Nematology, 2000, Vol. 2(1), 17-30

Symposium

Genisys and computer-assisted identification of nematodes

Jim DIEDERICH [1], Renaud FORTUNER [2,*] and Jack MILTON [1]

[1] *University of California, Department of Mathematics, Davis, California 95616, USA*
[2] *La Cure, 86420 Verrue, France*

Presented at the symposium '*Biodiversity in the phylum Nematoda*', Gent, Belgium, 17 September 1999

Summary – Identification of species for the study of biodiversity in the phylum Nemata is defined as being able to identify any of the 12 000 described species of nematodes, and being able to recognise when the specimen under study belongs to a species unknown to science. Moreover, this large-scale identification must be done by biologists who are not expert taxonomists. It is argued that printed identification aids are too rigid and unreliable, and molecular identification aid is too limited, to be used in this context. Only computer identification tools are flexible, reliable, and general enough to give a non-taxonomist the possibility to identify any species. However, a review of existing tools and approaches shows that they fall short of reaching this objective. The functionality of a general identification system that could overcome the problems and difficulties encountered in the use of computerised tools is then described. This system (which has been described by the authors in several articles under the name of Genisys) would include a general database with all the characters from all the published descriptions of nematodes species and a set of integrated identification tools. Some of these tools would export the data into formats usable by existing identification and systematics tools. Other tools would have to be developed to support new approaches. A guidance tool would help the non-expert user by suggesting the successive computerised tools to be used during an identification session or by directing the user to other approaches, such as molecular identification in the case of heteroderids or some *Meloidogyne* species.

Résumé – ***Genisys et l'identification des nématodes assistée par ordinateur*** – L'identification des espèces pour l'étude de la biodiversité dans le phylum Nemata est définie comme la possibilité d'identifier n'importe laquelle des 12 000 espèces décrites du groupe et d'être capable de reconnaître que le spécimen étudié appartient à une espèce non encore décrite. En outre, ce type d'identification à grande échelle doit pouvoir être pratiqué par des biologistes qui ne sont pas des taxinomistes. Il est montré que les aides imprimées à l'identification sont trop rigides et trop peu fiables et que l'identification moléculaire est encore de portée trop limitée pour être utilisables dans ce contexte. Seuls les outils d'identification assistée par ordinateur sont suffisamment flexibles, fiables et généraux pour permettre à un non taxinomiste d'accomplir les identifications définies plus haut. Pourtant, une revue des outils existants et des approches utilisées montre qu'ils sont incapables d'atteindre l'objectif fixé. La fonctionnalité d'un système d'identification général capable de résoudre les problèmes soulevés par l'utilisation d'outils informatisés est ensuite décrite. Ce système (décrit par les auteurs du présent article sous le nom de Genisys) doit comprendre une base de données avec tous les caractères inclus dans les descriptions publiées des espèces connues et un ensemble d'outils d'identification intégrés. L'un de ces outils devrait permettre d'exporter les données sous des formats utilisables par les outils existants d'identification et de systématique. D'autres outils devraient être développés pour mettre en œuvre de nouvelles approches. Un outil de guidage aiderait le non-spécialiste en lui suggérant les outils informatiques à utiliser successivement dans le cadre d'une session d'identification ou en le dirigeant vers d'autres approches telles que l'identification moléculaire dans le cas de certains Hétérodérides et d'espèces de *Meloidogyne*.

Keywords – anatomy, biological characters, biological database, computerised identification tools, description, morphology, organs, representation, systems, uniformity.

Biodiversity is the study of the variety of living beings.

In light of this definition, the first step in the study of biodiversity is the identification of the various kinds of organisms present in the biotope being studied. For the study of biodiversity in the phylum Nematoda (or Nemata as argued by Maggenti *et al.*, 1987), we need to be able to identify nematodes.

For biodiversity, identification of nematodes is not limited to the recognition of a few economically important species but it means being able to put a name on any

* Corresponding author, e-mail: fortuner@wanadoo.fr

species observed. It also means being able to recognise that a specimen belongs to a new species, never described before.

In the present article, we will describe various identification methods, focusing more on the problems and difficulties raised by each method than on the details of the method itself, as these are available from many general reviews (Pankhurst, 1975, 1978; Fortuner, 1993a; Bridge *et al.*, 1998). We will see that, while all approaches and most tools implementing these approaches have value, none is able to allow a biologist to identify any form encountered in a sample.

We will then summarise the drawbacks of the existing efforts and describe the requirements for a general identification system. The concepts we will present here have been developed over the last 10 years by a team consisting of the present authors, two computer scientists from UC Davis (JD and JM) and one nematologist (RF). The project was called at first Nemisys (Nematode Identification System), but it was soon modified to Genisys (General Identification System) when we saw that our findings could readily be generalised to identification in any biological group (Diederich *et al.*, 1989, 1997, 1998, 1999; Diederich & Milton, 1989, 1991, 1993a,b; Fortuner, 1989b, 1993b; Diederich & Fortuner, 1996, 1998; Diederich, 1997). Genisys is currently a set of high-level principles and specifications for biological databases and identification. A summary of these principles and a list of articles already published on Genisys can be seen on the Web (http://math.ucdavis.edu/~milton/genisys.html). These principles have not yet been implemented into available tools, but the project is seeking funding.

The various approaches to identification and their limitations

The most ancient method of identification is instant recognition, and it remains the most practised form of identification by the general public. However, instant recognition quickly reaches its limits when we try to identify less well known species and as we discover and describe more and more species. With nematodes, instant recognition is still practised by all nematologists who are able to identify on sight the few species they study and by taxonomists who can recognise a larger number of species in the groups where they specialise. However, there is no nematologist who can recognise on sight each of the thousand species described, and the nematode taxonomist is himself an endangered and fast-disappearing species. Also, the main concern of taxonomists is to do taxonomy and they are not interested in identification *per se*, which means that taxonomists are not available to do the large scale identifications that would be needed for a comprehensive study of nematode biodiversity.

As a result, nematode identification must rely on some sort of identification aids, which can be classified into three basic categories: *i)* molecular identification methods, *ii)* printed identification aids, and *iii)* computerised identification aids. (There are other approaches, such as host-range differentials, but their scope of application is very limited.)

Molecular identification methods

Molecular identification certainly holds promise and it is probably the only possible approach to reliable identification in special cases, such as those *Meloidogyne* species that are almost identical to one another in morphology but have very different host-plant lists.

However, we are not concerned here with identification of a few economically important parasites but with identification for biodiversity purposes, which includes all species, known and unknown. In this sense, we do not see much hope for general molecular identification for two major reasons.

First, the definition of molecular markers for a particular species is a very long, difficult, and costly process. Understandably, most efforts have concentrated so far on the most economically significant species, most of them in the Heteroderidae family, a few aphelenchs, particularly the closely related species *Bursaphelenchus xylophilus*/*B. mucronatus*, some pratylenchids (particularly *Radopholus similis*/*R. citrophilus*), the *Xiphinema americanum* group, and a few other species (Table 1). This is very useful for plant nematology but we are far from our goal in the field of biodiversity.

Second, in order to characterise marker molecules in a species, it is necessary to start from well identified specimens: how are we going to identify them in the first place? This is not a joke or an idle question. It has been said that half of the contents of the various gene banks is useless because the molecular data applies to misidentified specimens. Even if molecular identification were a workable prospect for biodiversity, it would still need specimens identified by traditional means, that is, from morphological and anatomical data.

Table 1. *Number of studies on molecular identification as reported in* Nematological Abstracts *1993-1999.*

Taxa studied	Number of references found
Heterodera/Cactodera	63
Meloidogyne	59
Bursaphelenchus mucronatus/xylophilus	10
Pratylenchus spp.	7
Xiphinema (mostly *americanum* group)	6
Ditylenchus (mostly *D. dipsaci*)	5
Radopholus (*similis/citrophilus*)	5
Aphelenchoides spp.	2
Longidorus elongatus	1
Anguina tritici	1
Nacobbus aberrans	1

Printed identification aids

The second and oldest approach includes all identification aids that are printed on paper.

Printed aids in their overwhelming majority are dichotomous keys, but the use of keys often results in an erroneous answer. This is because keys proceed by elimination and they often use unreliable characters. A dichotomous key for a group of n species requires n-1 steps, which does not translate into n-1 characters because the same character can be used several times in the key. On the other hand, keys often include lines with more than one character so the total number of characters used may exceed the n-1 number of steps. The point is that keys to the species of medium to large genera require several dozen characters. Such a large number of reliable identification characters is not available in most nematode genera.

Keys are powerful tools that can be used safely as long as they rely on primary identification characters (Fortuner, 1989b), *i.e.*, characters that are: *i)* able to differentiate species or groups of species (there are clearly marked gaps in the data), and *ii)* recordable without risk of error (clearly seen, not ambiguous, not variable within a species or with well-defined variability).

The danger is that, in most nematode genera of medium to large size, the identifier quickly runs out of primary identification characters and elimination of species and groups of species must continue based on unreliable characters. Obviously, the result also is unreliable unless the identification is done by an expert. It is unfortunate that expert identifiers are fast disappearing: certainly there are not enough of them to run a full-scale biodiversity programme.

Keys do have their uses. They are very helpful to the persons who created them, because they have an intimate knowledge of the genus and they are able to avoid the pitfall of unreliable characters by using their own expertise. In fact, taxonomists continue to publish a large number of keys: keys are mentioned at least 72 times in *Nematological Abstracts* during the period we scanned for molecular identification studies (1993-1999).

Another limitation of a purely dichotomous key is that each successive dichotomy relies on a single character. If the character used at line L of the key cannot be seen, the identification cannot proceed past line L. Guessing the answer at line L is risky, with a good chance of reaching an incorrect answer, particularly if the identifier is not an expert in the genus.

These difficulties led to the second major approach to paper identification aids: tabular keys. There, elimination of species and groups of species can rely on several characters at a time, which somewhat protects against misuse of one unreliable character. In general, tabular keys work well at eliminating all but one group of a few species, using the first three or four characters at the left of the table. Tabular keys are easy to use with small genera but they quickly become too cumbersome when the number of species increases.

Another problem with identification aids printed on paper is that the user must follow the original choices made by the author of the key. If a user doesn't agree with the choice of the sequence of characters in a dichotomous key, or with the first few characters at the left of a tabular key, nothing can be done.

A final problem is that paper is a dead medium that cannot be updated. When new species are described, a new key needs to be printed, which rarely happens. Only an electronic medium offers the flexibility needed to give at least some control to the user and to make it easier to update a key. Computer identification also solves some of the problems we encountered with printed keys.

Computerised identification aids

Computer identification aids have existed since the 70's, but they are not being regularly used by people other than the authors of the various identification tools that have been created during the past 30 years. In particular, they are not being used by people who are not themselves specialised identifiers (basically taxonomists and laboratory technicians). In this sense, and in spite of very impressive achievements, it is a fact that computerised identification remains limited in its use. We need to look into

the reasons for this relative failure of computer identification and try to propose a way to solve the problems we'll uncover.

This evaluation will be presented as a short review of the computerised identification tools that have been proposed so far in plant nematology, not so much to describe the approaches used but to show why they failed to provide the perfect identification tool.

Table 2 is the result of a quick check of computer identification in CAB's *Nematological Abstracts*. This list may not be complete prior to 1993, but it gives an interesting overview of the field. It does not include the works on visual recognition of nematodes using computers (Fernandez-Valdivia *et al.*, 1989, 1992; Palhares-Melo & Bastos, 1997).

Similarity

Computer identification of nematodes started in 1983-1984 with two simultaneous projects, one by Boag and Smith (1983) (see review in Boag *et al.*, 1989), the other (called Nemaid) by Fortuner (1983) and Fortuner and Wong (1983; 1985). The two teams worked independently from each other, but they happened to select the same genus, *Helicotylenchus*, and the same approach, similarity, based either on a direct matching technique (Boag & Smith, 1983) or on the general coefficient of similarity of Gower (Nemaid). In these first two projects we see a good illustration of both the leap in reliability of computer identification and the limitations of this technique.

Table 2. *Major computer identification tools developed in plant nematology.*

Year	Authors	Name of tool	Approach	Scope
1983	Boag, B. & Smith, P.		Similarity (matching)	*Helicotylenchus*
1983	Fortuner, R.		Similarity coefficient (Gower)	Theoretical study
1983; 1985	Fortuner, R. & Wong, Y.	NEMAID	Similarity coefficient (Gower)	*Helicotylenchus*
1988	Rey, J.M. *et al.*		Similarity coefficients	*Longidorus*
1988	Rey, J.M. & Mahajan, R.		Cluster analysis	*Tylenchorhynchus* and *Merlinius*
1989	Tarjan, A.C.	MARINEMA	Multiple entry key	Marine nematodes
1989	Tarjan, A.C.	PARANEMA	Multiple entry key	Plant-parasitic nematode
1989	Fortuner, R. *et al.*	Nemisys	Expert workstation	Theoretical study
1989	Zullini, A. & Marganelli, G.	DORY	Key	Dorylaimida
1991	Fortuner, R. & Maggenti, A.R.		Discriminant Function Analysis	*Hirschmanniella oryzaelbelli*
1992	Ivezic *et al.*		?	*Pratylenchus*
1993	Fortuner, R. *et al.*	Nemisys	Expert workstation	Theoretical study
1993	Viscardi, T. & Brzeski, M.W.	DITYL	Similarity coefficient (Gower)	*Ditylenchus*
1995	Viscardi, T. & Brzeski, M.W.		Similarity coefficient (Gower)	*Cactodera*, *Globodera*, *Heterodera*, and *Punctodera*
1995	Gallsova, V.V. & Kulangieva, K.V.		Expert system	Freeliving nematodes
1996	Ye-WeiMin		Works spreadsheet	Nematodes
1997	Ryss, A.	Use of BIKEY 7	Elimination using Excel spreadsheet	Pratylenchidae; *Radopholus*
1997	Eisenback, J.D.		Reproduction of articles on CD-ROM	*Meloidogyne*

Both applications used a measure of similarity, which is something that is easily done with a computer but that is very tedious to do on paper. Basically, similarity between an unknown specimen and each of the various species in the genus is computed from a set of 20 to 25 characters, which is about as many as a dichotomous key would use, but these characters are considered all at the same time to compute a single measure of similarity, instead of one at a time as in a key. This means that the species are ranked based on all the available data. This somewhat protects against errors: if one mistake has been made in the recording and entry of 20 characters, the similarity computed is still 95% accurate. This is what is called graceful degradation: the system performance does decrease with errors, but the decrease is gradual, not catastrophic as with a key where one mistake leads to a completely wrong answer (or to a dead end).

Similarity is more reliable than dichotomy, and as such it is a big improvement over that method. However, similarity does not give a single answer, it only ranks the candidate species in order of resemblance to the specimen. Identification must end with just one name, so it is obvious that similarity alone is not enough and that some other method is needed.

In the early 80's, the two teams worked on the same group and used the same approach, but there was no concerted effort. Both teams developed their own database, which represents a wasteful duplication of efforts. Also, at least one of the systems (Nemaid) was not kept up to date and is no longer available. These are shortcomings that are often seen in the development of computer identification tools.

Gower's similarity coefficient has been used by other nematologists, in particular Rey *et al.* (1988) and Viscardi and Brzeski (1993; 1995). It is one of the most powerful of the dozens of similarity coefficients that have been proposed over the years and similarity tools are certainly very useful for identification, as long as they are not used alone.

Multivariate statistics

In 1988 we saw the introduction of a new approach, multivariate statistics with cluster analysis (Rey & Mahajan, 1988). This statistical approach is somewhat similar to similarity coefficients in the sense that all of the characters are used simultaneously to determine the relative proximity of the unknown to all the species considered, but the statistical multivariate approach is much more powerful than a simple matching coefficient and it makes it possible to give a measure of the distances between the various entities.

Like any other approach, a cluster analysis is only as good as the data it uses, and the results are affected by the choice of characters. Many statistical tools, in identification and other fields, have been proposed and have been used by people who had very little knowledge of the underlying assumptions. For identification, the fact that the specimen falls closest to species x does not necessarily mean that it belongs to this species. Wrong answers obtained with statistical systems and the reluctance of some nematologists to use a method they do not fully understand contributed to the relative failure of this approach.

Even with a good choice of characters and with no errors in the data entered, the specimen might belong to a new species, in which case the only right answer would be that there is no answer. This may be indicated by the cluster analysis (when the distance of the specimen to all species considered is statistically significant), but this would not be a definite answer, just a possibility (with only a certain percentage of chances that it is true). Lamberti and Ciancio (1994) used cluster analysis to study the relationships between species in the *Xiphinema americanum* group, but they concluded that this method is impractical for identification.

Fortuner and Maggenti (1991) used another multivariate approach, a discriminant function analysis (DFA), with better results to differentiate *Hirschmanniella oryzae* from *H. belli*. They succeeded in selecting seven characters and a multivariate space that could be used to identify an unknown specimen belonging to one of these two very similar species. Still, it took a 6-month study to develop this method for the differentiation of only two species. This is certainly not an approach that could be extended to the thousands of existing nematode species.

Computerised keys

The next two tools, MARINEMA and PARANEMA, developed in 1985 by Tarjan (1989), are good examples in nematology of what became the most commonly used approach to computer identification — the computerisation of the traditional dichotomous key. Computers do allow for a major improvement over printed keys: multiple entry. The user can now select which characters to enter, and in which sequence. Hopefully, he will use reliable characters first. This does not solve the problem of lack of a sufficient number of reliable identification characters, but it reduces the risk of early catastrophic error. Computers also make it possible to introduce some measure of degradation in that some systems require two (or more) character

mismatches before rejecting a species (Pankhurst, 1993). The user is allowed to make one or more mistakes before the identification fails. This is not true graceful degradation (when the user enters one error too many, the system fails totally) but the risks of a wrong answer are indeed reduced. However, this approach requires even more characters than a regular key, which means that the probability of making errors is in fact increased.

Multiple entry keys (aka polyclaves) are very much in favour inside and outside nematology. Introducing the 1996 Cambridge symposium on information technology, plant pathology and biodiversity, Scott (1998) gives a list of 11 'diagnostic tools': all are various sorts of keys (multi-entry keys, graphical keys or dichotomous keys).

Graphical keys are a special subclass of dichotomous keys where the choices are presented in the form of images illustrating the various options instead of text, as in traditional keys. A good example of graphical keys for nematology can be seen in the web page (http://ianrwww.unl.edu/ianr/plntpath/nematode/key/nemakey.htm) prepared by Tarjan, Esser and Chang.

Graphical keys can be used in two modes. The user can browse through the key, hoping to see an image that 'looks like' the specimen to be identified, or can use the graphical key in a dichotomous manner, in which case the graphical key is only a variant of the dichotomous key, with data presented in a graphical instead of a textual way. Graphical keys are aesthetically pleasing and they elicit great initial interest from potential users, but they suffer from the same limitations as the other elimination tools.

Outside nematology, graphical keys are used by ETI, a general identification project based in Amsterdam that plans to publish identification aids on CD-ROMs for various biological groups (Schalk, 1998). In spite of their polished appearance, ETI identifications aids rely on a dichotomous approach, with all the limitations and dangers of this method.

Table 2 shows that keys have been used extensively in nematology, either as stand-alone programs such as DORT (Zullini & Marganelli, 1989) or as applications built on top of commercial spreadsheets (Yo-WeiMin, 1996; Ryss, 1997a, b). It would seem that biologists who create computer identification systems tend to use the electronic medium to do what they used to do on paper. In some cases, printed descriptions and keys are bodily transferred to an electronic medium, such as the CD-ROM on *Meloidogyne* systematics and identification prepared by Eisenback (1997).

It may be that it takes a computer scientist to take full advantage of the potential of computers and the next three approaches have been developed outside of biology.

Expert systems

The term 'expert system' is used with many different meanings. Pankhurst (1993) calls expert systems applications such as his own PANKEY, Dallwitz's CONFOR or Lebbe's XPER although they are not based on rules but on a data matrix.

A true expert system is based on rules, such as the one proposed in Table 3, entered into a general-purpose application called an expert system shell (which includes the expert system engine but no rules). We have not been able to locate any rule-based identification systems in nematology. What was originally conceived as an expert system *sensu stricto* by Diederich and Milton (1989) evolved into the more suitable concept of expert workstation described by these authors in Diederich and Milton (1993a, b). The 'expert system' of Galtsova and Kulangieva (1995) is a complete system with some built-in expertise, but its identification approach is based on a polytomial key.

It would be theoretically possible to design a tool that would construct the rules automatically from the published diagnosis, as most expert system shells include verification functions that check the pertinence and consistency of the rules and make sure there are no incompatibilities.

However, the example in Table 3 shows that an expert system is a fancy way to write a key. It is true that it offers several important functions unknown in a key: it uses all characters simultaneously instead of sequentially, it provides extra functions such as the ability to explain to some extent its reasoning in a diagnosis or identification, and it gives a degree of confidence it has in the answer.

Table 3. *Example of a rule for a hypothetical expert system for the identification of* Pratylenchus *spp.*

If
it is a *Pratylenchus*
it has two lip annuli
it has an empty spermatheca (males absent)
it has a smooth tail end
it has four lines in the lateral field
it has a stylet more than 19 μm long
it has a vulva at 82-89% of body length
it has low angular lip region
Then
it is *P. brachyurus*

However, it would be difficult to differentiate from the interface alone a rule-based expert system from a matrix-based multiple entry key. In both cases the system asks the user to enter the characters it needs, and in both cases it is susceptible to faulty data.

Neural networks

Neural networks work in a very different way since they are modelled on connections of nodes in a manner suggested by the connection of neurones in the brain. Typically, a neural network includes several levels of nodes, often three levels, and the nodes in one level are all interrelated to all the nodes in the next level(s). Neural networks can be trained to find correct answers (using a 'training data set') or they can learn by themselves by trying to find some pattern in the training data set.

Reviewing neural networks, Boddy *et al.* (1998) conclude that, while this approach has a huge potential for identification, it should not be considered as a panacea because of *i)* limitations in the data (if there is no pattern within the data used to train the network; it cannot be expected to discern one!), *ii)* lack of suitable data sets for training (suitable data are data that reflect the extent of individual variation), and *iii)* the necessity to enter the specimen values for all the characters that were used to train the network.

However, the main drawback of neural networks is that it is not clear how they reach their conclusions. In particular, the choice of the number of nodes in the middle layer is selected from hit and miss experiments rather than from design principles and logic. This 'black box' aspect makes most biologists wary of neural networks, all the more so when they consider the performance level stated by proponents of this approach: Boddy *et al.* (1998) indicate 10-20% wrong rejections of known taxa. As neural networks are not always 100% right, and as their logic is not clear to the user, they are not the general identification tool we are looking for, particularly when used alone.

Bayes rule and probabilistic systems

Bayesian systems are based on Bayes rule, which is a mathematical way to identify the species to which a specimen belongs. The rule says that the probability of having species x given the available evidence (character data) is equal to the probability of observing the evidence if we are in fact looking at species x, times the probability of observing species x in a sample, divided by the probability of observing the evidence:

$$P(Species/Evidence) = P(Evidence/Species).P(Species)/P(Evidence)$$

This rule is said to be basic to all identification approaches. For example, many biologists are reluctant to use probabilities in species identification, but a deterministic approach such as a dichotomous key is in fact using Bayes rule with all probabilities equal to 100%.

Problems crop up when Bayes rule is used, not as a theoretical framework, but for doing practical identifications. One of the stumbling blocks in the real world is the definition of the *a priori* probability for observing a particular species — *P*(Species) in Bayes rule. Walker (see Appendix A in Fortuner, 1993a), explains that 'if I have to identify a nematode from California, the prior probability that it belongs to a species known only from Manchuria is low'. Prior probabilities can be given when identifications are made in a well known environment by experts who have already done many identifications in this context and who know what to expect. However, the study of biodiversity requires a system able to recognise all the species present in any environment, including biotopes never studied before. Walker indicates that 'if you do not know prior probabilities, you say they are all equal', but many biologists will contend that there is no point in using for practical identification a method that relies on precisely defined probabilities if it is impossible to give these probabilities and if artificial values (all priors equal) are used. In answer to Walker, Pankhurst said that 'You are not going to get probability data from most biologists. We can't get it for you and we don't need to, we can do better than that. We don't need rules, we can't give you probabilities, but we can provide data matrices' (Appendix A in Fortuner, 1993).

After this brief review of the existing approaches to computer identification and the difficulties that are apparent to those who try to use them, we can now look at these problems and try to propose ways to resolve these difficulties. These problems can be separated into two categories: those that are linked to the database and those that relate directly to the identification tools themselves.

Problems associated with computer identification and possible solutions

Problems linked to the database

The major problem is that, so far, each database linked to a particular tool includes only a small subset of characters as selected by the author of the tool. Not only does this

limit the freedom of the user for selecting the characters to be used but it causes duplication of effort. Even when the same approach is followed for several tools the authors of the different tools often select different sets of characters. This means that a separate and independent database must be created for each tool.

Obviously, these character sets are more different when different identification approaches are used, and even more different when we compare identification databases to taxonomic databases. While it is true that identification characters and taxonomic characters should be selected according to different criteria, it remains that many characters can be used in both cases. Selection of a small character set for each tool means that separate databases must be created each time at great cost.

A second problem is that the format under which the characters are recorded is far from uniform. In fact, each author is free to create his own format. For example, even quantitative data can be recorded in many different ways: as measurements with or without statistical parameters, as approximate values, as fuzzy values, or as coded classes of data (Table 4).

Most character formats do not support metadata (data about the data) nor general statements about the data. An example of a general statement would be: 'in a cylindroid organ, the width in lateral view is equal to the diameter in cross section'. General statements and metadata are needed for embedding expertise in the identification tools, which is the only way we can create identification tools that can be used by non-taxonomists.

Most identification databases are not kept up to date. They are often created by one author or one team for a specific project, then the author or team generally goes on to other projects and has neither the time nor the interest in the large effort needed to add new descriptions as they are published.

Table 4. *Various ways to record the same character: body length of a hypothetical nematode species.*

Body length	
Approximate value	*ca* 750 μm
Range	658-840 μm
Statistical value	mean 747 μm standard deviation 38 μm
Class	2-3 (with class 2 = 500-750 μm and class 3 = 750-1000 μm)
Fuzzy value	body long

POSSIBLE SOLUTIONS TO DATABASE PROBLEMS

For nematode biodiversity studies, we need to be able to identify any nematode encountered, which means that we need a database that holds the description of all the species described in this biological group. This database must solve the database problems we just described.

This means that:

- The characters in the database must not be selected beforehand. Instead, all characters that have been used and described for any purpose, be it taxonomy, identification or just description, must be included.
- The characters must be usable by any identification program, which of course means that they must be reformattable into formats usable by the various programs. This means that the general database must record the characters in their most basic form, because it is easier to create a complex character from basic data than the other way around, even if the latter were always possible.
- The database must include descriptions of all the species in the group considered. These descriptions must be not only at the species level (*e.g.*, the data found in published species descriptions) but also at the population level (*e.g.*, descriptions of other populations of existing species) and at the specimen level (actual data used by taxonomists for describing species and populations). It is true that taxonomists have the weird habit of spending a lot of time and efforts in gathering data, and then of throwing it all away after they have published a summary account of these data (in the form of species descriptions), but this tradition must be changed if we are serious about the study of biodiversity. To study diversity, diversity must be recorded, and diversity exists not just between one species and the next, but also between populations of each species and between specimens of each population. The general database must serve as a depository for specimen descriptions, as this is the best way to account for the whole extent of specific diversity.
- The database must support metadata and general statements. The only way to cope with the declining population of experts is to capture their expertise and record it in the database so it can be embedded into the identification tools we create.
- The database must be managed by a person or a group of persons specifically appointed for this purpose.

IMPLEMENTATION OF THE GENERAL DATABASE

Before a general database can be created properly, two questions must be answered: *i)* what data should be

recorded? and *ii)* under what format should the data be recorded?

Source of data

In spite of several drawbacks, the only way to create a general database is to use published data (Diederich *et al.*, 1999). These drawbacks are:

- Published data are only as reliable as the authors of each publication — but this is true also of new data that would be gathered for the purpose of creating the database: even if it were possible to gather new data for the general database, there is no guarantee that it would be more reliable than at least some of the published data.
- Each publication includes values for only a small fraction of all possible characters, typically 50-80 characters out of several thousand possible characters (nematodes include about 227 structures described by up to 20 properties), which means that there are many missing data in published descriptions. It should be noted that it would be impossible in practice to record thousands of characters for each of the 12 000 or so species of nematodes. Also, missing characters that are needed for a particular application can always be recorded later and added to the database. This would ensure that the character would no longer be missing the next time it is needed.
- There is no uniform format in published descriptions, but it is possible to create uniformity by reformatting the characters.

The advantages of using published data are:

- It is an easy way to start a general database housing basic descriptions for the thousands of existing species. More data can be added later but we can start with a complete database (in the sense that it will include all the described species).
- It is possible to extract, reformat, and store data much faster than it would take to find, process, measure, and record characters from new specimens. Using an extraction/formatting tool such as Terminator (Diederich *et al.*, 1999), one species description can be processed in less than 1 hour (including 0.5 hour for putting the printed description into an electronic format and half an hour for the actual data extraction), whereas it would take several weeks to obtain and process new specimens.

A general format for the data

The data extracted from published descriptions must be put into a single format and this format must be a very basic one. This can be done by using an already existing format or a new format.

For identification, a 'descriptive language for taxonomy', or DELTA, has been developed by Dallwitz (1980) and it is used by many biologists, mainly botanists. Using DELTA, any character can be coded in a uniform way. For example, in a Delta code for Graminae, if character number 5 refers to culms (whether woody or herbaceous) and state number 1 to 'woody and persistent', this character will be coded '5,1' for the genus *Phragmites* (Dallwitz, 1993). The major problem with DELTA is that, while it is true that the code itself is uniform and can be used by generic identification programs, nothing has been done to enforce uniformity in the characters represented by the codes. Another author could very well have proposed state 1 for 'woody' and state 2 for 'persistent'. To support general statements about the data, it is necessary that the characters themselves are uniform. In addition, DELTA does not support metadata (except as 'comments', which cannot be easily exploited by applications).

In computer science, data is classically represented as entity-attribute-value. This was used, *e.g.*, by Lebbe (1991) to represent taxonomic characters. The decomposition of traditional characters into three basic components makes it theoretically possible to support general statements. Still, this decomposition alone is not sufficient to assure uniformity of characters. For example, Lebbe (1991) decomposed the character 'number of teeth of the pod' into an entity — 'pod' — and an attribute — 'number of teeth'. This 'attribute' is in fact a mixture of an entity (teeth) and a true attribute (number).

Strict rules and guidelines are necessary to create a truly uniform character set and to make sure that non-uniformities do not creep back into the database. Diederich *et al.* (1997) proposed to follow a strict decomposition into organs, properties and states or values. Diederich (1997) determined that all morphological-anatomical properties we have encountered to date belong to a small set of about 20 'basic properties' that are the same for all organs in all nematode species and in fact in any biological species. Other rules and guidelines are needed for the definition of a list of structures, which includes the organs and organ parts arranged under traditional systems.

The task of extracting and reformatting the data from published descriptions and redescriptions of all nematodes species can be seen as daunting, but Diederich *et al.* (1999) have demonstrated proof of concept for an electronic tool that they designed for that purpose. The prototype tested (called Terminator because it works from terms) was able to extract and reformat the data in about 0.5 hour per description. Currently, there are about 12 000

described species of nematodes plus a number of re-descriptions of known species that might double this total. Processing 24 000 descriptions would take 12 000 h, which is about 12 person-years. This is a significant amount of time, but it can be done in stages (focusing first on the major groups) and it can be entrusted to several persons who are not necessarily experienced taxonomists (because the final version of Terminator would have embedded expertise that can assist the operator). The end results would be well worth the effort, as the resulting database could be used by any identification tool as well as purely taxonomic tools.

The database must also include metadata, *i.e.*, data about the data (Diederich *et al.*, 1989; Diederich & Milton, 1991) and support relationships between data at the entity, property, and state or value levels. An example of the use of metadata is provided by the endorsement of the data entered by the user. Any identification tool or approach is only as reliable as the data it uses. Therefore, the application must be able to evaluate the reliability of the data. Diederich and Fortuner (1996) have proposed an endorsement factor based on the values of a dozen types of metadata. The endorsement factor can be used, *e.g.*, to select those characters that are reliable enough to be trusted to eliminate candidate species (using a key) or to weigh characters for the computation of a similarity coefficient. Primary identification characters (Fortuner, 1989) are characters that are easy to see (organ well visible, property not ambiguous) and that are not variable (or with well defined variability) within the taxon while making it possible to differentiate between separate taxa. Visibility, ambiguity, variability are some of the dozen or so types of metadata proposed by Diederich *et al.* (1989).

Finally, the administration of the database must be entrusted to a person or a team for which this would constitute one of the primary duties. This would ensure that the database is permanently kept up-to-date. The administrator would have to enter each new published description. This would represent a few hundred descriptions per year or about one person-month of work. The database could also serve as a repository for individual specimen descriptions provided by authors before publication. Also, every time the database is used during an identification session, the user would have to enter the description of the specimens to be identified. These data can be saved, validated, then added to the list of descriptions. At the moment, a majority of nematode species have been described from their type population only. This is not good enough to study biodiversity but a general database would certainly boost the quantity of data available. The database administrator would have the duty to ensure the reliability of the data with the assistance of an editorial board composed of eminent taxonomists. Because the creation of the database is such a huge job it will need special funding, which means there will be resources available to keep it up to date.

A final point is that the database should be available online over the Internet, which would assure a general and world-wide distribution of knowledge.

Problems linked to identification tools

The difficulties that hinder the development of computerised identification will not all be solved by a general database, even if it is based on sound principles. As we have seen, a number of difficulties stem from the identification tools themselves.

The first and most significant drawback seen in all existing tools is the fact that each tool relies on a single approach, be it elimination (keys), comparison (similarity coefficients), or other approaches such as expert systems, neural networks, or Bayesian systems. A typical identification session uses in fact several approaches to reach a conclusion. The identifier first uses instant recognition to select a group of candidates, which can be anything from a family or order to a genus or even to a possible species. Then, if the group recognised on sight includes several species, the identifier eliminates the species that are obviously different from the specimen. If only reliable characters (primary identification characters) are used, chances are that this elimination process will not end with a single species and the remaining candidates must be processed further using other approaches. In theory, it is possible to use several of the existing tools during a particular identification, but these tools are not compatible and the data would have to be entered again each time (plus the partial results obtained with one tool would not be available with the next tool).

Second, the approaches and tools used do not depend only on the circumstances of each identification session. We have seen that some biologists strongly reject any probabilistic tool while others argue that Bayes' rule is the only game in town. A general identification system must accommodate the preferences of all users, which means that no single approach should be forced onto all users.

The most common approach — elimination — also carries the greatest risk of reaching a wrong answer. In spite of specific improvements made possible by computerised

keys, this very powerful method must be made more reliable.

Computer tools are often seen as slower and requiring more data entry than corresponding paper aids. For example, Tardivel and Morse (1998) compared a printed dichotomous key to the computerised version of the same key and to a multiple entry key. They found that both computer keys were slower than the paper key (because many more choices had to be made before an answer could be reached) and that, while the multi-access key was slightly more accurate than the printed key, the other computer key was the least accurate of the three. Understandably, most people are reluctant to spend more time reaching a less accurate answer!

Solutions to identification tool problems

A general identification system must not rely on a single approach but must help the identifier with all possible approaches by providing him with a set of tools (Diederich & Milton, 1989, 1993). The user will then be free to select the tool deemed most appropriate at each step of an identification session (but a guidance tool should also be available upon request to recommend a course of action to less experienced users). The various tools must be integrated, in the sense that the data and the results from one tool must be available with all the other tools.

The set-of-tools approach will solve the problem caused by the reluctance of some biologists to use unfamiliar methods. When all possible approaches are available, each user will be free to select those liked best. Neural networks, Bayesian systems, *etc.*, will be included (if and when the necessary training data or priors have been obtained), but other more traditional approaches will also be available.

The problems linked to the use of unreliable data, particularly with non-degradable approaches such as dichotomous keys, can be solved in two ways: first, an evaluation system such as the endorsement factor could be used to select only reliable characters for degradation-sensitive tools. Second, the availability of other approaches would make it unnecessary to continue using a key when all reliable characters have been exhausted.

The set-of-tools approach also obviates the problem linked to the large amount of data required by some systems. The experienced user of a printed key can proceed directly to the part of the key that holds the likely answer. On the contrary, most computerised keys do not offer such short cuts, and the user must enter data that are needed only to eliminate the non relevant portions of the set of species. Using a tool that supports instant recognition would provide electronic short cuts equivalent to the practice of experts using a familiar printed key.

The concept of 'promorph' (Fortuner, 1989b) can be used to implement these short cuts. A promorph is a form that can be recognised at first glance, before observation of detailed morphology. The definition of promorphs depends on the level of expertise of the user. Where the general public will see only a 'worm', any plant nematologist can recognise a 'tylench' (stylet-bearing nematode), and an expert taxonomist can say that the specimen is, *e.g.*, a 'helico' of even a member of the '*dihystera*' promorph. Each promorph is associated with a list of possible species that are quite independent of taxonomical considerations. For example, the promorph 'scutello' includes not only the species in the genus *Scutellonema*, but also members of other genera that, at first glance, may be confused with this genus, *e.g.*, *Helicotylenchus vulgaris* (Fortuner, 1991).

The tools of a general identification system

Export tool

The first tool developed for a general identification system must be an export tool able to extract selected characters from the general database and reformat them so they can be used by existing identification tools. For example, a DELTA format export function would make it possible to use the various DELTA tools. Other export functions would make it possible to use the general database with various taxonomic tools. This would ensure that the implementation of a general identification system would not mean that all the time and effort spent over the years in developing tools that all have some use, albeit limited, have been wasted.

Also, this would assure that the general identification system is at least as good as the best available independent tool. If a powerful new method is ever published, better in some aspect than all existing tools of the general identification system, a new export function would make it possible to use the general database with this new tool.

Genisys tools

Including an export tool in the general identification system would give access to all the approaches that have been used in the past for identification and taxonomy, but it would not solve all the problems we listed. For this, we would need new tools and new functions, specifically designed and built for this purpose.

For example, a tool based on instant recognition could be used to narrow the field of candidate species to a particular promorph. If the user is an expert in the field, this promorph can be defined at the lowest level, even a particular species. This would cut down tremendously the amount of data entry needed and it would make sure that only relevant data is required from the user (in fact, this could be used to enhance the existing tools: if an expert claims to recognise species x, it can be assumed that the primary identification characters of species x are in fact present in the specimen, and an export tool could enter these characters in, say, Pankhurst's PANKEY as if they were characters entered by the user).

The set of tools should implement some approaches that need very little input from the user (but input-intensive approaches must be available as well, if needed). For example graphical identification, browsing textual and graphical descriptions, and instant recognition should be supported.

An elimination tool using a multi-entry key or expert system approach is a very powerful way to eliminate all the species obviously different from the specimen, as long as it is used with primary identification characters only. A new elimination tool could be built based on the endorsement factor we described earlier. This could be used to select a nest of species, which is a group of species that share the same set of primary identification characters (Fortuner, 1989b).

After elimination of the obviously irrelevant species, a comparison tool could rank the remaining species according to their similarity with the specimen, as measured by various similarity coefficients. With a browsing tool the user could look at the actual descriptions (by text or images) of the top-ranked species. Finally, a diagnostic tool could verify that all the key characters for the selected species are present with the correct state or value in the specimen.

Guidance tool

Every identification session is different. An expert may recognise a species at first glance and go directly to the diagnostic tool to verify that his hunch was right. At the opposite end, a beginner may have no idea where to start and ask the help of the guidance tool. This tool would operate in the background and keep track of the history of each identification session. Based on the point reached by the user, it would propose the most logical next step using one of the available computer tools or directing the user to an alternative method such as molecular identification for the few species where identification markers have been developed. This tool would probably include expert system rules and heuristics. However, the user must always be in charge and must have full freedom for the selection of the tools, the data to enter, the candidate species to consider, the threshold values, and all the other aspects of identification. At the same time, the user would also be able to call the guidance tool for help, if so desired.

Obviously, the data and the partial results obtained with one tool would be available with the other tools, as they would all be integrated within a single system.

Conclusion

Will the Genisys concepts and tools be sufficient to make general practitioners in biology use electronic identification? Only their implementation can answer this question, but we believe that necessity caused by increasing problems in managing biodiversity studies will dictate that scientists use the best tools possible or fall behind. There is sometimes a lag between the availability of a method and its widespread use, but we believe that pressing problems will yield no choice.

It should be noted that a Genisys set of tools would include all available approaches and this should help even an expert identifier do better. This alone should ensure a wider usage. Non-expert users may be reluctant at first to trust themselves to do identifications, but the disappearance of taxonomists in many laboratories will force them to take the plunge. It is hoped that they will find the guidance tool friendly and helpful and that they will obtain correct results.

If the biodiversity of nematodes is to be studied, nematode identifications will have to be performed on a large scale, but traditional identifiers (nematode taxonomists) will not be available for this duty. Existing printed, molecular, and computerised identification aids cannot be trusted to non-taxonomists for identification of all of the specimens collected. Only a general identification system such as the one outlined here can solve this problem.

References

BOAG, B. & SMITH, P. (1983). Computer assisted identification of nematodes. *Systematic Parasitology* 5, 235-239.

BOAG, B., TOPHAM, P.B., BROWN, D.J.F. & SMITH, P. (1989). The use of microcomputers for the identification of plant-parasitic nematodes. In: Fortuner, R. (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 9-18.

BODDY, L., MORRIS, C.W. & MORGAN, A. (1998). Development of artificial neural networks for identification. In: Bridge, P., Jeffries, P., Morse, D.R. & Scott, P.R. (Eds). *Information technology, plant pathology and biodiversity*. Wallingford, UK & New York, NY, USA, CAB International, pp. 221-231.

BRIDGE, P., JEFFRIES, P., MORSE, D.R. & SCOTT, P.R. (1998). *Information technology, plant pathology and biodiversity*. Wallingford, UK & New York, NY, USA, CAB International, xiv + 478 pp.

DALLWITZ, M.J. (1980). A general system for coding taxonomic descriptions. *Taxon* 29, 41-46.

DALLWITZ, M.J. (1993). DELTA and INTKEY. In: Fortuner, R. (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 287-296.

DIEDERICH, J. (1997). Basic properties for biological databases: character development and support. *Journal of Mathematical and Computer Modelling* 25, 109-127.

DIEDERICH, J. & FORTUNER, R. (1996). Endorsement of observations in identification. *Fifth IEEE International Conference on Fuzzy Systems, 8-11 September 1996, New Orleans, LO. USA*, pp. 175-179.

DIEDERICH, J. & FORTUNER, R. (1998). Classification using small fuzzy biological data sets. *IEEE International Conference on Fuzzy Systems, 4-9 May 1998, Anchorage, Alaska, USA*, pp. 1429-1434.

DIEDERICH, J., FORTUNER, R. & MILTON, J. (1989). Building a knowledge base for plant-parasitic nematodes: description and specification of metadata. In: Fortuner, R., (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 65-76.

DIEDERICH, J., FORTUNER, R. & MILTON, J. (1997). Construction and integration of large character sets for nematode morpho-anatomical data. *Fundamental and Applied Nematology* 20, 409-424.

DIEDERICH, J., FORTUNER, R. & MILTON, J. (1998). A general structure for biological databases. In: Bridge, P., Jeffries, P., Morse, D.R. & Scott, P.R. (Eds). *Information technology, plant pathology and biodiversity*. Wallingford, UK, CAB International, pp. 47-58.

DIEDERICH, J., FORTUNER, R. & MILTON, J. (1999). Computer-assisted data extraction from the taxonomical literature. *Virtual publication on http://math.ucdavis.edu/~milton/genisys.html*.

DIEDERICH, J. & MILTON, J. (1989). NEMISYS, an expert system for nematode identification. In: Fortuner, R. (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 45-63.

DIEDERICH, J. & MILTON, J. (1991). Creating domain specific metadata for scientific data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering* 3, 421-434.

DIEDERICH, J. & MILTON, J. (1993a). Expert workstations: a toolbased approach. In: Fortuner, R. (Ed.). 1993. *Advances in computer methods for systematic biology: artificial intelligence, databases, computer vision*. Baltimore, MD, USA, The Johns Hopkins University Press, pp. 103-123.

DIEDERICH, J. & MILTON, J. (1993b). NEMISYS: an expert workstation, a computer science perspective. In: Fortuner, R. (Ed.). 1993. *Advances in computer methods for systematic biology: artificial intelligence, databases, computer vision*. Baltimore, MD, USA, The Johns Hopkins University Press, pp. 165-179.

EISENBACK, J.D. (1997). *Root-knot taxonomic database*. Wallingford, UK, CAB International, CD-ROM.

FERNANDEZ-VALDIVIA, J., CASTILLO, P. & GOMEZ-BARCINA, A. (1989). Use of image analysis for automatic recognition of nematodes. In: Fortuner, R., (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 293-299.

FERNANDEZ-VALDIVIA, J., PÉREZ DE LA BLANCA, N., CASTILLO, P. & GOMEZ-BARCINA, A. (1992). Line detection and texture analysis for automatic nematode identification. *Journal of Nematology* 24, 571-577.

FORTUNER, R. (1983). Computer assisted semi-automatic identification of *Helicotylenchus* species. *California Plant Disease Reporter* 2, 45-48.

FORTUNER, R. (1989a). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, ix + 386 pp.

FORTUNER, R. (1989b). A new description of the process of identification of plant-parasitic nematode genera. In: Fortuner, R. (Ed.). *Nematode identification and expert system technology*. New York, NY, USA, Plenum Publishing Corp, pp. 35-44.

FORTUNER, R. (1991). The Hoplolaiminae. In: W.R. Nickle (Ed.). *A manual of agricultural nematology*. New York, NY, USA, Marcel Dekker, pp. 669-719.

FORTUNER, R. (1993a). *Advances in computer methods for systematic biology: artificial intelligence, databases, computer vision*. Baltimore, MD, USA, The Johns Hopkins University Press, xiv + 560 pp.

FORTUNER, R. (1993b). The NEMISYS solution to problems in nematode identification. In: Fortuner, R. (Ed.). *Advances in computer methods for systematic biology: artificial intelligence, databases, computer vision*. Baltimore, MD, USA, The Johns Hopkins University Press, pp. 137-164.

FORTUNER, R. & MAGGENTI, A.R. (1991). A statistical approach to the objective differentiation of *Hirschmanniella oryzae* from *H. belli* (Nemata: Pratylenchidae). *Revue de Nématologie* 14, 165-180.

FORTUNER, R. & WONG, Y. (1983). *NEMAID. Computer program for identification of nematodes. User's manual*. Sacramento, CA, USA, California Department of Food and Agriculture, Publication No. 640, 44 pp.

FORTUNER, R. & WONG, Y. (1985). Review of the genus *Helicotylenchus* Steiner, 1945. I. A computer program for

identification of the species. *Revue de Nématologie* 7, 385-392.

GALTSOVA, V.V. & KULANGIEVA, K.V. (1995). Expert system for identification of freeliving nematodes. *Ninth International Meiofauna Conference, Perpignan, France, 3-7 July 1995*, 9, 62.

IVEZIC, M., VUKADINOVIC & HORVAT, D. (1992). [A computer program for the determination of the nematode genus *Pratylenchus* (Filipjev, 1936).] *Znanost i Praksa u Poljoprivredi i Prehrambenoj Tehnologiji* 22, 191-197.

LAMBERTI, F. & CIANCIO, A. (1994). The relationships between species within the *Xiphinema americanum* group (Nematoda: Dorylaimida). *Bulletin OEPP* 24, 475-484.

LEBBE, J. (1991). *Représentation des concepts en biologie et en médecine. Introduction à l'analyse des connaissances et à l'identification assistée par ordinateur*. Thèse de doctorat, Université Pierre et Marie Curie, Paris, xii + 282 + xxiv pp.

MAGGENTI, A.R., LUC, M., RASKI, D.J., FORTUNER, R. & GERAERT, E. (1987). A reappraisal of Tylenchina (Nemata). 2. Classification of the suborder Tylenchina (Nemata, Diplogasteria). *Revue de Nématologie* 10, 135-142.

PANKHURST, R.J. (1975). *Biological identification with computers*. London, UK, Academic Press, 333 pp.

PANKHURST, R.J. (1978). Biological identification. In: *The principles and practices of identification methods in biology*. London, UK, Edward Arnold, 104 pp.

PANKHURST, R.J. (1993). Principles and problems of identification. In: Fortuner, R. (Ed.). 1993. *Advances in computer methods for systematic biology: artificial intelligence, databases, computer vision*. Baltimore, MD, USA, The Johns Hopkins University Press, pp. 125-136.

DE PALHARES-MELO, L.A.M. & BASTOS, R.C. (1997). A preliminary study for constructing a computational procedure for nematodes identification based on morphological aspects. *Proceedings of the 1st European Conference for Information Technology in Agriculture, Copenhagen, Denmark, 15-18 June 1997*, pp. 405-408.

REY, J.M., ANDRES, M.F. & ARIAS, M. (1988). A computer method for identifying nematode species 1. Genus *Longidorus* (Nematoda: Longidoridae). *Revue de Nématologie* 11, 129-135.

REY, J.M. & MAHAJAN, R. (1988). A computer method for identifying nematode species 2. Genera *Tylenchorhynchus* and *Merlinius* (Nematoda: Tylenchina). *Revue de Nématologie* 11, 269-277.

RYSS, A. (1997a). Computerized identification of species of the genus *Radopholus* (Tylenchida: Pratylenchidae). *Russian Journal of Nematology* 5, 137-142.

RYSS, A. (1997b). Computerized identification of the plant nematodes of the family Pratylenchidae. *Proceedings of the 49th International Symposium on Crop Protection, Gent, Belgium, 6 May 1997, Part III*, 62, 3a, 701-711.

SCHALK, P.H. (1998). Archiving biodiversity: information technology applied to biodiversity information management. In: Bridge, P., Jeffries, P., Morse, D.R., & Scott, P.R. (Eds). *Information technology, plant pathology and biodiversity*. Wallingford, UK & New York, NY, USA, CAB International, pp. 213-220.

SCOTT, P.R. (1998). The incredible pace of change: information technology in support of plant pathology. In: Bridge, P., Jeffries, P., Morse, D.R. and Scott, P.R. (Eds). *Information technology, plant pathology and biodiversity*. Wallingford, UK & New York, NY, USA, CAB International, pp. 1-13.

TARDIVEL, G.M. & MORSE, D.R. (1998). The role of the user in computer-based species identification. In: Bridge, P., Jeffries, P., Morse, D.R. and Scott, P.R. (Eds). *Information technology, plant pathology and biodiversity*. Wallingford, UK & New York, NY, USA, CAB International, pp. 247-259.

TARJAN, A.C. (1989). Multiple entry identification keys. In: Fortuner, R. (Ed.). *Nematode identification and expert system technology*. New York, USA, Plenum Publishing Corp, pp. 27-30.

VISCARDI, T. & BRZESKI, M.W. (1993). DITYL: computerized key for species identification of *Ditylenchus* (Nematoda: Anguinidae). *Fundamental and Applied Nematology* 16, 389-392.

VISCARDI, T. & BRZESKI, M.W. (1995). [Computerized key for eelworms identification]. *Materialy Sesji Instytutu Ochrony Roslin* 35, 374-375.

YO-WEIMIN (1996). Applying Microsoft Works spreadsheet in statistics for morphometric data of nematode identification. *Afro-Asian Journal of Nematology* 6, 203-211.

ZULLINI, A. & MARGANELLI, G. (1989). A new computer program for nematode identification. *Computer Applications in the BioSciences* 5, 243-244.