

PPNID: a manually curated reference database and identification tool for plant-parasitic nematode



Prof. dr. Qing Xue
Nanjing Agricultural University, China

Why we need a program for nematode DNA barcoding?



How do most user identify nematode?
Simple BLAST and pick up the highest match

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST® >> blastn suite

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query sequence.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

Or, upload file [选择文件](#) [未选择任何文件](#)

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):
Nucleotide collection (nr/nt)

Organism

Exclude ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to ☐ Sequences from type material



Max score	Total score	Query cover	E value	Id	Accession
1683	1683	100%	0.0	100.00%	EU284023.1
1652	1652	99%	0.0	99.67%	EU284024.1
865	865	99%	0.0	84.23%	EU284030.1
863	863	89%	0.0	85.99%	EU284029.1
863	863	88%	0.0	86.18%	AY443351.1
817	817	88%	0.0	85.15%	AY443354.1
802	802	88%	0.0	85.12%	AY443355.1

Why we can not rely on BLAST for nematode identification?



Reason 1: Current NCBI database is problematic from Nematoda

Heterodera latipons 28S ribosomal RNA, partial sequence

GenBank: FJ151164.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS FJ151164 314 bp mRNA linear INV 30-AUG-2008
DEFINITION Heterodera latipons 28S ribosomal RNA, partial sequence.
ACCESSION FJ151164
VERSION FJ151164.1
KEYWORDS .
SOURCE Heterodera latipons
ORGANISM Heterodera latipons
Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Tylenchida;
Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.

Tylenchorhynchus sp. 1 CCN-2014 isolate SUB304 28S ribosomal RNA gene, partial sequence

GenBank: KJ461559.1

[FASTA](#) [Graphics](#) [PopSet](#)

[Go to:](#)

LOCUS KJ461559 706 bp DNA linear INV 07-OCT-2014
DEFINITION Tylenchorhynchus sp. 1 CCN-2014 isolate SUB304 28S ribosomal RNA gene, partial sequence.
ACCESSION KJ461559
VERSION KJ461559.1
KEYWORDS .
SOURCE Tylenchorhynchus sp. 1 CCN-2014
ORGANISM Tylenchorhynchus sp. 1 CCN-2014

Fungal *Malassezia* sp. sequences contamination

GenBank

Atetylenchus minor 18S ribosomal RNA gene, partial sequence

GenBank: KP730045.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS KP730045 1089 bp DNA linear INV 22-SEP-2015
DEFINITION Atetylenchus minor 18S ribosomal RNA gene, partial sequence.
ACCESSION KP730045
VERSION KP730045.1
KEYWORDS .
SOURCE Atetylenchus minor

Heterodera cynodontis isolate CD274 clone 2 actin gene, partial cds

GenBank: EU284023.1

[FASTA](#) [Graphics](#) [PopSet](#)

[Go to:](#)

LOCUS EU284023 911 bp DNA linear INV 26-JUL-2016
DEFINITION Heterodera cynodontis isolate CD274 clone 2 actin gene, partial cds.
ACCESSION EU284023
VERSION EU284023.1
KEYWORDS .
SOURCE Heterodera cynodontis
ORGANISM Heterodera cynodontis
Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Tylenchida;
Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
REFERENCE 1 (bases 1 to 911)

Sequence are labelled as unidentified species/genus while they have been formally described to species level

Raw data were not properly assembled or trimmed

The identity of the sequence has changed, but the old name still remained



Reason 2: Not all user known how to choose a gene to BLAST. Not all species can be identified by BLAST

<input type="checkbox"/>	Meloidogyne javanica isolate R21 large subunit ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	MF673747.1
<input type="checkbox"/>	Meloidogyne morocciensis isolate MOR1 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KY882485.1
<input type="checkbox"/>	Meloidogyne polycephannulata isolate POLY 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KY882480.1
<input type="checkbox"/>	Meloidogyne incognita isolate INCS2 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KY882479.1
<input type="checkbox"/>	Meloidogyne incognita isolate INC1 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KY882477.1
<input type="checkbox"/>	Meloidogyne arenaria 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KY293688.1
<input type="checkbox"/>	Meloidogyne javanica isolate XSBN-1 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KX646188.1
<input type="checkbox"/>	Meloidogyne javanica voucher VW4 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KP901084.1
<input type="checkbox"/>	Meloidogyne javanica voucher GuMj 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KP901083.1
<input type="checkbox"/>	Meloidogyne incognita voucher 13-509 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KP901072.1
<input type="checkbox"/>	Meloidogyne incognita voucher 12-502 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KP901070.1
<input type="checkbox"/>	Meloidogyne sp. WW-2014 clone 3 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	KM100870.1
<input type="checkbox"/>	Meloidogyne javanica strain WF 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	JX100423.1
<input type="checkbox"/>	Meloidogyne arenaria 28S ribosomal RNA gene, partial sequence	1083	1083	83%	0.0	100.00%	EU364889.1
<input type="checkbox"/>	Meloidogyne arenaria isolate PQ20 28S ribosomal RNA gene, partial sequence	1081	1081	83%	0.0	100.00%	MK026624.1
<input type="checkbox"/>	Meloidogyne arenaria isolate PQ9 28S ribosomal RNA gene, partial sequence	1081	1081	83%	0.0	100.00%	MK026622.1

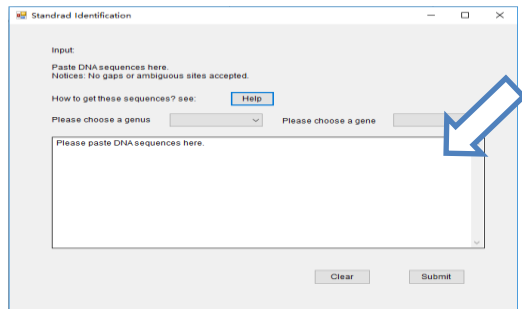
Several RKN can not be identified by rRNA genes!



PPNID: a user friendly solution for plant-parasitic nematode identification



User interface of homepage



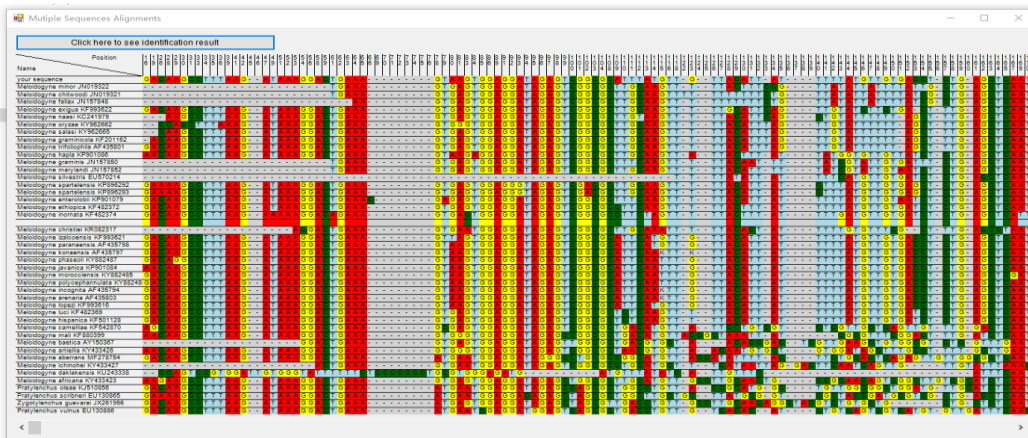
Window for user sequence input



Help page: information for primer, PCR condition, and evaluation of sequencing quality are included.

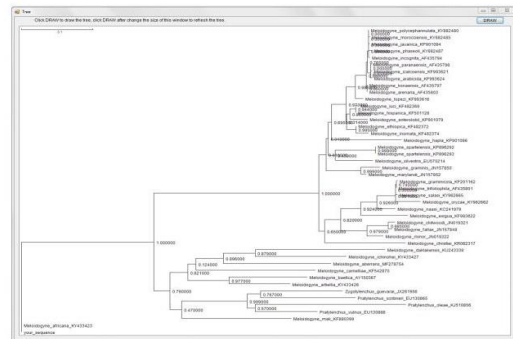
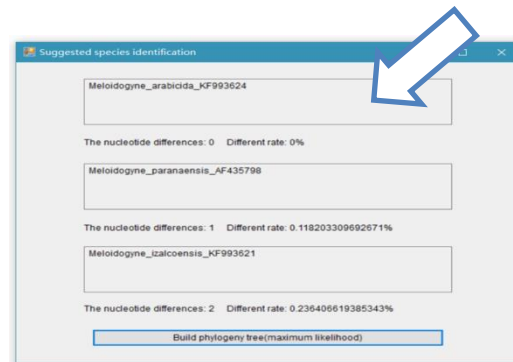


MUSCLE based alignment, p-distance based similarity ranking, maximum likelihood phylogeny reconstruction



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

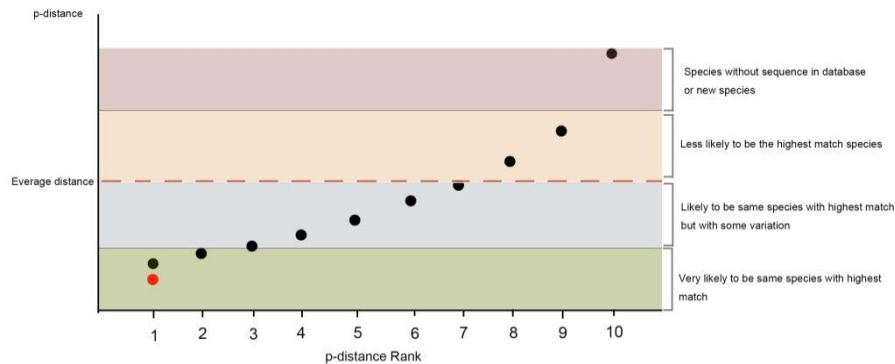
Database	p-distance matrix			
sp. 1: A T T G C C G T T A	sp.1	sp.2	sp.3	sp.4
sp. 2: - T T G C A G T T -	0	0.125	0.2	0.4
sp. 3: A T T G G - G T T A	0.125	0	0.25	0.5
sp. 4: A T T A T T C T T A	0.2	0.25	0	0.25
	0.4	0.5	0.25	0



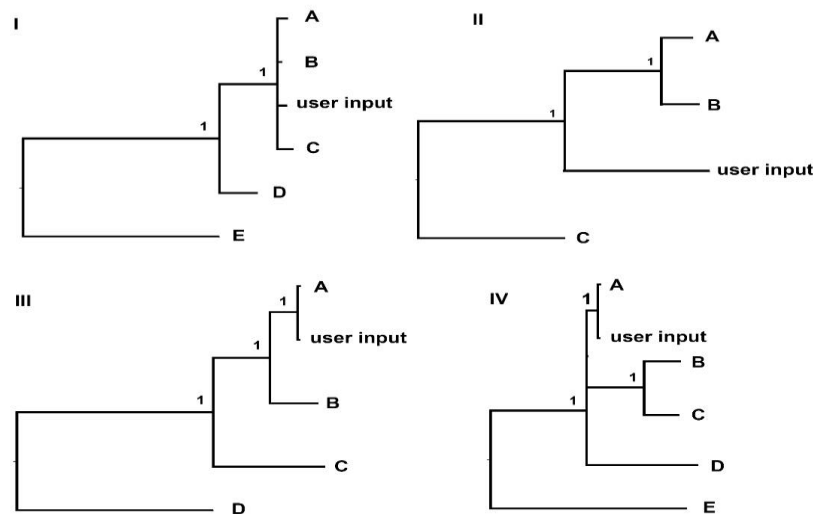
Qing et al., 2019 *Bioinformatics*



Automatic species identification: improved phylogenetic species concept



$$\overline{x_j} = \frac{1}{5} \sum_{k=1}^m d_{j,k}$$





SNP based tropical root-knot nematodes identification

利用线粒体单倍型SNP鉴定根结线虫

Quick start: tropical root-knot nematodes

Input

Paste DNA sequences here.
Notices: Only NAD5 is accepted here. No gaps or ambiguous sites accepted.

How to get NAD5 sequences? see: [Help](#)

Please paste DNA sequences here.

Clear Submit

User input sequence



Automatic alignment and SNP searching



Iden

Your input is:

Meloidogyne enterolobii

Ok

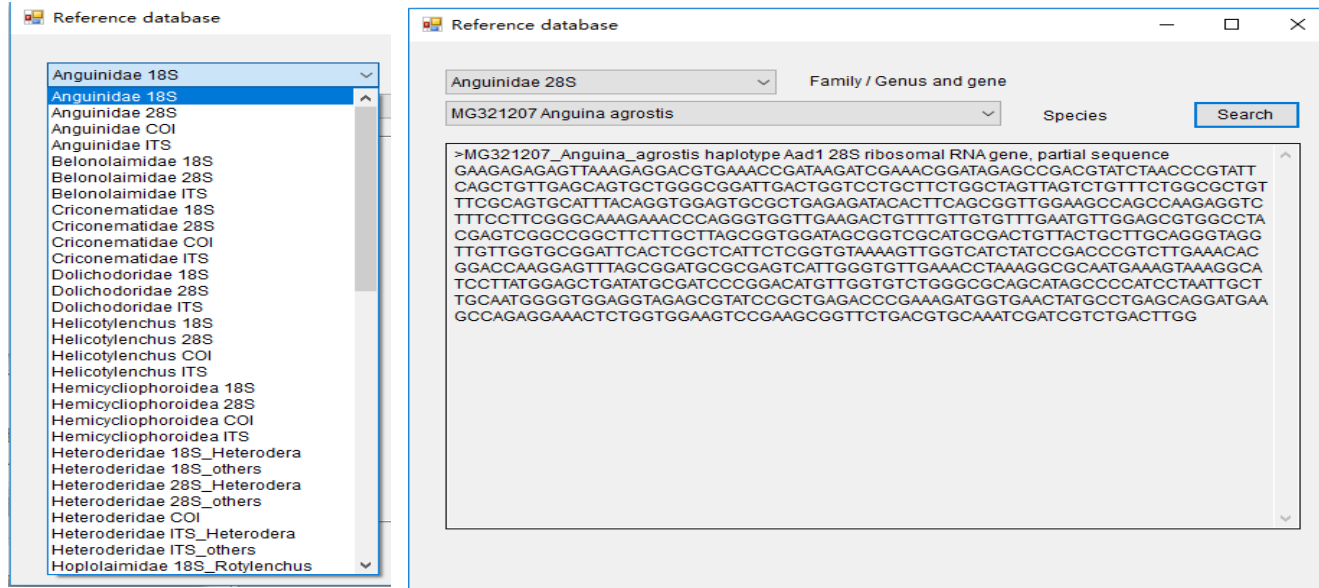


The executable file along with tutorials is available at

<https://github.com/xueqing4083/PPNID>. 程序包及其开源代码可在Github中免费下载



Manually curated reference database



The database includes all current available barcoding sequences for PPN that passed verification. It is freely available and will update two time a year.

Troubleshooting 1: path error



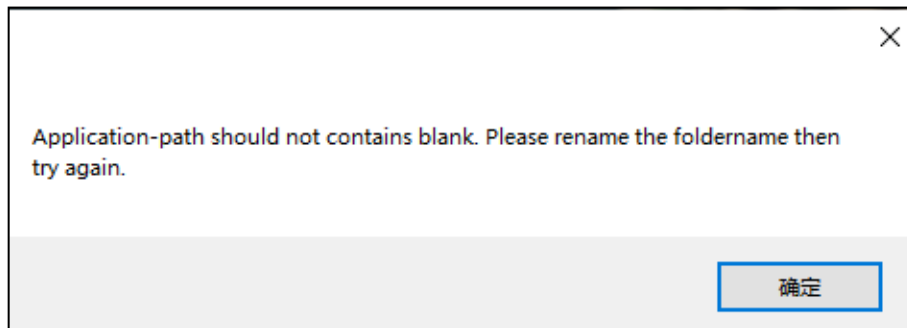
Path with a illegal space:

C:\Users\xue qi\PPNIDQing2.24



Correct path:

C:\Users\xueqi\PPNIDQing2.24

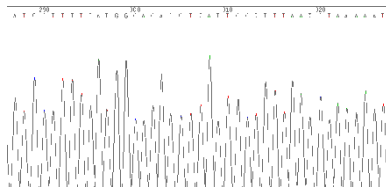
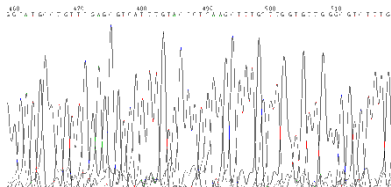


Solution: remove the space in path

Troubleshooting 2: gaps/sequence quality error

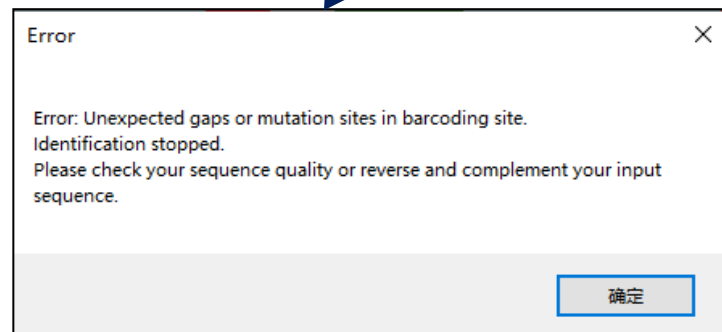
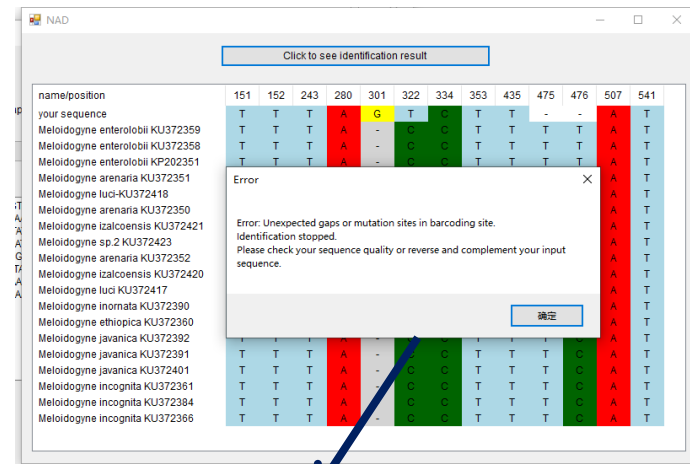
This error means:

- (1) The input sequence is in negative-sense (most likely)
- (2) The quality of sequence is low, result in a unknown pattern in barcoding sites
- (3) You find a new NAD5 pattern (less likely)



Solution: first check the quality of raw sequence, If ok then reverse and complement your input

Online tool for reverse and complement:
<http://reverse-complement.com/>





WeChat

For more information please contact Dr. Qing Xue

 **YouTube** <https://www.youtube.com/watch?v=txqkDtRBsNk>

Researchgate: www.researchgate.net/profile/Qing_Xue6

E-mail: xueqing4083@gmail.com ; qingxue@njau.edu.cn