



Short communication

NINJA: An automated calculation system for nematode-based biological monitoring



Bogdan Sieriebriennikov^{a,*}, Howard Ferris^b, Ron G.M. de Goede^c

^a Ghent University, Nematology Unit, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium

^b University of California, Davis, Department of Entomology and Nematology, One Shields Avenue, Davis, CA 95616, USA

^c Wageningen University, Department of Soil Quality, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:

Received 21 October 2013

Received in revised form

15 January 2014

Accepted 7 February 2014

Available online 25 February 2014

Handling editor: Christoph Tebbe

Keywords:

Nematodes

Biological monitoring

ABSTRACT

Monitoring of soil quality and health provides critical insights into the performance of ecosystems. Nematodes are useful indicators of soil condition because they are ubiquitous, represent different trophic levels of a soil food web and are convenient to work with. Several quantitative analyses of nematode assemblages have been developed and used in monitoring programs and by individual researchers. However, the calculations of the metrics involved are quite complicated. Since they are done manually using spreadsheet software, the calculations are time-consuming and error-prone and usually involve a significant learning curve for the user. We have developed an R code to perform these calculations. The code is compiled in html and deployed over the web. It is and will remain freely accessible and has a user-friendly interface. It requires only an input table with taxonomic inventory data and provides output within a few seconds.

© 2014 Elsevier Masson SAS. All rights reserved.

1. Introduction

Soil represents one of the most complex and diverse habitats in the biosphere: it is an important site for some essential ecosystem services, including nutrient cycling and sequestration [1]. Monitoring of soil quality and soil health provides critical insights into the condition and performance of the ecosystem. Because of the great variety of organisms inhabiting soil, and the diversity of methodologies necessary for assessing their abundance, it is impractical to conduct a complete inventory as a basis for biological monitoring [2]. Hence, it is most convenient to select one group that is sufficiently representative of the whole system. Ideally, the organisms of such a group should be ubiquitous, represent different trophic levels of a soil food web, be easy to extract from the soil and convenient to classify into taxonomic and functional groups; all those features are exhibited by nematodes [3].

Nematodes are small (typically less than 2.5 mm long) invertebrates that include free-living inhabitants of soil and sediment as well as parasites of plants and animals. They are found in

practically all types of terrestrial and aquatic habitats and respond in quantity and community structure to changes and perturbations of ecosystems [4,5]. Several approaches have been developed for the biological monitoring of soil quality based on the abundance and structure of nematode assemblages. Some of these methods have been used in monitoring programs on a national scale, including the NSMN program of The Netherlands [2] and the US-EMAP program of USA [6].

The calculations of the indices and other metrics (as discussed below) that are proposed and used in nematode faunal analysis are quite complicated. They are usually performed in Microsoft Excel or other spreadsheet software and may require manual adjustment of formulae and data format during the process. The approach is rather time consuming and error-prone, and may involve a significant learning curve for the user. A more advanced computational example is a SAS code developed by Neher & Campbell [7] which calculated four of the indices and also required much manual adjustment of the data.

We have developed an R code [8] to perform the faunal analysis calculations. The code is compiled in html and deployed over the web. It is and will remain freely accessible and has a user-friendly interface. The tool permits the flexibility to override the coded default feeding habits of individual taxa. It requires only an input table with taxonomic inventory data and provides output within a few seconds.

* Corresponding author.

E-mail addresses: sieriebriennikov@yandex.com, silvern@bigmir.net (B. Sieriebriennikov), hferris@ucdavis.edu (H. Ferris), ron.degoede@wur.nl (R.G. M. de Goede).

2. Theoretical background

Biomonitoring using nematode taxa is built upon a classification based on life-cycle characteristics. Specifically, the families, genera and species of free-living (i.e., all but herbivores and animal parasites) and herbivore nematodes are distributed among five classes in a so called coloniser–persister, or c–p, series. Nematodes in c–p-1 are regarded as “enrichment opportunists”, being able to complete their life cycle in some cases in as short as two days and their population sizes responding rapidly to pulses of nutrient input. As a result, c–p-1 nematodes are often dominant in disturbed environments. In contrast, c–p-5 nematodes may be considered “extreme persisters” that are generally intolerant of disturbance and inhabit stable, mature ecosystems; they have life cycles that may last several months. The weighted mean of the proportion of a free-living nematode assemblage in each c–p class is designated the Maturity Index (MI) and is considered representative of the degree of maturity of an ecosystem [4,9]. The same index calculated for herbivore nematodes is called the Plant Parasitic Index (PPI), although its interpretation may be somewhat ambiguous as it responds to enrichment and disturbance in a more complex manner than MI [10–13]. Nematodes in class c–p-2 (and not c–p-1) were shown to be the most resistant to pollution with heavy metals, so the weighted mean of the proportions of nematodes in all c–p classes except c–p-1 is referred to as Maturity Index 2–5 (MI_{2–5}). This measure is considered an indicator of heavy metal pollution, whereas MI tends to react primarily to nutrient enrichment, although to some extent both indices respond to various kinds of ecosystem disturbances [1,11]. A graphical illustration of the c–p structure of nematode assemblages is a c–p triangle, where the composition is depicted in triangular coordinates with the three corners representing absolute prevalence of c–p-1, c–p-2 and c–p-3–5 (combined number of nematodes in groups c–p-3, -4 and -5), respectively [14].

Food web diagnostics, which characterise the functional aspects of the nematode assemblage, consider the functional roles of different taxa in a food web based on their feeding habits. Thus, all free-living nematode taxa are allocated to three components: a) basal component comprising bacterivores of c–p-2 and fungivores of c–p-2; b) enrichment component including all c–p-1 nematodes and fungivores of c–p-2; c) structure component represented by all c–p-3–5 nematodes as well as by predators of c–p-2. Each combination of c–p class (life-history characteristic) and feeding habit is considered a functional guild. Accordingly, fungivores of c–p-2 constitute one functional guild and fungivores of c–p-3 another. A weight is assigned to each guild based on the hypothesis of constant connectance in community food webs. Weighted proportions of functional guilds are used to infer various attributes of the food web and give rise to an Enrichment Index (EI) and a Structure Index (SI). EI and SI were shown to correlate respectively with the intensity of nutrient enrichment and the degree of ecosystem maturity. The Channel Index (CI) is calculated from the weighted proportions of fungivores of c–p-2 and bacterivores of c–p-1 and is considered to indicate whether the “fast” bacterial channel or “slow” fungal channel of energy transformation prevails in an ecosystem [3,15].

A step further is the calculation of metabolic footprints as indices of Carbon utilisation by different nematode guilds and food web components. In effect, while the indices characterise ecosystem attributes based on proportion of nematodes in various categories, metabolic footprints attempt to quantify the magnitude of ecosystem services and functions. The generalised formula takes into account Carbon turnover due to both production and respiration and only requires the mass of an individual and the number of individuals for the calculation of the metabolic footprint of each taxon [16]. The mass of a nematode is based on morphometric

parameters using the formula of Andrassy [17]. Metabolic footprints are partitioned with regard to various ecosystem functions [16]. The composite footprint is a metabolic footprint of the whole nematode assemblage. Herbivore, bacterivore, fungivore, predator and omnivore footprints are the footprints calculated for corresponding trophic groups. Enrichment and structure footprints are the footprints of nematodes belonging to enrichment or structure components respectively.

For more details concerning the calculation of the indices and the rationale behind them the reader is referred to the original publications.

3. Programming implementation

The web interface is accessible via the next link: <http://spark.rstudio.com/bsierieb/ninja/>.

The R code is available upon request.

3.1. Data input

The input is a table filled with the numbers of individuals of each taxon in a sample, where columns correspond to different taxa and rows represent treatments or sampling locations. Consequently, the first row contains the taxa names. The first column contains individual names of samples which are utilised by a user only to correctly arrange their data and are not taken into analysis. The second column holds the names of the treatments or sampling locations which must be identical for different samples from the same treatment or location. Taxa are nematode families, genera or species, and different taxonomic levels (e.g., families and genera) may be combined in a single input table. There is an option to include dauerlarvae, they should be named “dauer” (without quotation marks). The quantities should be expressed as number of individuals per unit of area, soil volume or mass. If the quantities are expressed as proportions of taxa, indices will be calculated correctly, but the calculation of metabolic footprints requires the actual abundance of each taxon. The example of an input is provided in Table 1.

The assembled data should be saved in the Microsoft Excel format, either .xls or .xlsx, with points as decimal separators, and then uploaded by means of the web interface.

In the “Start” tab (opened by default) the results of a procedure are shown, which checks whether taxa names from the input file

Table 1

An example of the input table. The first column is filled with individual sample names that are intended for user's convenience only and are not used in the analysis. The second column contains names of treatments or sampling sites which must be ordered alphabetically. If indices should be calculated for each separate sample, then all names in the second column should be unique. If averages per treatment or sampling site are required, then the same treatments or sampling sites must bear identical names. First line contains taxa names that can be the names of families, genera or species of nematodes. Dauerlarvae can also be included and they should be encoded as “dauer” (without quotation marks). Numbers in the table represent counts of individuals normalised per unit of area, soil volume or mass. Decimal separators are points.

		Plectidae	Plectus	Plectus parvus	dauer
<Your name>	Control	5.5	25	0	5
<Your name>	Control	6	40	0	10
<Your name>	Control	4	10	1	0
<Your name>	Site A	15	50	5	2
<Your name>	Site A	18	60	9	3
<Your name>	Site A	12	40	1	1
<Your name>	Site B	60	200	20	8
<Your name>	Site B	72	240	36	12
<Your name>	Site B	48	160	4	4

correspond to those retained in the database and suggests corrections when possible. Additionally, in the case that some columns have identical taxa names, such columns are summed and a warning is provided for the user. After necessary corrections are performed, the input must be uploaded again. When the data check has not identified any mismatches, the user may proceed to the next tab.

3.2. Database

A database contains the vast majority of nematode families, genera and species which can be found in soil (excluding most of animal parasites) with the corresponding values of *c–p* following Bongers [4] and subsequent publications [18,19]. The *c–p* values are provided for both free-living forms and herbivores, in the latter case being called *p–p* (plant parasites). Other parameters in the database are feeding type according to Yeates et al. [20] and an estimate of individual body mass taken from the Nemaplex website (http://plpnemweb.ucdavis.edu/nemaplex/Ecology/nematode_weights.htm). Information for different taxonomic levels is concatenated in a single table which allows using taxa of any level as an input. The database table is stored on the server and is loaded automatically. We plan to maintain and regularly update the database and any suggestions (e.g., data for taxa not yet included) are very welcome.

In actual fact, feeding types designation may be ambiguous. This is the reason why we added an option to adjust feeding types assigned to taxa by default (i.e. based on the classification of Yeates et al. [20] and subsequent publications; for a complete list of these publications follow <http://spark.rstudio.com/bsierieb/ninja/mutations.html>). As soon as a user's input file is uploaded, a list of taxa from the uploaded file appears in the sidebar panel together with default feeding types assigned to them. At this stage, the user may adjust the feeding types according to their own knowledge or leave the values as set by default.

3.3. Output

3.3.1. Summary

The table contains calculated values of estimates of the Maturity Index, Maturity Index 2–5, Plant Parasitic Index, Channel Index, Enrichment Index, Structure Index, total biomass, composite footprint, enrichment footprint, structure footprint, herbivore footprint, fungivore footprint, bacterivore footprint, predator footprint, omnivore footprint as well as fractions of different feeding types (both percentages from the total numbers and from the numbers of free-living nematodes) and fractions of different *c–p* classes amongst herbivores and free-living forms. For all measures but the percentages the means and standard deviations are provided, calculated per treatment or sampling site and accompanied by corresponding *p*-values of ANOVA.

3.3.2. Feeding types and *c–p/p–p*

At the top of the page there is a table with the values extracted from the database, provided for a user to know which information has been used in calculations. The table is followed by five stacked column charts that represent the feeding type composition of the whole nematode assemblage, the free-living and the herbivore nematode assemblage as well as the *c–p* structure of the free-living nematode assemblage and of the herbivore assemblage.

3.3.3. MI family indices

This tab contains a set of box & whisker plots representing Maturity Index, Maturity Index 2–5 and Plant Parasitic Index. The last figure is a *c–p* triangle.

3.3.4. Food web diagnostics

There are two charts on the tab – a box & whisker plot for the values of Channel Index and a plot with the values of Enrichment and Structure index scattered along the two corresponding axes.

3.3.5. Metabolic footprints

This tab contains a plot of the type described by Ferris [16], with an additional feature of illustrating standard deviation as a dotted line on both sides of the full line that represents mean values for a treatment. Briefly, the point in the middle of a depicted rhombus represents the intersection of Enrichment Index and Structure Index and length of vertical and horizontal axes of the rhombus corresponds to enrichment and structure footprints respectively. It is important to mention that the graphic representation of the footprints shows relative proportions of components inside a single graph and cannot be used to compare those across different graphs, as a different scaling coefficient is used each time. Thus, for comparison of footprints from different graphs absolute values from the summary table should be used.

3.4. References to other R packages

The R code [8] is compiled in html using the shiny package [21] and deployed via the web using an RStudio-hosted shiny server. To read in tables in the Excel format, the package *gdata* [22] is used. Other functions used include those from the shinyIncubator [23] and base [8] packages.

3.5. System requirements

NINJA requires a web browser supporting the WebSocket protocol. The following web browsers provide this feature: Internet Explorer version 10 and higher, Firefox version 6 and higher, Safari version 6 and higher, Google Chrome version 14 and higher, Opera version 12.10 and higher.

4. Summary and future perspectives

The represented system provides a free, flexible and easily accessible tool for quick calculation of all the principal indices used in nematode-based biomonitoring. Limiting a user's role to only creating an input file greatly simplifies the procedure so the tool does not require a high level of computational expertise. Moreover, this approach protects against calculation errors. Primary statistical treatment and automatic creation of graphs facilitate the processing of the results obtained.

The rapidly evolving development in nematode-based biomonitoring is the introduction of molecular techniques for species and functional identification, which would improve economic efficiency and resolution of the analyses. The catalogue of diagnostic sequences for nematodes is already being developed [5,24]. Simple expansion of the database described in this article by including sequence accession numbers will also allow processing of the results obtained via molecular identification.

Acknowledgements

We would like to acknowledge classes and lectures in the EUMAINE (Erasmus Mundus European Master of Science in Nematology) program for inspiring the creation of the tool. We thank three anonymous reviewers for providing valuable comments. We are also grateful to Amra Kujundžić, Israel Onoja Ikoyi and Tatsiana Lipinskaya for the initial testing of the web-interface and providing their feedback.

References

- [1] H. Ferris, T. Bongers, Indices for analysis of nematode assemblages, in: M. Wilson, T. Kakouli-Duarte (Eds.), *Nematodes as Environmental Bio-indicators*, CABI, Wallingford, 2009, pp. 124–145.
- [2] M. Rutgers, A.J. Schouten, J. Bloem, N. van Eekeren, R.G.M. de Goede, G.A.J.M. Jagers op Akkerhuis, et al., Biological measurements in a nationwide soil monitoring network, *Eur. J. Soil Sci.* 60 (2009) 820–832.
- [3] H. Ferris, T. Bongers, R.G.M. de Goede, A framework for soil food web diagnostics: extension of the nematode faunal analysis concept, *Appl. Soil. Ecol.* 18 (2001) 13–29.
- [4] T. Bongers, The maturity index: an ecological measure of environmental disturbance based on nematode species composition, *Oecologia* 83 (1990) 14–19.
- [5] M. Vervoort, J. Arie Vonk, P. Mooijman, S. van den Elsen, H. van Megen, P. Veenhuizen, et al., SSU ribosomal DNA-based monitoring of nematode assemblages reveals distinct seasonal fluctuations within evolutionary heterogeneous feeding guilds, *PLoS One* 7 (2012) 1–13.
- [6] C. Campbell, J. Bay, A. Hellkamp, G. Hess, M. Munster, K. Nauman, et al., *Environmental Monitoring and Assessment Program: Agroecosystem Pilot Field Program Report – 1992, 1994.*
- [7] D. Neher, C. Campbell, Sampling for regional monitoring of nematode communities in agricultural soils, *J. Nematol.* 28 (1996) 196–208.
- [8] R Core Team, *R: A Language and Environment for Statistical Computing*, 2013.
- [9] T. Bongers, H. Ferris, Nematode community structure as a bioindicator in environmental monitoring, *Trends Ecol. Evol.* 14 (1999) 224–228.
- [10] T. Bongers, H. van der Meulen, G. Korthals, Inverse relationship between the nematode maturity index and plant parasite index under enriched nutrient conditions, *Appl. Soil. Ecol.* 6 (1997) 195–199.
- [11] T. Bongers, G. Korthals, The behaviour of maturity index and plant parasite index under enriched conditions, in: *Proc. 22nd Int. Symp. Eur. Soc. Nematol.*, Ghent, Belgium, 1994, p. 39.
- [12] D.W. Freckman, C.H. Ettema, Assessing nematode communities in agroecosystems of varying human intervention, *Agric. Ecosyst. Environ.* 45 (1993) 239–261.
- [13] G.W. Yeates, Modification and qualification of the nematode maturity index, *Pedobiologia (Jena)* 38 (1994) 97–101.
- [14] R.G.M. de Goede, T. Bongers, C.H. Ettema, Graphical presentation and interpretation of nematode community structure: c–p triangles, *Med. Fac. Landbouww. Univ. Gent* 58 (1993) 743–750.
- [15] H. Ferris, M.M. Matute, Structural and functional succession in the nematode fauna of a soil food web, *Appl. Soil. Ecol.* 23 (2003) 93–110.
- [16] H. Ferris, Form and function: metabolic footprints of nematodes in the soil food web, *Eur. J. Soil. Biol.* 46 (2010) 97–104.
- [17] J. Andrassy, Die Rauminhalt und Gewichtsbestimmung der Fadenwürmer (Nematoden), *Acta Zool. Acad. Sci. Hung.* 2 (1956) 1–15.
- [18] T. Bongers, R. Alkemade, G.W. Yeates, Interpretation of disturbance-induced maturity decrease in marine nematode assemblages by means of the Maturity Index, *Mar. Ecol. Prog. Ser.* 76 (1991) 135–142.
- [19] T. Bongers, R.G.M. de Goede, G. Korthals, G.W. Yeates, Proposed changes of c–p classification for nematodes, *Russ. J. Nematol.* 3 (1995) 61–62.
- [20] G.W. Yeates, T. Bongers, R.G.M. de Goede, D. Freckman, S. Georgieva, Feeding habits in soil nematode families and genera – an outline for soil ecologists, *J. Nematol.* 25 (1993) 315–331.
- [21] RStudio, shiny: Web Application Framework for R, 2013.
- [22] G. Warnes, B. Bolker, G. Gorjanc, G. Grothendieck, A. Korosec, T. Lumley, et al., *gdata: Various R programming tools for data manipulation*, 2013.
- [23] RStudio, shinyIncubator: Unofficial add-ons for Shiny, 2012.
- [24] M. Vervoort, J. Arie Vonk, K. Brolsma, W. Schutze, C. Quist, R.G.M. de Goede, et al., Release of isothiocyanates does not explain the effects of biofumigation with Indian mustard cultivars on nematode assemblages, *Soil. Biol. Biochem.* 68 (2014) 200–207.