# *PPNID*: a manually-curated reference database and molecular identification pipeline for plant-parasitic nematodes

## User Manual v1.0

Xue Qing[1], Meng Wang[1], Patricia Bucki[1], Wim Bert[2] & Sigal Braun Miyara[1]

[1]Department of Entomology, Nematology and Chemistry units; Agricultural Research Organization (ARO), the Volcani Center, Bet Dagan 50250, Israel

[2]Nematology Research Unit, Department of Biology, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium


Please contact Dr. Qing Xue for more details: xueqing4083@gmail.com

The program is available at https://github.com/xueqing4083/PPNID

# Contents

# 1. User guideline

## 1.1 Get required sequences in this program

1.1.1 *rRNA* sequence

18S, 28S and ITS rRNA are the most widely used barcoding genes for nematodes. The list of commonly used rRNA primers:

*-18S*

Primers pair for nearly full length

G18S4 (5′-GCT TGT CTC AAA GAT TAA GCC-3′) Forward
18P (5′-TGA TCC WMC RGC AGG TTC AC-3′) Reverse

Or some of shorter primers:

2FX (5′-GGA AGG GCA CCA CCA GGA GTG G-3′) Forward
23R (5′-TCT CGC TCG TTA TCG GAA T-3′) Reverse
13R (5′-GGG CAT CAC AGA CCT GTT A-3′) Reverse
23F (5′-ATTCCG ATA ACG AGC GAG A-3′) Forward
9FX (5′-AAG TCT GGT GCC AGC AGC CGC-3′) Forward
9R (5′-AGC TGG AAT TAC CGC GGC TG-3′) Reverse
26R (5′-CAT TCT TGG CAA ATG CTT TCG-3′) Reverse

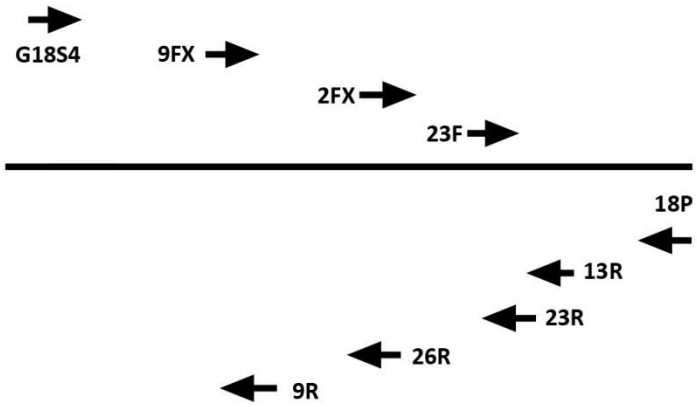



Fig. 1 The scheme for primer binding regions in 18S rRNA gene

*-D2D3 domain of 28S:*

D2A 5′- ACAAGTACCGTGAGGGAAAGTTG -3′ Forward
D3B 5′- TCGGAAGGAACCAGCTACTA -3′ Reverse

-*ITS:*

VRAIN2F 5′-CTTTGTACACACCGC CCGTCGCT-3′ Forward
VRAIN2R5′-TTTCACTCGCCGTTACTAAGGGAA TC-3′ Reverse

or

5367 5′- TTGATTACGTCCCTGCCCTTT -3′ Forward
F195 5′- TCCTCCGCTAAATGATATG -3′ Reverse

1.1.2 Mitochondrial genes

Mitochondrial haplotype from Nad 5 and Cytb genes are used for tropical root-knot nematode identification in this program. COI gene is widely used for barcoding in different taxa.

- *NADH dehydrogenase subunit 5 (Nad 5)*

Nad 5 primers:

NAD5F2 5′- TATTTTTTGTTTGAGATATATTAG -3′ Forward
NAD5R1 5′- CGTGAATCTTGATTTTCCATTTTT -3′ Reverse

(Recommend PCR amplification conditions: initial desaturation 2 min at 94 ℃, 40 cycles of 60s at 94 ℃, 60s at 45 ℃, 90s at 72 ℃, and finally an extension for 10 min at 72 ℃. )

- *Cytochrome b (Cytb)*

Cytb primers:

CYTBF 5′- TGAGGTTAATAATGGTTGGTTAATTCG -3′ Forward
CYTBR 5′- GGGAGCCAAGAACCAGTTTT -3′ Reverse

Recommend PCR amplification conditions: initial desaturation 2 min at 94 ℃, 40 cycles of 60s at 94 ℃, 60s at 55 ℃, 90s at 72 ℃, and finally an extension for 10 min at 72 ℃.

- *Cytochrome c oxidase subunit 1 (COI)*

Universal primers:

JB3 5’- TTTTTTGGGCATCCTGAGGTTTAT -3’ Forward
JB4.5 5’- TAAAGAAAGAACATAATGAAAATG -3’ Reverse

Primers for RKN

COX1F 5′- ATCCTCCTTTGATGATTGATGG -3′ Forward

COX1R 5′- AACTCAATAAAGAACCAATAGAAG -3′ Reverse

### 1.2 Program installation

Download PPNID.zip to any temporary directory. Extract all files contained in PPNID.zip in the directory of your choice. Start program by double-clicking on the file PPNID.exe, which is the main executable file.

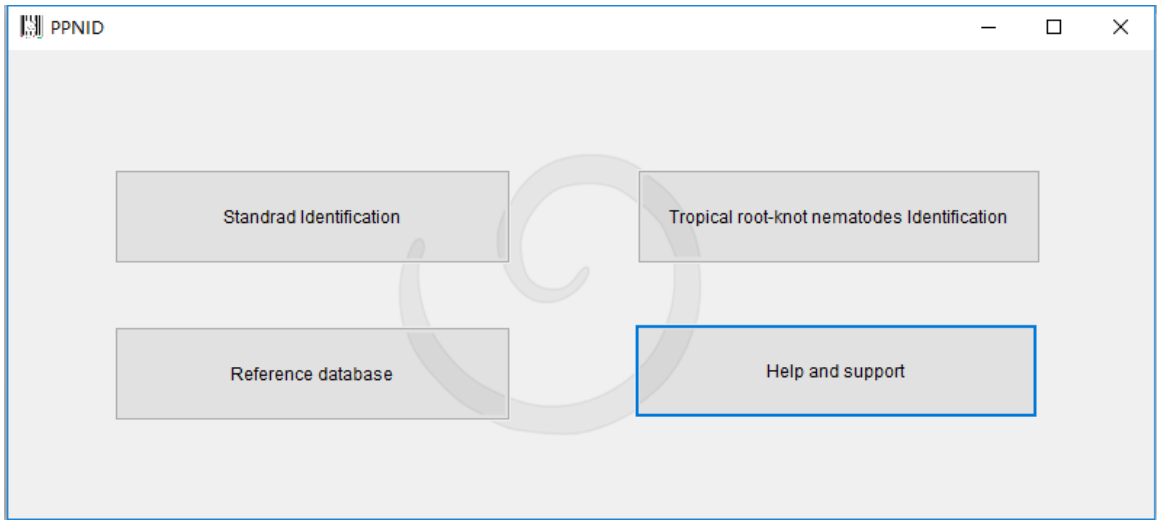



Fig.2 The main interface of PPNID

### 1.3 Which pipeline to choose?

-For most cases, please start with “Standard Identification”. This includes different plant-parasitic nematodes (PPN) taxa and four barcoding genes.

-If you want to identify tropical RKN (*M. incognita*, *M. javanica*, *M. arenaria* and other similar species in the *M. incognita* clade *e.g. M. inornata*) based on the haplotype method (Janssen et al., 2016), please use “Tropical root-knot nematodes Identification”.

### 1.4 Standard identification

#### 1.4.1 Input sequence and multiple sequence alignment

The user can start identification by choosing a taxonomic group (at family or superfamily level) and a barcoding gene (18S, 28S, ITS and COI), then paste the sequence into textbox (Fig. 3). Only one sequence is allowed each time. To have more powerful identification result,

PPNID only accept sequences without ambiguous sites. Help page provides some of the commonly used barcoding primers and corresponding PCR conditions. When your sequence is ready, select the gene database you want to search and click “Submit”

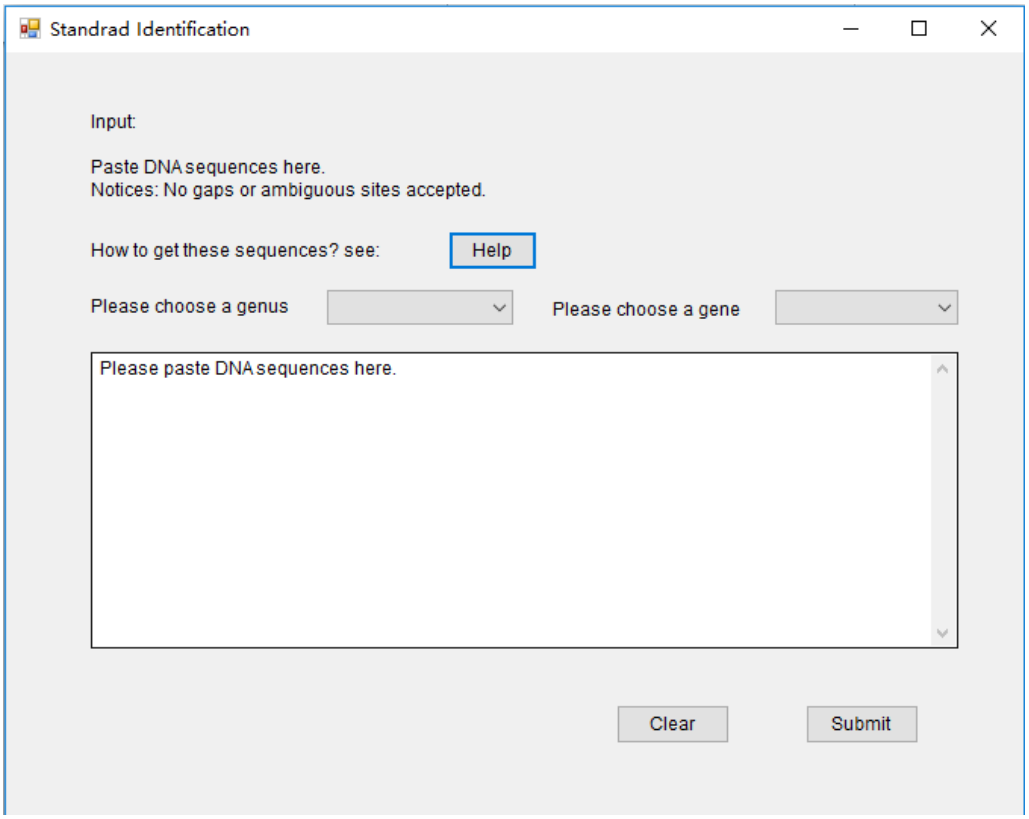



Fig.3 The Standard identification page.


Fig. 4 Multiple sequences alignment page (user input sequence displayed in first line)

If your sequence is correct, PPNID automatically aligns the input sequence with the reference database. In a new window, the multiple sequences alignment is visualized with each nucleotide marked in the same color (Fig.4). By dragging the scroll bar, the user can check the mutations in the sequences.

1.4.2 Similarity comparison

Similarity comparison of user inputted sequence to reference database can be achieved by clicking “Click here to see identification result” (Fig. 4). PPNID use p-distance to search the most similar species. The results are expressed in nucleotides differences and difference rate, and three most similar species are showed in a pop-out window (Fig. 5).

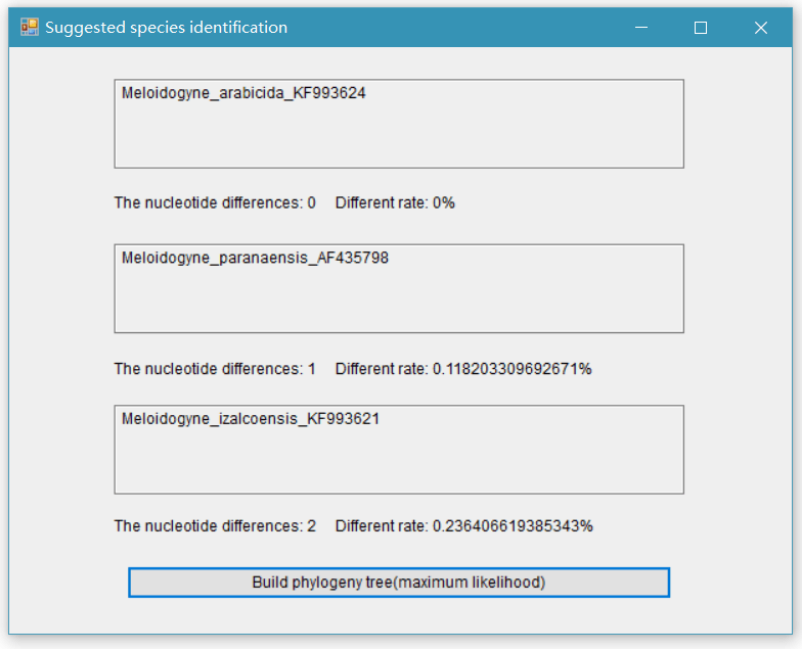



Fig. 5 The identification result showing three most similar species

1.4.3 Maximum likelihood tree reconstruction

To further evaluate the phylogenetic placement of the input sequence, PPNID build a maximum likelihood tree with other known species of a given genus. Bootstrap values are given at branch nodes and, in general, a value greater than 0.7 (or written as 70 sometimes) can be considered as a reliable clade (Fig. 6). The program automatically generate *.bmp* format image for the tree and it is saved in “result” file of PPNID directory. By clicking “Draw” button, the user can refresh the tree and overwrite the image file.

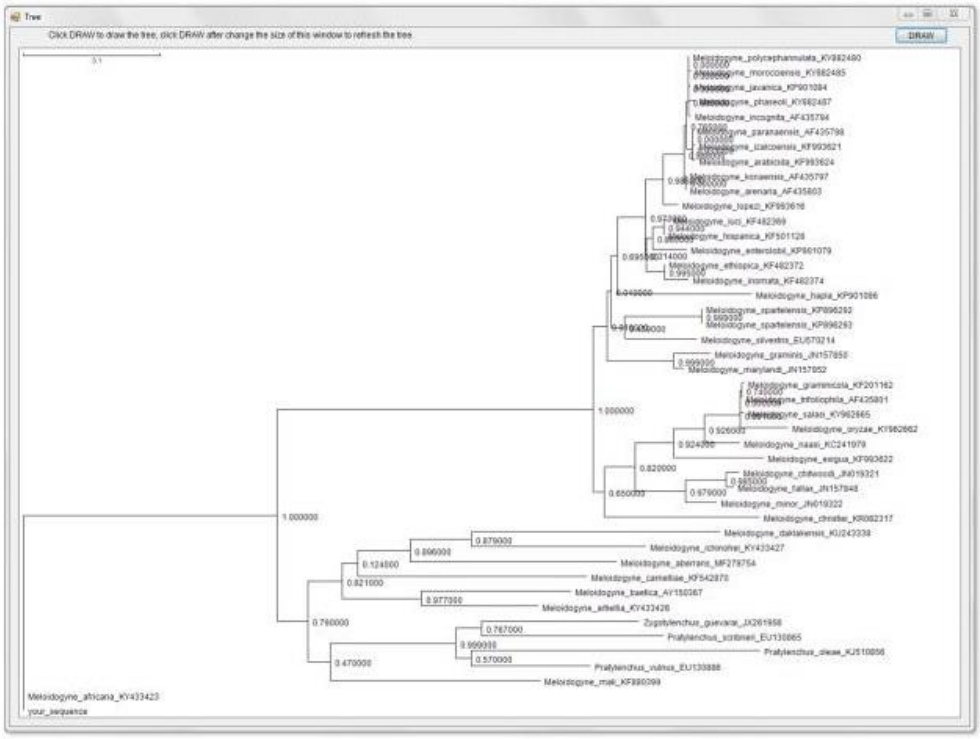



Fig.6 The maximum likelihood phylogeny tree of user input species and other reference species in genus *Meloidogyne*.

### 1.5 How to make a conclusion for identification?

Notice that the identification of PPN is difficult, based on expertise and a consensus among taxonomists is often lacking, therefore a standardized identification pipeline arriving to a definitive conclusion is inherently not possible. Thus, the goal of this program is an improved interpretation of molecular data and replace the widely used BLAST search among non-specialist, but it can not be replace conclusions based on a comprehensive taxonomical investigation. Here we provide some suggestions how to interpret the PPNID program.

*-How to interpret similarity comparison?*

If the highest match is reached in a similarity comparison (chapter 1.4.2) one of the sequences is usually the same as the input sequence, however, it may **still be a different species! i**f your sequence is a 100% match (e.g. Fig. 5) with one of the sequences in the

reference database, your sequence likely belongs to be the same species, however, this is not the case in tropical RKN. If sequences are not 100% identical, further maximum likelihood analyses is needed to allow interpretation in a phylogeny topology with support value.

-*Combine similarity search and phylogeny analysis*

If your sequence is 95%-100% similar to one of sequence in reference database, your input sequence has high probability to be same with the highest matched one, but maximum likelihood analysis is needed. Below are different scenarios users may encounter:

(1) When your sequence is in a well-supported monophyletic clade together with other unresolved species, and all species have short branches (Fig. 7, I), your input sequence can be different to highest match species (input sequence can be same species to any one of A, B, C in Fig. 7). In this case suggests currently applied barcoding gene is insufficient to different this taxonomic group. You may need a less conserved gene, e.g. COI, ITS.

(2) When your sequence has quite long branch to other clades, the inputted sequence can be different to highest match species (Fig. 7, II). This is either because your inputted species do not have molecular representative in database (more likely) or your species is new to science (less likely).

(3) When your sequence and one sister species is nested in a supported clade, both branch lengths are very short (Fig. 7, III, VI), and sister species has highest similarity match, your inputted species is likely same to sister species (input sequence can be identified as A in Fig, 7).

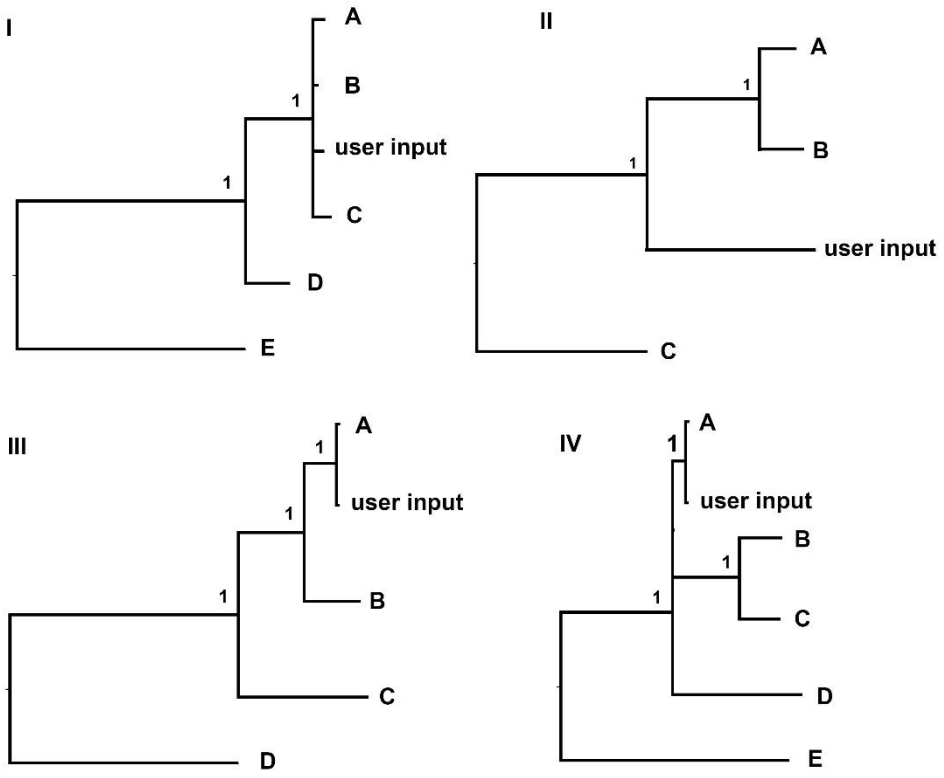



Fig. 7. The examples of species identification in given different tree topologies. I: barcoding gene is less informative and user input sequence can be same to A, B and C; II: the molecular data is missing for user input species or it is new species; III and IV: user input species can be identified as species A.

### 1.6 Tropical root-knot nematodes

When your sequence belongs to tropical RKN (here defined as *M. incognita*, *M. javanica*, *M. arenaria* and other similar species in the *M. incognita* clade *e.g. M. ethiopica*, *M. inornata* and *M. luci*), rRNA cannot provide sufficient resolution for appropriate identification result. In this case user needs to provide mitochondrial gene Nad5. You can start this process by shortcut clicking “Quick start: tropical root-knot nematodes” in main menu page, or follow the instruction in “Standard identification” when result fall in these similar species. Similar with standard identification, user need to paste Nad5 sequence in textbox and click “submit” (Fig.8).

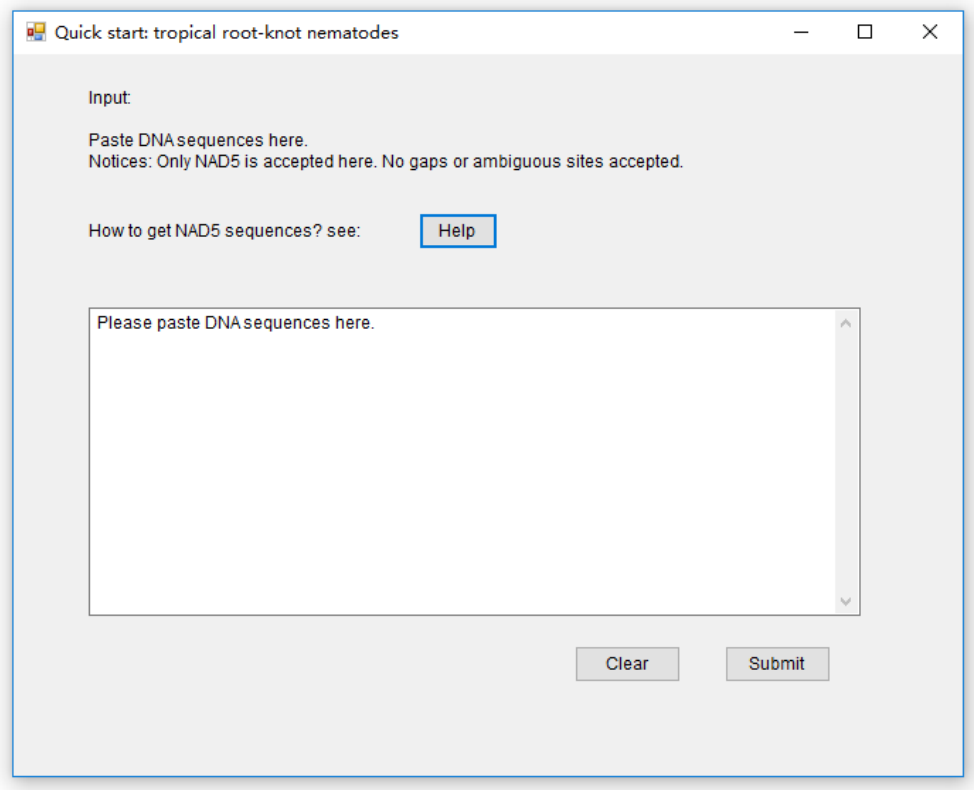



Fig.8 The tropical RKN identification page using Nad5 gene

PPNID extracts the informative polymorphic positions of input sequence and compare with database. The results will show in a new window with each nucleotide marked in different colors (Fig. 7). You can manually compare the results by checking the polymorphic sites or click button "Click here to see identification result" (Figs.9, 10).

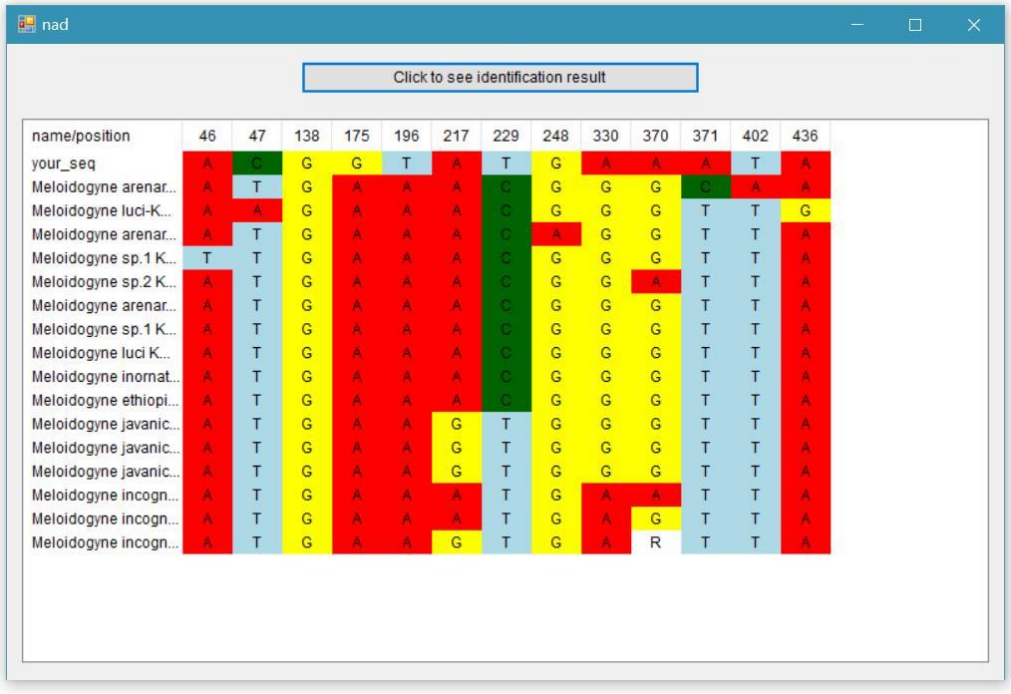


Fig. 9 The multiple sequences alignment of Nad5 gene shows mutations in the barcoding polymorphic sites.

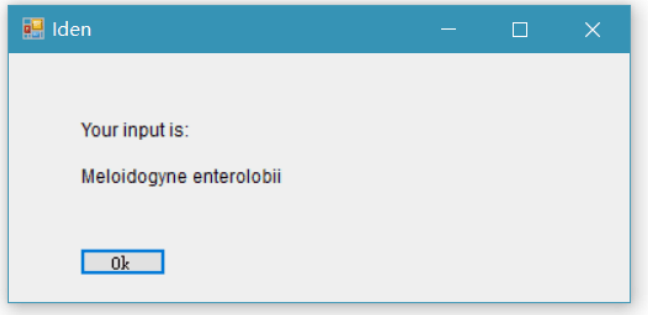



Fig. 10 The result of tropical RKN identification

In few cases, you may get the warning for gaps and unexpected sites (Fig. 11). This is because your input sequence differs from any of known haplotypes in database, while the barcoding polymorphic sites still fit some species in database. More details see method and algorithm part. When you click “Yes” to continue, PPNID will ignore those unexpected sites and only take barcoding polymorphic sites into consideration. In this scenario, you may get a correct identification by, but the reliability is unverified.

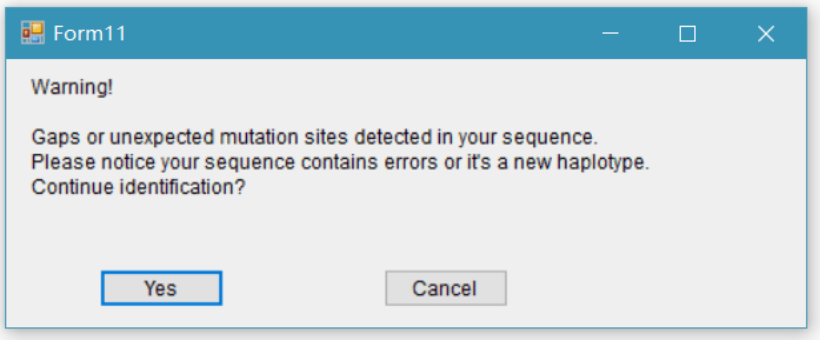



Fig. 11 The error page when user input sequence fails to match database

More rarely, you may need extra mitochondrial gene Cytb if your sequence mataches Nad5 haplotype share by *M. arenaria* H3, *M. ethiopica*, *M. inornata* and *M. luci* H1 (Fig. 12). Similar with Nad5 gene identification window, just simply paste Cytb sequence and Click “Submit” (Fig. 12).

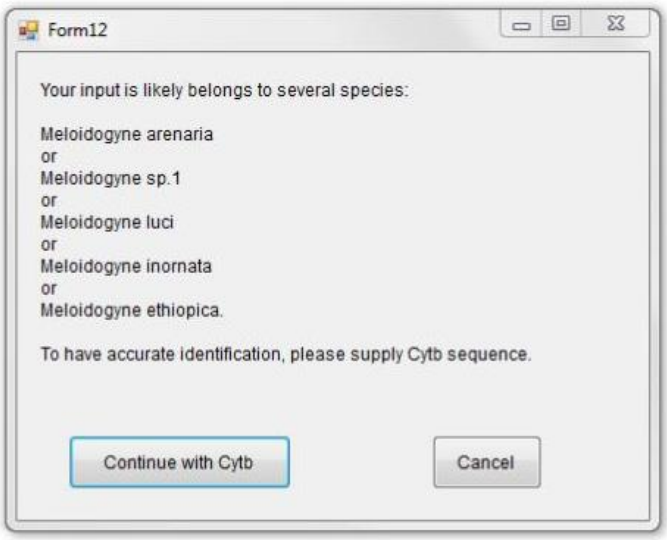



Fig. 12 The identification page when user input Nad5 gene match more than one species

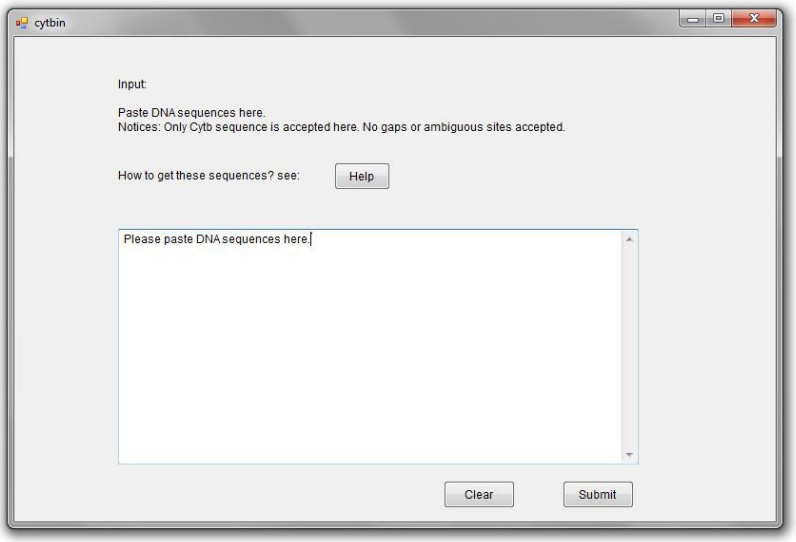



Fig. 13 The tropical RKN identification page using Cytb gene.

### 1.7 Plant-parasitic nematodes database

The PPNID database includes PPN barcoding sequences that were manually curated and updated. By select Taxa-gene database name in left drop-down menu and species name in right drop-down menus, user can search for the single reference sequence. In the case only database is selected but not for species name, all sequences from that selected database will be listed. The sequences are presented in FASTA format with GenBank accession number and other original information in first row (Fig. 14).

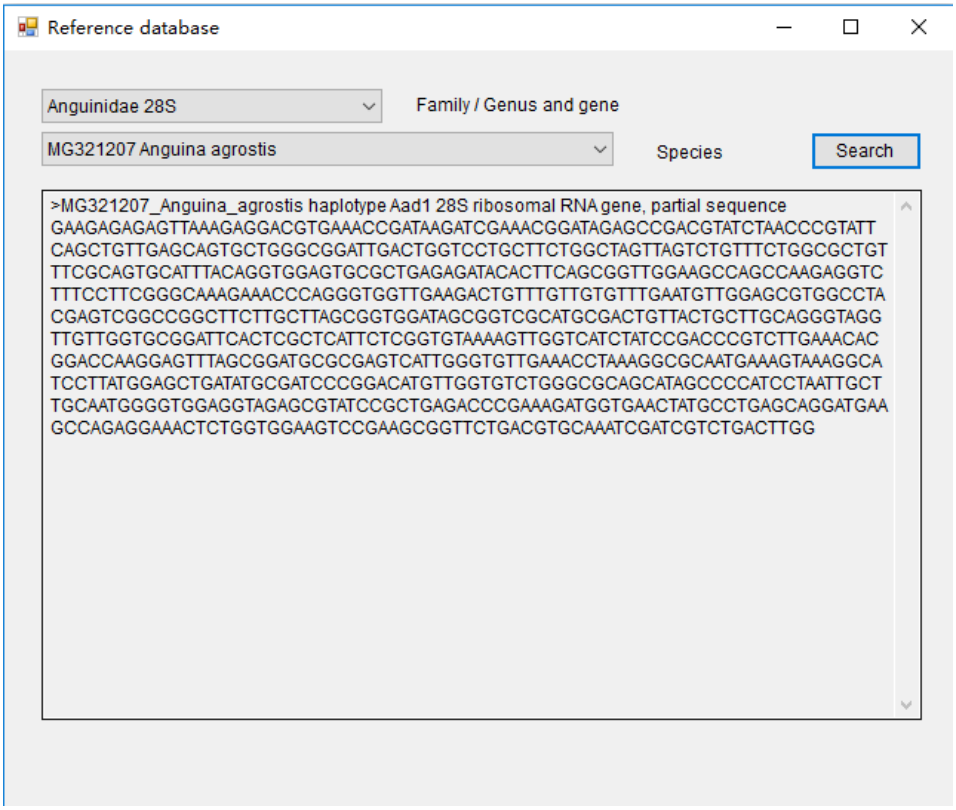



Fig. 14 The page of reference database

### 1.8 Rewrite tool

PPNID also provides tool for sequence reformatting (Fig. 15) and duplicates searching (Fig. 16). This tool is not necessarily needed for identification but may be useful for sequences preparation.

Basically, re-format tool can format any FAST file into FAST file without gaps, line breaks and all nucleotides will appears in single line. The newly formatted file will be saved in same directory of PPNID with name “form.txt”. Repeats searching tool will detect FAST file with duplicate sequences, but not change original file. The same sequence will be indicated as “>species1 == >species2”, while “species1” and “species2” are names of two sequences.

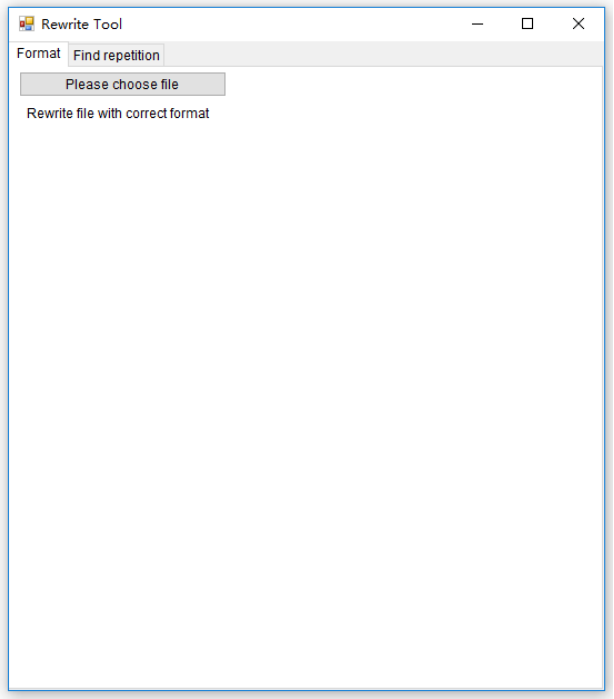



Fig. 15 The tool to remove gaps, and reformat line breaks in single line in FASTA format sequences file

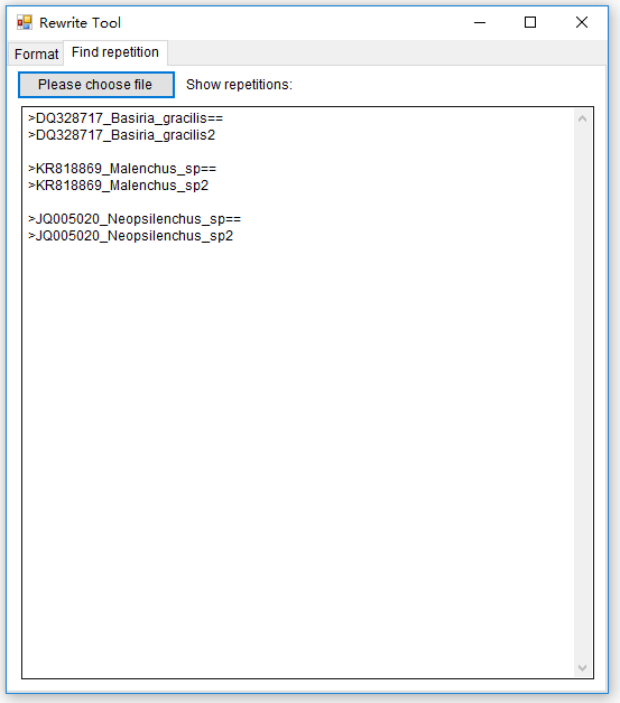



Fig. 16 The tool to find duplicate sequences in FASTA format sequences file

### 1.9 Help and support

In help and support page user will get information on the version of the program and citation.

Help and support page is also linked with other helps button in each of identification page. There we provide recommendations how to get sequences and how to solve the possible errors.

A proper PPN identification is difficult and DNA based identification remain problematic. Therefore the identification result should be carefully interpreted. If any bug is found in the PPNID, or any problem on identification, please contact Dr. Qing Xue for consultation.

## 2. Outline and background

### 2.1 Why PPNID?

In short: (1) widely used PPN sequence repository (INSD: International Nucleotide Sequence Databases, e.g. GenBank, EMBL, DDBJ) is problematic as it contains numerous errors, duplications, synonyms and unlabeled sequences. Comparisons between phylogenetic studies may be tricky as different authors include different reference species or sequences sources; (2) to compile and interpret molecular data involves a significant amount of manual intervention is practically difficult for non-expert, and solely using DNA-similarity searches like BLAST can be misleading; (3) the demand of RKN identification is high. However, a proper RKN identification is complicated, need background knowledge on nematology and phylogeny, laborious in data process and may introduce error. A straightforward tool that incorporates all steps in a single run is essential especially for non-specialists.

### 2.2 Methodological outlines

#### 2.2.1 Programming language

The PPNID is written in vb.net winform, developed on Microsoft visual studio 2017 and compiled versions are available for 64-bit and 32-bit of Windows.

#### 2.2.2 Input sequence

PPNID only accept assembled DNA sequences. The quality should be carefully checked from

chromatography file (.ab1). Ambiguous nucleotide sites were automatically converted as unknown (N). The input only allows one sequence at each time.

### 2.2.3 Multiple sequences alignment

PPNID use MUSCLE method proposed by Edgar (2004) for multiple sequences alignment. In short, this method use k-mer counting for fast distance estimation, log-expectation score for progressive alignment and tree-dependent restricted partitioning for refinement, and consider as one of best alignment algorithm combining speed and accuracy.

### 2.2.4 Standard identification

The genetic distances together with phylogenetic tree are used for species identification. This p-distance is the proportion of nucleotide sites at which two sequences being compared are different. It is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared. It does not make any correction for multiple substitutions at the same site, substitution rate biases (for example, differences in the transitional and transversional rates), or differences in evolutionary rates among sites (Fig. 17). The program create the p-distances based on user input sequence and database sequences. The sequence with lowest p-distance is selected and assigned as the most similar species to user input.

$$p\text{-distance} = \frac{no.\ of\ nucleotide\ differences}{Total\ no.\ nucleotide\ compared}$$

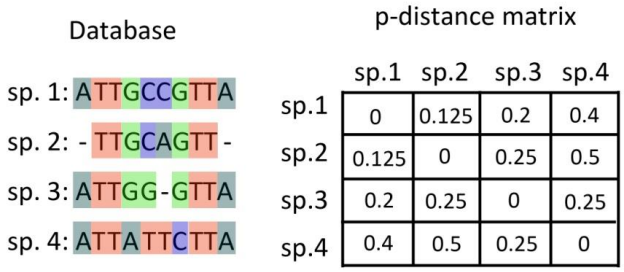



Fig. 17 The scheme of p-distance calculation method used in PPNID

To compensate the shortage of p-distance (*e.g.* without model on among-site rates and substitution rate) and indicate the phylogenetic placement of input sequence, PPNID introduces phylogenetic tree as identification reference. Here we use maximum likelihood

method employed in the PhyML package (Guindon and Gascuel, 2003) with all default settings except for a GTR model.

### 2.2.5 Identification of tropical RKN

PPNID use the method proposed by Janssen et al. (2016) for tropical RKN identification. From user input Nad5 sequence a total of 13 polymorphic sites are extracted and compared to the database. The polymorphic sites used in PPNID listed in Fig. 18.

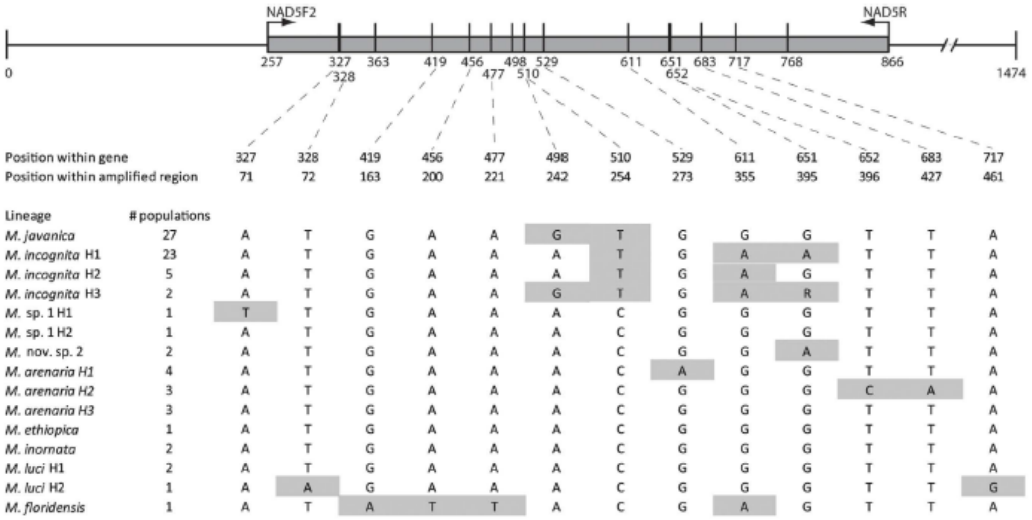
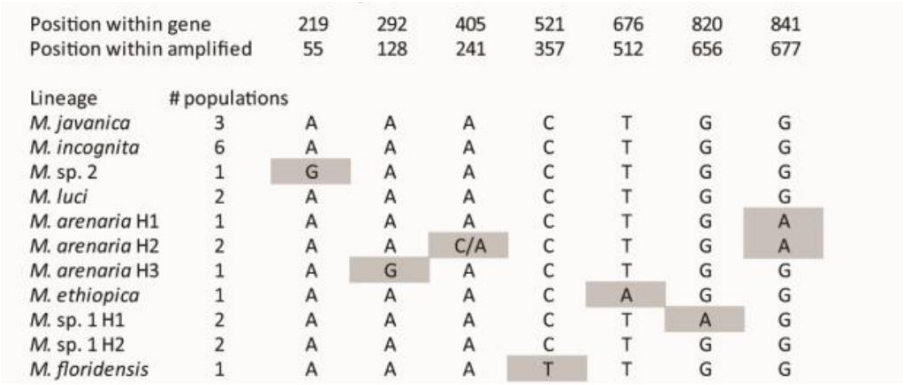



Fig. 18 The polymorphic barcoding sites used for species identification in Nad5 gene. Figure original describe by Janssen et al. (2016).

Although Nad5 gene can delimitate most common tropical RKN species, *M. arenaria* H3, *M. ethiopica*, *M. inornata* and *M. luci* H1 share same Nad5 haplotype. When user input sequence fits any of these four species, Cytb gene is needed for further identification. A total of seven polymorphic sites are used and listed in Fig. 16.



| Position within gene | | 219 | 292 | 405 | 521 | 676 | 820 | 841 |
|---|---|---|---|---|---|---|---|---|
| Position within amplified | | 55 | 128 | 241 | 357 | 512 | 656 | 677 |
| Lineage | # populations | | | | | | | |
| *M. javanica* | 3 | A | A | A | C | T | G | G |
| *M. incognita* | 6 | A | A | A | C | T | G | G |
| *M.* sp. 2 | 1 | G | A | A | C | T | G | G |
| *M. luci* | 2 | A | A | A | C | T | G | G |
| *M. arenaria* H1 | 1 | A | A | A | C | T | G | A |
| *M. arenaria* H2 | 2 | A | A | C/A | C | T | G | A |
| *M. arenaria* H3 | 1 | A | G | A | C | T | G | G |
| *M. ethiopica* | 1 | A | A | A | C | A | G | G |
| *M.* sp. 1 H1 | 2 | A | A | A | C | T | A | G |
| *M.* sp. 1 H2 | 2 | A | A | A | C | T | G | G |
| *M. floridensis* | 1 | A | A | A | T | T | G | G |

Fig. 19 The polymorphic barcoding sites used for species identification in Cytb gene. Figure original describe by Janssen et al. (2016).

2.2.6 Examples for tropical RKN identification

If we define the sequences of three species as:

Species 1：TTAAC

Species 2：TTTAC

Species 3：TTCAC

The barcoding position is the third nucleotide.

Here are algorithms used in PPNID:

**Case 1**. When user input is exactly same with species in database, program will return corresponding species as identification result, *e.g.*:

User input: AAATTAACGGG

Return identification：Species 1 (non-overlapping parts in input sequence are ignored, here AAA, GGG)

User input: TAAC (non-overlapping parts in database are ignored, here T)

Return identification：Species 1

**Case 2**. When user input is different to any of sequences in database, but barcoding sites fit one of species, program will return identification based on barcoding sites while show warning information on those unexpected mutations, *e.g.*:

User input: ATTAC

Return identification：Species 2 (mutation A ignored)

User input: TTTC and TTT-C after alignment (gap ignored)

Return identification：Species 2

**Case 3.** When user input is different to any of sequences in database, including barcoding

sites, program will report error, *e.g.*

-User input: TTGAC (mutation at barcoding site)

Return error: “unexpected gaps or mutation sites in barcoding site. Identification stopped. Please check your sequence”

-User input: TTAC, after alignment TT-AC (contain gap in barcoding site)

Return error: “unexpected gaps or mutation sites in barcoding site. Identification stopped. Please check your sequence”

-User input: TTRAC (ambiguous site) or TTNAC (Unknown site)

Return error: “unexpected gaps or mutation sites in barcoding site. Identification stopped. Please check your sequence”

### 2.2.7 Sequences database

-*Criterions for sequences selection*:

Firstly, the sequences in PPNID database were selected form GenBank based on following criterions:

1: Sequences from its original species description, or topotype.

2: Sequences from a taxonomic research article with and/or with detailed morphology backup.

3: Sequences from a non taxonomic research article but published in peer-viewed journal with sufficient supports for species identification.

4: When not sequences available from criteria1-3, other available sequences in GenBank are also used, but only when its identity can be confirmed (by alignment and placement in phylogeny)

Then, we manually examined all candidate sequences quality by alignment and compared their phylogeny placement with some of available references. We also checked the possible synonyms and species name of unlabeled sequences.

Finally, the database will be manually updated in twice a year frequency according latest publications/researches.

-*Database for RKN*

The RKN sequences used in PPNID database are listed in following table. The three sequences from genus *Pratylenchus* and *Zygotylenchus* are also included as outgroup for

phylogeny tree reconstruction. The list of sequences included in database of genus *Meloidogyne* (with *Pratylenchus* and *Zygotylenchus* as outgroup)*:*

| Species name | 18S | 28S | Cox1 | ITS |
|---|---|---|---|---|
| *Zygotylenchus guevarae* | AF442189 | JX261956 | / | FJ717817 |
| *Pratylenchus oleae* | KJ510864 | KJ510856 | KJ510866 | KJ510862 |
| *P. vulnus* | KC875383 | EU130886 | GQ332425 | FJ713007 |
| *P. scribneri* | EU130811 | EU130865 | KY424093* | KT873860 |
| *M. ichinohei* | AF442191 | KY433427 | KY433448 | |
| *M. africana* | KY433422 | KY433423 | KY433446 | KY433429 |
| *M. artiellia* | KC875392 | KY433426 | KY433447 | AF248478 |
| *M. mali* | KJ636400 | KF880399 | KY433450 | KR535971 |
| *M. baetica* | KP896296 | AY150367 | / | AY150366 |
| *M. camelliae* | JX912884* | KF542870 | KF542873 | KF542872 |
| *M. coffeicola* | HE667739 | / | / | |
| *M. hapla* | AY593892 | KP901086 | KU517171 | AY281854 |
| *M. incognita* | AY284621 | AF435794 | KU372164 | KY882489 |
| *M. fallax* | AY593895 | JN157848 | KU517182 | KP825332 |
| *M. graminicola* | KR234083* | KF201162 | KJ139963 | KY020414 |
| *M. chitwoodi* | KJ130033 | JN019321 | KJ476150 | JN241865 |
| *M. enterolobii* | KP901058 | KP901079 | KU372161 | JF309158 |
| *M. naasi* | KJ636260 | KC241979 | KM491211 | JN157862 |
| *M. aberrans* | MF278755 | MF278754 | MF278757 | MF278757 |
| *M. luci* | LN713298 | KF482369 | MF280975 | KF482363 |
| *M. daklakensis* | | KU243338 | MG341340 | MG266058 |
| *M. kralli* | KJ636370 | | | |
| *M. spartelensis* | KP896295 | KP896292 | | KP896294 |
| *M. spartinae* | EF189177* | | | |
| *M. graminis* | JN241838 | JN157850 | | JN241896 |
| *M. duytsi* | KJ636385 | | KU517177 | |
| *M. microtyla* | AF442198 | | | |
| *M. partityla* | KT825143 | | | KR047556 |
| *M. oryzae* | AY942631 | KY962662 | | KY962654* |
| *M. minor* | EU669937 | JN019322 | | KC241953 |
| *M. exigua* | AF442200 | KF993622 | | |
| *M. hispanica* | HE667741 | KF501128 | JX683713 | JX885741* |
| *M. ethiopica* | FJ559408 | KF482372 | KU372162 | KF482366 |
| *M. arenaria* | AY942623 | AF435803 | NC026554 | GQ395518 |
| *M. izalcoensis* | HE667743 | KF993621 | | |
| *M. konaensis* | HE667744 | AF435797 | | KY911100 |
| *M. paranaensis* | AY942622 | AF435798 | | KY911111 |
| *M. javanica* | AY942626 | KP901084 | KU372169 | AF387094 |
| *M. floridensis* | AY942621 | AY194853 | | |

| | | | | |
|---|---|---|---|---|
| *M. ardenensis* | EU669946 | | KY433452 | |
| *M. arabicida* | KF993648 | KF993624 | | |
| *M. lopezi* | KF993645 | KF993616 | | |
| *M. inornata* | | KF482374 | KU372168 | KF482368 |
| *M. silvestris* | EU570215 | EU570214 | | EU570216 |
| *M. marylandi* | JN241856 | JN157852 | | JN157855 |
| *M. christiei* | KR082316 | KR082317 | | |
| *M. phaseoli* | | KY882487 | | KY882498 |
| *M. polycephannulata* | | KY882480 | | KY882491 |
| *M. morocciensis* | AY942632 | KY882485 | | KY882497 |
| *M. cruciani* | HE667740 | | | |
| *M. salasi* | | KY962665 | | KY962656 |
| *M. trifoliophila* | | AF435801 | | JX465593* |
| *M. haplanaria* | | | KU174206 | KU174207 |
| *M. megadora* | | | | KU559896 |
| | | | | |

*-Database for other PPN*

For the PPN other than RKN, only included genera is listed:

## Tylenchomorpha

**Anguinidae** (18S, 28S, ITS and COI): *Anguina Ditylenchus Subanguina Mesoanguina*

**Belonolaimidae** (18S, 28S): *Belonolaimus Morulaimus Carphodorus Ibipora*

(18S, 28S, ITS and COI): *Criconema Criconemoides Ogma Mesocriconema Lobocriconema*

(ITS): *Belonolaimus Morulaimus Ibipora*

**Criconematidae** (18S, 28S) *Crossonema Hemicriconemoides*

(18S, 28S, ITS) *Xenocriconemella*

(18S) *Discocriconemella*

(ITS, COI) *Neobakernema*

**Dolichodoridae** (18S, 28S, ITS and COI): *Tylenchorhynchus*

(18S, 28S, ITS): *Bitylenchus Pratylenchoides*

(18S, 28S): *Scutylenchus Merlinius Paratrophurus Trophurus Nagelus*

(28S): *Dolichodorus Amplimerlinius*

(18S): *Geocenamus*

**Hemicycliophoroidea** (18S, COI): *Hemicycliophora*

(18S) *Hemicaloosia, Caloosia*

**Heteroderidae** (18S, 28S, ITS, COI): *Heterodera Globodera Cactodera Punctodera Cryphodera*

(28S, ITS, COI): *Atalodera Cryphodera Meloidodera*

(18S, 28S, ITS): *Vittatidera*

(28S, ITS): *Paradolichodera*

(ITS, COI): *Rhizonemella*

(ITS): *Betulodera*

(28S): *Verutus*

(18S): *Ekphymatodera*

**Pratylenchidae** (18S, 28S, ITS,) *Radopholus Hoplotylus Radopholoides Nacobbus*

(18S, 28S, ITS, COI): *Pratylenchus Hirschmanniella Zygotylenchus*

(COI): *Pratylenchoides*

**Rotylenchulidae** (18S, 28S, ITS, COI): *Rotylenchulus*

(18S, ITS): *Bilobodera*

**Hoplolaimidae** (18S, 28S, ITS, COI): *Helicotylenchus*, *Rotylenchus*, *Hoplolaimus*, *Scutellonema*

(28S) *Peltamigratus*, *Aorolaimus*

**Tylenchuloidea** (18S, 28S, ITS): *Paratylenchus Tylenchulus Meloidoderita*

(18S, ITS): *Tylenchuloidea*

(18S, 28S): *Sphaeronema*

(ITS): *Trophotylenchulus*

## Dorylaimida

**Longidoridae** (18S, 28S, ITS, COI): *Longidorus*, *Paralongidorus*, *Xiphinema*

## Triplonchida

**Trichodoridae** (18S, 28S, ITS, COI): Trichodorus Paratrichodorus

(18S, ITS, 28S): Monotrichodorus Nanidorus

## 3. References

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797

Guindon S., Gascuel O. (2003) PhyML: A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood, *Systematic Biology*, 52(5):696-704

Janssen, T., Karssen, G., Verhaeven, M., Coyne, D., & Bert, W. (2016). Mitochondrial coding genome analysis of tropical root-knot nematodes (Meloidogyne) supports haplotype based diagnostics and reveals evidence of recent reticulate evolution. *Scientific Reports*, 6, 22591.

Janssen, T., Karssen, G., Couvreur, M., Waeyenberge, L., & Bert, W. (2017). The pitfalls of molecular species identification: a case study within the genus Pratylenchus (Nematoda: Pratylenchidae). *Nematology*, 19, 1179-1199