Rey *et al.*, 1988; Boag *et al.*, 1988). All these are based on Gower's (1971) General Coefficient of Similarity and its various modifications. This is especially helpful when the number of known species of a genus makes preparation of a dichotomous key difficult. A computer enables simultaneous comparison of a number of characters of an unknown population with data of known species introduced into a program. The genus *Ditylenchus* appears to be especially suited for making use of this technique because of the many known species, which have small differences between each other and the relatively few characters useful for species identification (Fortuner, 1982; Brzeski, 1991; Sturhan & Brzeski, 1991). This paper gives a description of a method used to produce a computerized key for the genus *Ditylenchus* to calculate coefficient values for qualitative characters, hierarchically divide the data base and to calculate coefficient values for quantitative characters when ranges of measurements of various characters strongly overlap for many species.

Materials

The characters used for this key are based on lattice produced by Brzeski (1991). To evaluate the stability of characters within population, the coefficients of variability were calculated for eight quantitative morphological characters of females from twelve populations (Ta-

ble 1). Position of vulva (V) is the most stable (mean CV = 1.8) and post-vulval uterine sac (PUS) is the most variable. However, the degree of stability of a character varies between populations and species.

Program description

The program was written in Pascal on a computer compatible with IBM PC with 640 kB RAM and 360 kB floppy disc drive. It takes 150 kB of memory of which about 30 kB is the data base with characteristics of known species. The equipment may also include an external printer. The data base is kept in a separate file. This makes it easy to append or correct data for any species.

Gower's General Coefficient of Similarity was taken as a measure of similarity. It allows one to consider both quantitative and qualitative characters, makes possible weighting of each character separately and allows for missing information.

Fourteen characters were used to characterize species: four qualitative and ten quantitative (Table 2). Patterns for quantitative characters were defined by a mean value and a range of variability. The mean values use are estimated from the means of many populations measured.

Table 1. Coefficients of variability of some morphometric characters of *Ditylenchus* spp. females.

Species, population (number of specimens)	L	a	c	c'	V	Stylet	PUS/VBD	PUS % VA
<i>D. adasi</i> , Stężycza ¹ (16)	3.3	4.2	4.4	8.2	1.5	2.9	12.2	13.5
<i>D. australis</i> , type ² (20)	10.4	10.1	5.6	9.0	0.8	5.2	21.0	–
<i>D. equalis</i> , Puszcza Białowieska ¹ (10)	11.4	7.4	11.5	8.4	2.0	5.8	14.7	22.3
<i>D. exilis</i> , type ² (21)	7.2	7.0	4.1	6.9	0.9	5.9	13.7	–
<i>D. longimatricalis</i> , Bory Tucholskie ¹ (24)	5.2	11.8	5.6	10.6	1.2	4.1	19.6	20.7
<i>D. longimatricalis</i> , Puszcza Biała ¹ (15)	5.6	8.2	5.6	11.0	1.7	2.1	9.6	11.1
<i>D. medians</i> , Dziekanów Leśny ¹ (13)	6.6	10.8	4.7	7.9	1.6	7.0	12.1	12.1
<i>D. medicaginis</i> , Rewa ¹ (11)	16.7	16.0	12.8	8.1	4.2	5.2	20.5	27.1
<i>D. parvus</i> , Podkowa Leśna ¹ (18)	11.1	8.4	8.7	11.0	1.9	6.2	17.9	15.5
<i>D. phyllobius</i> , Texas ¹ (10)	10.8	13.7	14.9	12.6	2.3	4.3	13.1	15.9
<i>D. terricoius</i> , type ¹ (24)	5.2	7.0	7.7	10.7	1.9	2.0	10.0	10.0
<i>D. sp. n.</i> ³ (30)	9.2	7.9	7.7	7.2	1.5	7.0	14.6	12.8
Mean CV	8.5	9.4	7.8	9.3	1.8	4.8	14.9	16.1
Range	13.4	11.8	10.8	5.7	3.4	5.0	11.4	17.1

1. Measurements in Brzeski (1991).
2. Measurements in Brzeski (1984).
3. Measurements not published.

Table 2. Characters and codes used in the DITYL programme.

Character	Code
1 Female body length, mm	
2 a (body length : body width)	
3 c' (tail length : body width at anus)	
4 c (body length : tail length)	
5 V (position of vulva in per cent of body length)	
6 Stylet, µm	
7 Spicule, µm	
8 PUS/VBD (posterior uterine sac : body width at vulva)	
PUS % VA (posterior uterine sac length in per cent of vulva-anus distance)	
10 Bursa (% of tail)	
11 Lateral lines (number)	
12 Median bulb valvae	Present 1 Absent 0
13 Tail tip	Pointed 1 Rounded 2 Dull 3 Sickle 4
14 Tail shape	Short, tapers evenly 1 Tapers more anteriorly, then almost cylindrical 2 Elongated, tapers more or less evenly 3 Elongated, posterior part almost cylindrical 4

General Coefficient of Similarity between unidentified sample and a pattern of the *j*th species is defined as the average score taken over all possible comparisons :

$$S_j = \frac{\sum_{k=1}^n s_{jk} w_k}{\sum_{k=1}^n \delta_{jk} w_k}$$

where :

- S_j = General Coefficient of Similarity
- n = number of considered characters
- s_{jk} = the score for *k*th character
- w_k = the weight for *k*th character
- δ_{jk} = value representing the possibility of making comparisons. It is equal to 1 when the *k*th character can be compared for the unidentified sample and the *j*th pattern and is equal to 0 when information is missing in the sample or in the pattern or in the case of dichotomous variables when the character is non-existent in both sample and pattern.

Calculation of the scores for a particular character depends on the character type. The program distinguishes three types of characters :

- a) *dichotomous characters*, where 1 denotes the presence of the character and 0 its absence. S_{jk} is equal to 1 for positive match (1-1) between a sample and a pattern and 0 otherwise, δ_{jk} is equal to 0 for negative match (0-0) and 1 for positive match (1-1) and mismatches (1-0, 0-1) between a sample and a pattern.

b) The programme distinguishes three types of *qualitative characters* :

- characters where no intraspecific variability is observed,
- characters with intraspecific variability, but no preferences are given to certain values,
- characters with intraspecific variability, where preferences are given to certain values.

The first type is represented by two characters : number of lateral lines and presence or absence of the thickenings of the oesophageal lumen walls in the median bulb called here median bulb valvae. They were used to divide hierarchically the data base. When values of these characters are entered, the program makes identification based only on the subset of the data base which means that only patterns with positive matches for lateral lines and median bulb valvae are used. For two remaining types of qualitative characters the comparison is based rather on common sense than strictly on mathematical rules. This is because percentage characterizations are not always available.

In DITYL the following procedure is used :

First there are indicated all possible values of qualitative variables for given species. Then they are sorted in descending order with respect to possibility of their occurrence.

Possibility is set in terms of : usually, rarely, often. The character tail tip has four possible values (Table 2). The description of a pattern may look like this : 1, 2, 3. This means that possible values are pointed, rounded and dull and moreover that most typical for this pattern is pointed, then rounded and then dull in descending order. During calculation of the Gower's Coefficient value the program scores positive match between entered values and any value in a pattern description as 1. For the third type of qualitative characters at final selection when some patterns have equal value of similarity coefficients preferences are given to that character with positive match for more typical value of this character.

c) *quantitative characters*. The score s_{jk} is calculated according to the formula :

$$S_{jk} = 1 - |y_k - x_{jk}| / R_{jk}$$

where :

- y_k = value of k th character for unidentified sample
- x_{jk} = value of k th character for j th pattern
- R_{jk} = variability range for k th character for the j th pattern.

Range of measurements of various characters used in the key overlap for many species. Therefore in the formula for calculation of Gower's Coefficient of Similarity, in the case of quantitative characters, a range was used for the character in species instead of the whole

range for all species included in data base. When $y_k = x_{jk}$, then $s_{jk} = 1$, but if y_k differs from x_{jk} more than R_{jk} , the value of s_{jk} is set to 0.

For every character missing values are allowed.

The program starts reading the data for all the species from the data base file and then displays a menu :

- 1 - Append or correct species data
- 2 - Erase species
- 3 - Data input and identification
- 4 - Weight setup
- 5 - Exit

Options 1 and 2 are used to handle the data base file. Option 3 is used for entering the data of an unknown population that has to be identified. These data can be introduced either as calculated means of measurements of some specimens, or as actual measurements allowing the program to average the data. When the input is finished the identification begins and five species with the highest values of the General Coefficient of Similarity are displayed. This may also include short additional pieces of information that supplement numerical data and help to take a final decision. Because users of the program represent widely varying levels of acquired computer skills the system interface had to be very user friendly. All keyboard inputs are mainly simple responses (Y/N, numeric) and require minimal computer literacy. Data are entered into the program in a concrete and easy way using forms that are displayed on the screen and are interactively filled out by the user. Responses to the questions are entered successively, after a cursor is transferred to an allocated place on the form. For qualitative characters additional menu windows are displayed. After entering all data the user can check character values and then confirm the whole form or make corrections.

Identification results

In practical identification of nematodes dichotomous or polytomous keys are not sufficient for final species identification. The role of the key is rather to indicate the group of species with key-characters the most similar to the unknown examined population. It may not imply the most related species. A final decision must be taken by comparison with a good description and, if possible, reference specimens. According to these obvious statements the similarity coefficients are shown in identification results of DITYL to allow judgment of the degree of similarity. The actual value of the calculated coefficients depends on the weight attached to each character. The weight depends on the degree of stability or variability of a character. This means that the weight given to each character is related to the coefficient of variability (CV) of an unknown population. Therefore DITYL has the option open for the user to change the weight according to the CV of each character of the population that is to be identified.

Table 3. Identification of *Ditylenchus* spp. population with various weights attached to measured characters.

Identified species, population (no. of specimens)	Species	Identification results				
		Coefficients of similarity				
		version ² :	1	2	3	4
<i>D. adasi</i> ¹ Steżyca (16)	<i>D. adasi</i>		0.936	0.936	0.936	0.936
	<i>D. acutus</i>		0.639	0.639	0.638	0.638
	<i>D. singhi</i>		0.577	0.582	0.584	0.587
<i>D. equalis</i> ¹ Puszcza Białowiecka (10)	<i>D. equalis</i>		0.801	0.799	0.797	0.796
	<i>D. acuminatus</i>		0.739	0.747	0.751	0.755
	<i>D. parvus</i>		0.733	0.737	0.739	0.741
<i>D. longimatricalis</i> ¹ Bory Tucholskie (24)	<i>D. longimatricalis</i>		0.845	0.851	0.855	0.859
	<i>D. parvus</i>		0.670	0.679	0.683	0.689
	<i>D. acuminatus</i>		0.580	0.605	0.619	0.634
<i>D. longimatricalis</i> ¹ , Puszcza Biała (15)	<i>D. longimatricalis</i>		0.847	0.849	0.849	0.850
	<i>D. parvus</i>		0.687	0.693	0.696	0.699
	<i>D. acuminatus</i>		0.560	0.570	0.576	0.581
<i>D. medicaginis</i> ¹ Rewa (11)	<i>D. medicaginis</i>		0.908	0.914	0.918	0.923
	<i>D. valveus</i>		0.797	0.792	0.788	0.784
	<i>D. myceliophagus</i>		0.778			
	<i>D. tenuidens</i>			0.790	0.798	0.807

1. Measurements in Brzeski (1991)

2. Weight attached to character :

	version	1	2	3	4
with cv < 10		1.0	1.0	1.0	1.0
cv 10-20		1.0	0.8	0.7	0.6
cv 20-30		1.0	0.6	0.4	0.2

cv values of these populations are shown in Table 1.

The calculations of the coefficients of similarity for five known populations, attaching various weights, were conducted (Table 3). These calculations show high values calculated for the species being identified and very little effect of changing the weights on the final results. This is because the key is based on eleven characters and makes use of hierarchical division of the genus into four groups. Nevertheless, the open option for changing the weight of each character might facilitate identification. Comparison of data from Tables 1 and 3 shows the increase of the difference between coefficients of similarity of the first two species shown by DITYL in the case of *D. medicaginis*, the species with the highest mean CV of all characters used for identification.

References

- BOAG, B., TOPHAM, P. B., SMITH, P. & FONG SUN PIN, G. (1988). Advances in computer identification of nematodes. *Nematologica*, 34 : 164-172.
- BRZESKI, M. W. (1991). Review of the genus *Ditylenchus* Filipjev, 1936 (Nematoda : Anguinidae). *Revue Nématol.*, 14 : 9-59.
- FORTUNER, R. (1982). On the genus *Ditylenchus* Filipjev, 1936 (Nematoda : Tylenchida). *Revue Nématol.*, 5 : 17-38.
- FORTUNER, R. & WONG, Y. (1984). Review of the genus *Helicotylenchus* Steiner, 1945. 1 : A computer program for identification of species. *Revue Nématol.*, 7 : 385-392.
- GOWER, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrika*, 27 : 857-874.
- REY, J. M., ANDRES, M. F. & ARIAS, A. (1988). A computer method for identifying nematode species. 1. Genus *Longidorus* (Longidoridae). *Revue Nématol.*, 11 : 129-135.
- REY, J. M. & MAHAJAN, R. (1988). A computer method for identifying nematode species. 2. Genera *Tylenchorhynchus* and *Merlinius* (Nematoda : Tylenchina). *Revue Nématol.*, 11 : 269-277.
- STURHAN, D. & BRZESKI, M. W. (1991). Stem and bulb nematodes, *Ditylenchus* spp. In : Nickle W. R. (Ed.). *Manual of agricultural nematology*. New York, Marcel Dekker, Inc. : 423-464.