# <Name of application> – A tool for helminth species delimitation at various taxonomic levels

Background information:

Defining species boundaries for helminths have confounded researchers for decades. The vast diversity of helminth species, along with the presence of species complexes and cryptic species contributes to challenges in species delimitation for helminths (1). With species as the basic unit in taxonomy, species boundaries are thus important for accurate species identification, disease diagnosis, and its role in the ecosystem (2,3). Genetic distances, through the difference in sequence variation between two sequences, are the simplest way to gauge if species are conspecific. A general value of 10% genetic variation using mitochondrial genetic markers between two sequences is deemed as belonging to different species (4). Researchers usually rely on previous genetic distance of organisms of related taxa that have been studied to estimate whether their target taxa of interest are conspecific or not. However, estimates to determine what constitutes ‘sufficient’ sequence variation varies among taxa and across taxonomic levels, and there is currently no estimate of sequence variation across taxonomic boundaries for helminths of medical importance to humans and animals.

To circumvent these gaps, the purpose of this application is to aid in helminth species delimitation through the cut-off genetic distance values that were estimated using the unsupervised machine learning K-means clustering algorithm. <Name of application> serves as a convenient tool for species delimitation across the taxonomic hierarchy levels for nematodes, trematodes, and cestodes.

What the application does:

<Name of application> uses the data obtained from Chan et al. (2021), where estimated cut-off genetic distance values for nematodes, trematodes, and cestodes were generated for ten genetic markers using the K-means clustering algorithm (5). The ten genetic markers are: nuclear 18S rRNA gene, nuclear 28S rRNA gene, nuclear ITS1 region, nuclear ITS2 region, mitochondrial 12S rRNA gene, mitochondrial 16S rRNA gene, and mitochondrial protein-coding genes *COI*, *COII*, *NAD1*, *cytB*. Briefly, cut-off genetic distances for each group of helminths per genetic marker were estimated using the K-means algorithm implemented in Wolfram Mathematica 12.1 (6). The minimum and maximum genetic distances were obtained for each taxonomic hierarchy level as a basis for taxonomic boundaries.

<Name of application> allows users to input their genetic distances obtained from their taxa of interest from the three groups of helminths. The queried value will be compared against the database of estimated cut-off genetic distance, and the user will be able to know where their queried value is positioned and which taxonomic hierarchy level it falls into. Additionally, depending on the genetic distance, users will be able to determine if their taxa are conspecific or not. The usefulness of <Name of application> lies in its convenience and simplicity, with just the input of genetic distances required for the taxa in question. Further recommendations may also be provided, depending on the outcome of the query.

How to use:

Users input the genetic distance value (between 0 to 1) obtained from the pairwise comparison between two taxa. The genetic distance value should ideally be obtained using p-distance, which is the proportion of nucleotide sites that are different for the two sequences being compared (7). The helminth group of interest and genetic marker used should be selected from the drop-down menu.

A graphical visualization of the queried genetic distance will be presented, with the dotted line representing the queried value. The range of genetic distances for each taxonomical hierarchy level is represent with the different colored bars. If the queried genetic distance falls within the estimated genetic distance range, the appropriate result will be presented. However, if the queried genetic distance falls out of the estimated genetic distance range, further recommendations on subsequent steps to take and other genetic markers is provided.

Assumptions:

1. The application only takes into account genetic distances, and does not use other information such as morphological characters or phylogenetic placements to determine taxonomic boundaries. However, the application can serve as a quick indicator, with recommendations on further steps to take.
2. The genetic distances obtained from Chan et al. (2021) were mined from the NCBI database, and covers only certain groups of medically important helminths (5). Since the initial aim of Chan et al. (2021) was to assess and evaluate the ten genetic markers, only species (or genera in some instances) that had sequences for all the ten genetic markers were used for comparison. More details can be found in the publication.
3. The data obtained assumes that the species identity for the sequences are correct based on the information from the NCBI database.
4. Recommendations on further steps to take are subjective and should be dependent on the aim of the user. The recommendations in the application are based on the results obtained from the queried genetic distance, the utility and limitations for helminth molecular-based studies from the authors’ comparison of genetic markers, and the authors’ opinions.

References:

1. Carlson CJ, Dallas TA, Alexander LW, Phelan AL, Phillips AJ. (2020) What would it take to describe the global diversity of parasites? Proceedings of the Royal Society B. Biological Sciences. 287(1939)
2. Wiens JJ. (2007) Species delimitation: new approaches for discovery diversity. Systematic Biology. 56(6):875-878
3. De Queiroz K. (2007) Species concepts and species delimitation. Systematic Biology. 56(6):879-886
4. Blouin MS. (2002) Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. International Journal for Parasitology. 32(5):527-531
5. Chan AHE, Chaisiri K, Saralamba S, Morand S, Thaenkham U. (2021) Assessing the suitability of mitochondrial and nuclear DNA genetic markers for molecular systematics and species identification of helminths. 14(233)
6. Wolfram Research Inc. Mathematica version 12.1. Champaign: Wolfram Research, Inc. 2020
7. MEGA software. Distance Estimation. <https://www.megasoftware.net/mega1_manual/Distance.html>