International Journal for Parasitology 56 (2026) 104702

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

# Variability among taxonomists in helminth species discrimination decisions: a noise audit

Robert Poulin *, Jerusha Bennett, Bronwen Presswell

*Department of Zoology, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand*

ARTICLE INFO

*Article history:*
Received 1 May 2025
Received in revised form 4 June 2025
Accepted 15 June 2025
Available online 18 June 2025

*Keywords:*
Biodiversity
Genetic data
Helminths
Morphology
Research experience
Trematodes

ABSTRACT

Determining whether or not superficially similar helminth specimens belong to the same species can be challenging, even for expert taxonomists. The possibility of cryptic species and host-induced morphological variation, combined with the lack of universally accepted thresholds for what can be considered intraspecific genetic variation, are largely to blame. In the end, decisions come down to the judgment of taxonomists. As with other domains of human judgment, however, taxonomic decisions are subject to noise, i.e., differences of opinions among taxonomists when presented with the same evidence. Here, we quantify this noise and test the role of past experience in taxonomic decision-making. We presented morphological, genetic and host data on 15 sets of hypothetical but realistic trematode specimens, each split into two groups, and asked many of the world's top trematode taxonomists to decide whether the two groups belonged to the same species, to different species, or they were not sure. Working independently on the exact same information, the taxonomists rendered species delimitation decisions that were largely inconsistent with each other, and unrelated to their past experience (measured as years of experience or numbers of published species descriptions). The inevitable conclusion is that whether two sets of trematode specimens are considered to represent the same species or two different species depends entirely on the particular taxonomist who examines them. We propose three strategies to reduce noise and achieve greater consistency and repeatability in species delimitation among different taxonomists: establishment of clear species discrimination guidelines, decomposition of the evidence into its separate components prior to a final decision, and aggregation of independent judgements from two or more experienced taxonomists. Limiting subjectivity in species delimitation decisions is essential if taxonomy is to continue underpinning other disciplines, from biodiversity and ecological research to conservation biology and wildlife management.

## 1. Introduction

Whenever hosts from a previously non-studied species or locality are necropsied and new helminth specimens are obtained, it is rarely immediately clear whether these helminths belong to a previously described species or they represent a new species. Making this distinction often poses challenges. For example, consider trematodes (Platyhelminthes: Neodermata). On one hand, there appears to be a higher frequency of cryptic species, i.e. species that are distinct genetically but nearly indistinguishable morphologically, among trematodes than in other helminth groups (Poulin, 2011; Pérez-Ponce de León and Poulin, 2018). On the other hand, substantial intraspecific morphological variation is not uncommon in trematode species, often associated with the identity of the definitive hosts from which different individuals are retrieved (Pérez-Ponce de León, 1995; Hildebrand et al., 2015; Presswell and Bennett, 2019; Cribb et al., 2022); thus, different-looking individuals may actually be conspecifics. When taxonomic experts are presented with a new set of specimens, making full use of all available evidence (the 'integrative taxonomic approach') and species delineation tools is crucial to reach the most robust decision regarding species identity. Assessments of parasite biodiversity depend on accurate taxonomic decisions (Carlson et al., 2020). The same is true for epidemiological monitoring and the tracking of invasive species (is this the same parasite species cropping up in different places, or are we dealing with multiple related species?), parasite conservation initiatives (is this declining parasite population unique, or does it belong to a widespread species?), and the management of diseases in aquaculture (which parasite are we looking at, the pathogenic species or its harmless relative?).

Recently, Cribb et al. (2025) have reviewed the 'pillars' of evidence based on which trematode species are characterised and identified. First, the existence of morphological differences

---

* Corresponding author.
*E-mail address:* robert.poulin@otago.ac.nz (R. Poulin).

https://doi.org/10.1016/j.ijpara.2025.06.003
0020-7519/

between individual specimens is usually indicative of different species, albeit cryptic species and intraspecific variation create difficulties. Second, the identity of the definitive host has also long been used to inform species identification and delineation; however, the lack of strict host specificity in many trematode species casts doubts over the reliability of host identity as a criterion for species recognition. Third, the geographic site of collection may also be informative, as trematode specimens collected in very distant localities are more likely to represent distinct species. However, reports of trematodes with cosmopolitan distributions (e.g., Aiken et al., 2007; Wee et al., 2022) weaken any argument in favour of geography as a reliable criterion. Lastly, DNA sequencing provides a source of information independent of morphology, host identity or geography (Blasco-Costa et al., 2016). Using a suitable nuclear or mitochondrial marker, conspecific individuals should show little to no sequence variation, whereas separate species are expected to show consistent sequence differences. Several algorithms can be applied to a large set of sequences to automate the species delimitation process (e.g., Puillandre et al., 2012, 2021; Fujisawa and Barraclough, 2013); these have been applied in some instances to trematode sequence data (e.g., Pérez-Ponce de León et al., 2016). However, these automated methods do not always provide congruent results (Carstens et al., 2013), and much user interpretation is still required.

Whether considered alone or combined, the above sources of evidence still require expert assessment and judgment (Cribb et al., 2025). However, wherever human judgment is involved, there will be bias and noise. Bias refers to systematic deviation from the correct decision in one particular direction, whereas noise refers to variation in judgment, as when different experts who are expected to agree when shown the same evidence end up reaching very different conclusions. The noise around expert decisions affects a broad range of real-world situations, including sentencing by court judges, diagnoses made by doctors, opinions of forensic experts, and recommendations of insurance claims assessors (see review in Kahneman et al., 2021). In the case of species recognition and delineation, bias manifests itself when certain experts have a tendency to either detect multiple species in the morphological and genetic data they examine, or a tendency to view the data as representing a single species, more frequently than the average across the community of taxonomic experts. This corresponds to the old ‘splitter versus lumper’ dichotomy affecting how living organisms are classified by experts with different tendencies (Simpson, 1945; Endersby, 2009). Noise, in contrast, is measured as the extent of the variation among the decisions of experts, regardless of bias. The existence of noise in taxonomic decisions has long been recognised (e.g., Moss, 1971). For instance, an earlier study has shown that there is much disagreement among taxonomists regarding the validity of putative new species after reading summaries of the relevant research (Conix et al., 2023). However, this focused on their opinions of past decisions made by others, without quantifying noise in the decision-making process itself.

To improve the quality of taxonomic decisions, we must overcome noise as well as bias. Here, we focus on noise. Perhaps counterintuitively, noise is not always bad: a certain amount of disagreement among experts can be beneficial if it leads to discussion among those with different perspectives and the refinement of objective decision-making processes. However, how much noise is there currently among parasite taxonomists? How much variation is there among taxonomists presented with the same set of specimens in deciding whether they represent the same or different species? How likely is it that the erection of a new species depends on which expert happens to be looking at the material? And can the nature of the evidence presented to taxonomists influence the noisiness of their decisions? Before we can take steps to reduce noise in taxonomic decisions, we must answer these questions. This is the goal of the present study. We conducted a noise audit (*sensu* Kahneman et al., 2021) of taxonomic decisions among a large selection of experts in trematode taxonomy and systematics. Trematodes are used here as a case study, however the implications of our findings and the ensuing recommendations apply more broadly to all other parasite taxa. As our audit reveals, taxonomic decisions suffer from a large amount of noise, with extensive disagreement among taxonomists. We follow the audit by proposing simple noise-reducing strategies aimed at remedying this situation.

## 2. Methods

We asked 40 experts in trematode taxonomy to participate in our study. They were chosen from the list of attendees at the Trematode 2024 conference (Brisbane, Australia), the list of authors of the Cribb et al. (2025) review article, and the list of prolific taxonomists identified by Poulin and Presswell (2022), but also included a few others in order to achieve broad geographic and gender representation. No information was given to each expert contacted regarding the identity of other participants.

When asked to participate, they were told that we were investigating how taxonomists make decisions regarding species discrimination when considering two sets of specimens, i.e. two sets of measurements and genetic data. Each taxonomist was sent 15 sets of hypothetical species descriptions, each consisting of a pair (subset A and subset B) of trematode descriptions that were clearly congeneric and plausibly, but not necessarily, the same species (see Supplementary Material). Although the data used are hypothetical, they were inspired from real species descriptions, with each set corresponding to a particular trematode family. In each case, the data included (i) morphometric measurements, sometimes supplemented with qualitative morphological characters (depending on the family), (ii) genetic data (percentage of base pair differences for different DNA markers), and (iii) information on how closely related the definitive hosts of A and B were (i.e., same genus, different genera but same family, etc.). The taxonomists were instructed to assume that all specimens were fixed, measured and genetically characterised with identical procedures, without errors, and that all sample sizes were equal. Finally, they were told that for each pair of descriptions, all specimens were collected from the same broad geographical region.

The data on morphology and genetic differences were deliberately chosen to create some ambiguity in species discrimination. The 15 pairs of hypothetical specimens were created to present the taxonomists with an approximate gradient ranging from very slight to more pronounced differences in morphology, DNA sequences and host use, but within a range where all pairs conceivably represented the same species (see Supplementary Material). In other words, differences between the two subsets in a pair ranged from trivial to relatively large with respect to the typical intraspecific variability observed in trematodes. The order in which the 15 sets were arranged was randomised for each participant. The taxonomists were asked to evaluate all 15 sets independently, i.e. using their own knowledge and experience, without consulting colleagues. Although they were not equally familiar or knowledgeable about all trematode families represented in our hypothetical data, they were asked to use their best judgement and decide whether the two subsets of descriptions, A and B, in each pair (i) belonged to the same species, (ii) represented two distinct congeneric species, or (iii) they were not sure.

Each taxonomist who responded was then sent a second questionnaire, in which they were asked (i) what was the main principle or method they used to reach a decision across all cases, i.e.

genetic divergence, morphological differences, both genetic and morphological differences, or other; (ii) how many trematode species they have described either as main taxonomist or co-author; and (iii) in what year they published their first description of a trematode species.

In addition to basic descriptive data, simple statistical tests were used to test for general patterns. Since respondents took from a few days to several weeks to reply, we used a Spearman correlation coefficient to test whether the number of "same species" decisions, the number of "different species" decisions, and the sum of same and different species decisions, were related to the time the experts took to reply. In addition, we used Spearman correlation coefficients to test for relationships between a taxonomist's experience, measured as both the total number of trematode species they have described and the number of years since their first published description, and either the number of "same species" decisions, the number of "different species" decisions, or the sum of same and different species decisions they rendered.

## 3. Results

We received responses from 22 expert trematode taxonomists (12 male, 10 female) from 11 countries, out of the 40 we initially contacted (three declined to participate, the rest never replied). Based on the information they provided, the number of trematode species they described ranged from fewer than 10 to over 160 trematode species per expert (average = 25.8 species). From the first trematode species description they published to the present, they have accumulated an average of 14.5 years of experience in taxonomy, ranging from less than 5 to more than 40 years.

With 22 taxonomists each examining 15 cases, we received 330 (22 × 15) expert judgments regarding trematode species discrimination (Fig. 1). Of these, 57 (17.3%) decisions were that the two subsets of specimens were the same species, 217 (65.7%) were that they were distinct species, and 56 (17.0%) were cases where the respondents were unsure. Strikingly, all 22 sets of decisions we received were unique (i.e., no two columns are identical in Fig. 1). In other words, if we take the decisions made by any expert and try to match them with those of any other expert, i.e., if we pair the experts two by two for a total of 231 possible pairs of experts among the 22 respondents, no two experts returned identical sets of decisions.

There were general differences among expert taxonomists with respect to their tendency to consider the data we sent them as representing the same or different trematode species (Fig. 2). Some individuals were more likely to conclude to "same species" (left hand side of Fig. 2), whereas others were much more likely to consider them as "different species" (right hand side of Fig. 2, notably experts #1 and 21, and especially #14). Most respondents provided a more balanced series of decisions, while still leaning toward "different species", with one expert (#7) being broadly indecisive based on the evidence provided.

Although for some families we had generated data for two subsets of specimens we thought were very similar and likely to be considered "same species" (top of Fig. 3) while for other families we felt the data were less similar and likely to be considered "different species" (bottom of Fig. 3), the decisions obtained from expert taxonomists did not really match this expectation. Not a single case out of the 15 presented to the experts received unanimous decisions (Fig. 3), although there was much support for the specimens in some families representing the same species (e.g., Derogenidae), and much support for the specimens in other families representing different species (e.g., Renicolidae). Every family case received at least seven "different species" decisions, whereas three families (Monorchiidae, Renicolidae and Azygiidae) received no "same species" decisions.

There was no correlation between the order in which the experts responded to our survey and the number of "same species" decisions ($r_s$ = 0.101, $P$ = 0.656), the number of "different species" decisions ($r_s$ = -0.096, $P$ = 0.670), and the total number of decisions they made (i.e., same or different species as opposed to "not sure") ($r_s$ = 0.068, $P$ = 0.765).

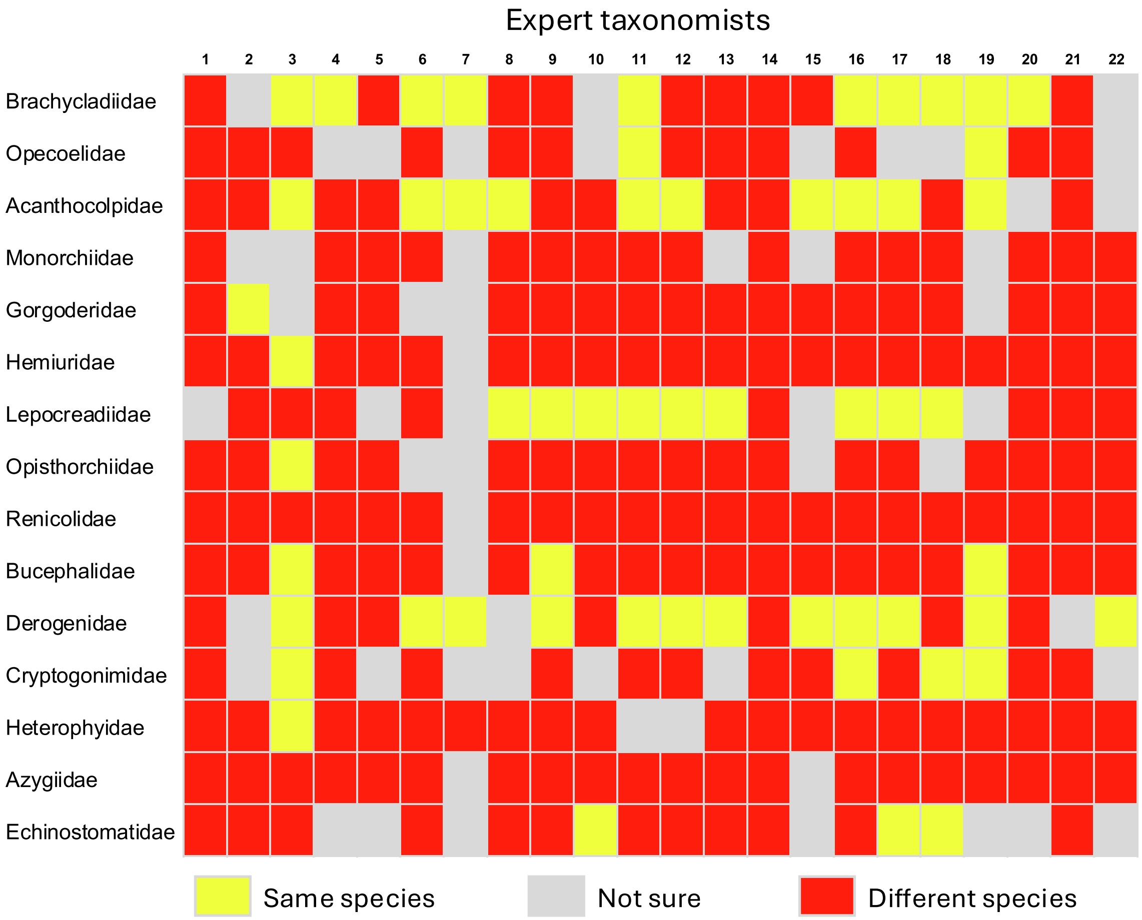
When asked what was the main principle or method they used to reach a decision across all cases, all experts replied that they used a holistic assessment considering all information (morphological, genetic, host identity) we provided, except one who stated focusing mostly on the genetic data.

With respect to the past experience of our respondents (Fig. 4), we found no correlation between the total number of trematode species they have previously described and the number of "same species" decisions ($r_s$ = –0.047, $P$ = 0.837), the number of "different species" decisions ($r_s$ = 0.038, $P$ = 0.867), and the total number of decisions they made (i.e., same or different species combined) ($r_s$ = 0.018, $P$ = 0.937). Similarly, we found no correlation between the number of years since they published their first trematode species description and the number of "same species" decisions ($r_s$ = –0.043, $P$ = 0.850), the number of "different species" decisions ($r_s$ = –0.224, $P$ = 0.315), and the total number of decisions they rendered ($r_s$ = –0.233, $P$ = 0.298).

## 4. Discussion

Taxonomy is the foundation of organismal biology. Accurate species delimitation is essential for research in disciplines such as ecology, biogeography, evolution and conservation biology, to name just a few. For instance, estimates of species diversification rates among branches of the tree of life vary as much as the taxonomic opinions on which they are based (Faurby et al., 2016). But what is a 'species', exactly? Whatever the species concept one chooses, in the end a species is what an expert taxonomist (or a small team of experts) says is a species. Some formally named and described species are later invalidated or synonymised (Poulin and Presswell, 2024). However, the majority of named and described species remain accepted as valid, the assumption being that decisions taken by vastly different individual taxonomists are equivalent and consistent. Here, we have shown that this is not the case, far from it. Many taxonomic decisions appear to be subjective, that is, not repeatable across different individual taxonomists. The key take-home message of our study is that whether two sets of trematode specimens are considered to represent the same species or two different species depends on the particular taxonomist who examines them. This has always been a tacitly recognised but inconvenient truth. Nevertheless, our findings unveil the magnitude of the noise in taxonomic decision-making, revealing it as an important source of uncertainty about 'true' trematode diversity.

In brief, our main findings were, first, confirmation that there is substantial variability (noise in decision-making) among experts in trematode taxonomy when they are presented with the exact same evidence. As a consequence, estimates of total species diversity among all our hypothetical datasets would have ranged from as low as 20 species (from expert #3), to the maximum of 30 species (expert #14 viewing all 15 datasets as representing two distinct species each). Thus, idiosyncratic differences among taxonomists working in different regions of the world can lead to vast differences in estimates of regional parasite diversity. Second, we confirm that trematode taxonomists range across the continuum from splitters to lumpers (*sensu* Simpson, 1945). Among all decisions received, there was a clear overall tendency for a higher frequency of splitting, i.e. considering the two subsets of specimens as distinct species, although this may be due to the hypothetical data we produced. Third, there was no relationship between experience



Expert taxonomists

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brachycladiidae | 🟥 | ⬜ | 🟨 | 🟨 | 🟥 | 🟨 | 🟨 | 🟥 | 🟥 | ⬜ | 🟨 | 🟥 | 🟥 | 🟥 | 🟥 | 🟨 | 🟨 | 🟨 | 🟨 | 🟨 | 🟥 | ⬜ |
| Opecoelidae | 🟥 | 🟥 | 🟥 | ⬜ | ⬜ | 🟥 | ⬜ | 🟥 | 🟥 | ⬜ | 🟨 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | ⬜ | ⬜ | 🟨 | 🟥 | 🟥 | ⬜ |
| Acanthocolpidae | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟨 | 🟨 | 🟨 | 🟥 | 🟥 | 🟨 | 🟨 | 🟥 | 🟥 | 🟨 | 🟨 | 🟨 | 🟥 | 🟨 | ⬜ | 🟥 | ⬜ |
| Monorchiidae | 🟥 | ⬜ | ⬜ | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 |
| Gorgoderidae | 🟥 | 🟨 | ⬜ | 🟥 | 🟥 | ⬜ | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 |
| Hemiuridae | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 |
| Lepocreadiidae | ⬜ | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | ⬜ | 🟨 | 🟨 | 🟨 | 🟨 | 🟨 | 🟨 | 🟥 | ⬜ | 🟨 | 🟨 | 🟨 | ⬜ | 🟥 | 🟥 | 🟥 |
| Opisthorchiidae | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | ⬜ | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 |
| Renicolidae | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 |
| Bucephalidae | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 |
| Derogenidae | 🟥 | ⬜ | 🟨 | 🟥 | 🟥 | 🟨 | 🟨 | ⬜ | 🟨 | 🟥 | 🟨 | 🟨 | 🟨 | 🟥 | 🟨 | 🟨 | 🟨 | 🟥 | 🟨 | 🟥 | ⬜ | 🟨 |
| Cryptogonimidae | 🟥 | ⬜ | 🟨 | 🟥 | ⬜ | 🟥 | ⬜ | ⬜ | 🟥 | ⬜ | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟨 | 🟥 | 🟨 | 🟨 | 🟥 | 🟥 | ⬜ |
| Heterophyidae | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 |
| Azygiidae | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 |
| Echinostomatidae | 🟥 | 🟥 | 🟥 | ⬜ | ⬜ | 🟥 | ⬜ | 🟥 | 🟥 | 🟨 | 🟥 | 🟥 | 🟥 | 🟥 | ⬜ | 🟥 | 🟨 | 🟨 | ⬜ | ⬜ | 🟥 | ⬜ |

🟨 Same species ⬜ Not sure 🟥 Different species

**Fig. 1.** Full decision matrix across 22 expert taxonomists (columns) judging 15 hypothetical sets of very similar trematode specimens from different families (rows), showing whether each expert thought that each set represented specimens of the same species or different species, or whether they were not sure. Taxonomists are numbered in the order in which they returned their responses. Trematode family sets are arranged from those thought *a priori* to be most similar (top) to least similar (bottom).

in taxonomy, measured either as the number of species descriptions published or the number of years of experience, and the taxonomists' confidence to reach decisions, or what those decisions were.

We acknowledge that our study had limitations. In a real situation, a taxonomist might have more available information prior to rendering a decision than what we provided. As pointed out by some respondents, a phylogenetic analysis testing whether the two subsets of specimens are reciprocally monophyletic would be useful, as would illustrations showing the shape, relative sizes and position of internal organs, and data on the genetic variation within each subset of specimens relative to the variation between the two subsets. More data are always better, however complete information is almost never available. For instance, the identity of intermediate hosts is unknown for the vast majority of described trematode species, and the availability of usable DNA sequences for other congeneric species is patchy for most trematode families. In the present study, expert taxonomists were given a reasonable amount of data about each subset of specimens and yet they showed considerable disagreement in their conclusions regarding species delimitation. It is likely that providing more information would have led to a greater degree of agreement, however not to total convergence on identical decisions. Furthermore, the participants were given the option of choosing "not sure" as a decision regarding whether or not two subsets of specimens represented the same species, if they felt insufficient information was available.

Another issue is that we used the number of species descriptions published by an expert and the time since their first description as measures of a taxonomist's experience. Of course, relevant experience also involves other taxonomic activities, such as species redescriptions and reclassifications. Nevertheless, we feel that the two measures we use do capture most of the differences in experience among taxonomists. Finally, it must be pointed out that the family names associated with the cases where decisions regarding species delimitation appeared more straightforward (e.g., Renicolidae and Azygiidae) or less clear-cut (e.g., Acanthocolpidae, Lepocreadiidae and Derogenidae) do not reflect the reality for these families; all cases used here were hypothetical and only associated with particular families to make them more realistic.

Disagreements among taxonomists regarding the distinct species status of newly-obtained parasite specimens are common. For instance, nearly one quarter of parasite species names proposed in the past century have later been invalidated, as the species they represented were synonymised with other species (Poulin and Presswell, 2024). Even many helminth species described and named in the past couple of decades have suffered this fate. Over time, new tools, such as electron microscopy and DNA sequencing, have become available to characterise parasite species. Indeed, these have been increasingly utilised in parasite species descriptions in recent decades (Poulin and Presswell, 2016). However, more information does not mean less room for disagreement in taxonomic decisions among experts. Here, we

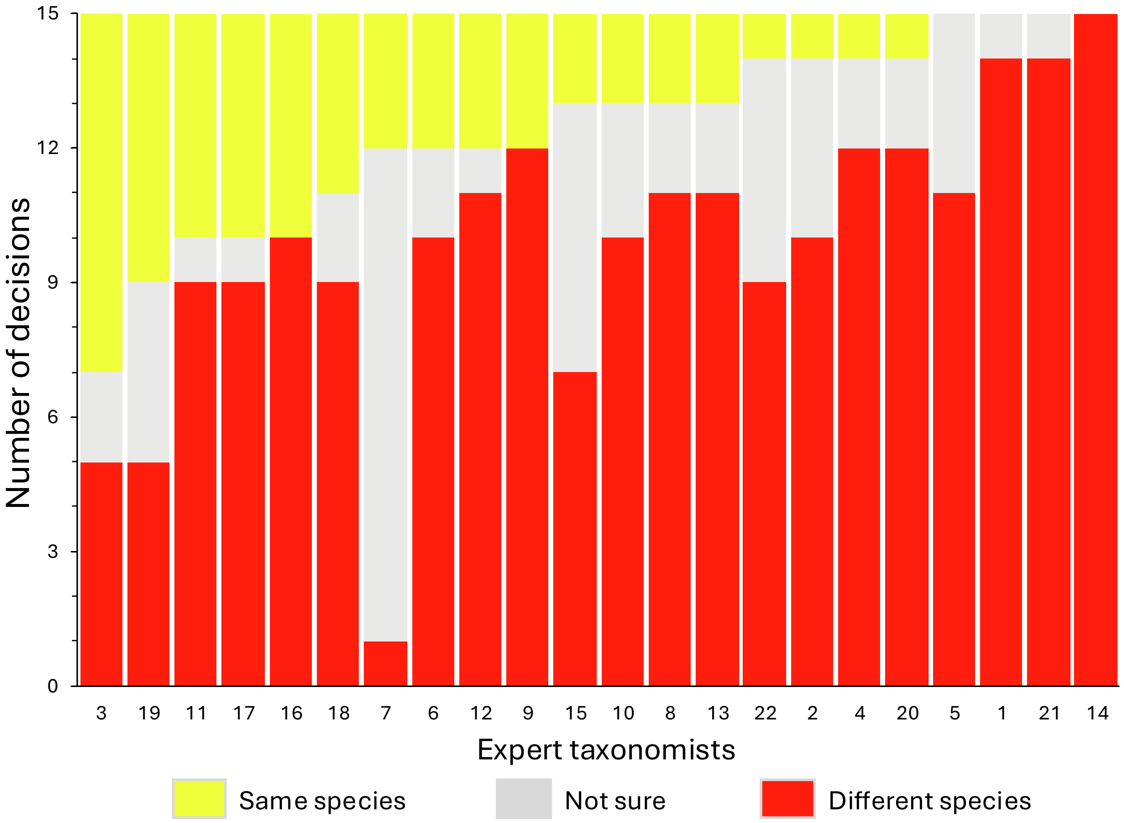


**Fig. 2.** Summary decision results from 22 expert taxonomists judging 15 hypothetical sets of very similar trematode specimens from different families. Taxonomists are ranked from those that rendered the most "same species" decisions (left) to those that rendered the most "different species" decisions (right); number codes for taxonomists reflect the order in which they returned their responses (same as in Fig. 1).

have shown that given the same evidence regarding the morphology, genetics, and host identity of two subsets of hypothetical parasites, no two experts could agree in all cases, and none of our 15 hypothetical cases received unanimous decisions across all experts.

For our purposes, it does not matter whether the 15 sets of pairs of closely-related specimens used in our study, if they were real, would actually represent distinct species or not. To some extent, this may depend on the species concept one chooses to adopt (Kunz, 2002; Tibayrenc, 2006; de Queiroz, 2007). There may be no right or wrong answers in the hypothetical cases we presented to our respondents. What matters is the noise among responses, i.e., the lack of consistent interpretation among expert taxonomists when presented with the exact same evidence. The total noise in decisions regarding species identification and delineation must be greater than what we demonstrate here. We only focused on what has been called system noise (among-expert variation): variability of the average judgments made by different individuals, due to personality, experience or idiosyncratic assessment of the same evidence (Kahneman et al., 2021). The other main component of total noise is occasion noise (within-expert variation), which corresponds to the same person making different judgments of the same evidence on different days, perhaps subconsciously influenced by irrelevant external factors. Therefore, total noise in taxonomic decision-making is likely greater than what we demonstrated here.

Cribb et al. (2025) provide a set of general guiding principles for trematode species delimitation. Here, we suggest the following three concrete steps, applicable not only for trematodes but all other taxa. They represent noise-reduction strategies that go beyond proper training in taxonomic science. First, we need clear practical guidelines for species delimitation, i.e., agreed-upon standardised criteria or reference levels, established at the same kind of international gathering of taxonomic experts that has led to the general principles put forth by Cribb et al. (2025). If expressed quantitatively (e.g., minimum genetic or morphological differences for species differentiation), these guidelines would allow much less room for noisy judgement in species identification and delineation. They should specify the weight to be given to each piece of evidence (morphological similarity, genetic divergence, host relatedness). This approach has been proposed for avian species discrimination, where phenotypic traits of multiple recognised bird species were used to establish species delimitation criteria based on data-driven thresholds (Tobias et al., 2010). The use of clear guidelines can increase the level of objectivity and transparency of taxonomic decisions, while still allowing case-by-case exceptions based on special circumstances.

Second, each species delimitation decision should follow the decomposition of the available evidence. This involves the *independent* assessment of each piece of evidence (morphology, sequence data, host identity, geographic origin), to avoid a quick, intuitive first impression based on one piece of evidence that may influence how the others are then viewed. For example, when no obvious morphological difference is apparent at first glance between two sets of specimens, this observation may then cause one to downplay data indicating some genetic divergence. Each line of evidence must therefore be assigned a pre-determined weight and assessed independently of the others. For this, it is best to structure judgments into separate, independent tasks, ideally separated by several days.

Third, nothing reduces noise more than the aggregation of multiple expert judgements prior to a final decision. For species delimitation, this would involve a discussion among two or more

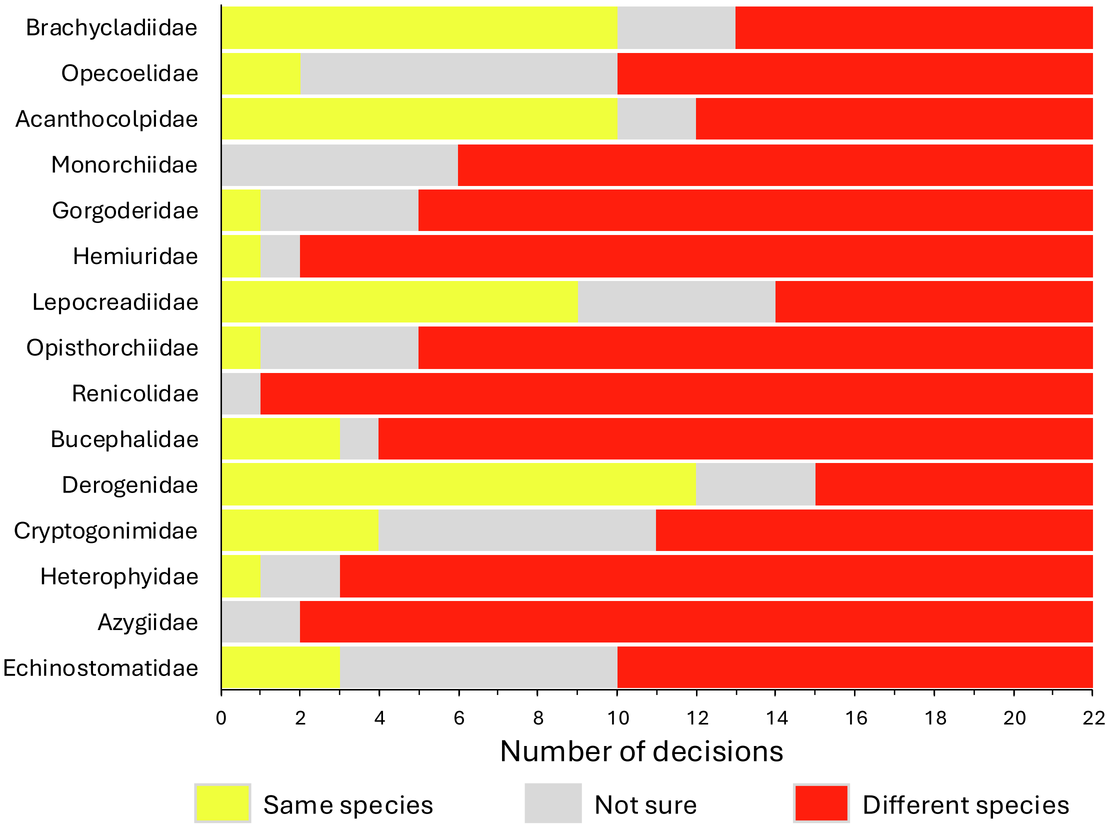


**Fig. 3.** Summary of decisions received from 22 expert taxonomists for 15 hypothetical sets of very similar trematode specimens from different families. Trematode family sets are arranged from those thought *a priori* to be most similar, i.e. expected to receive more "same species" decisions (top) to those thought to be least similar, i.e. expected to receive more "different species" decisions (bottom).

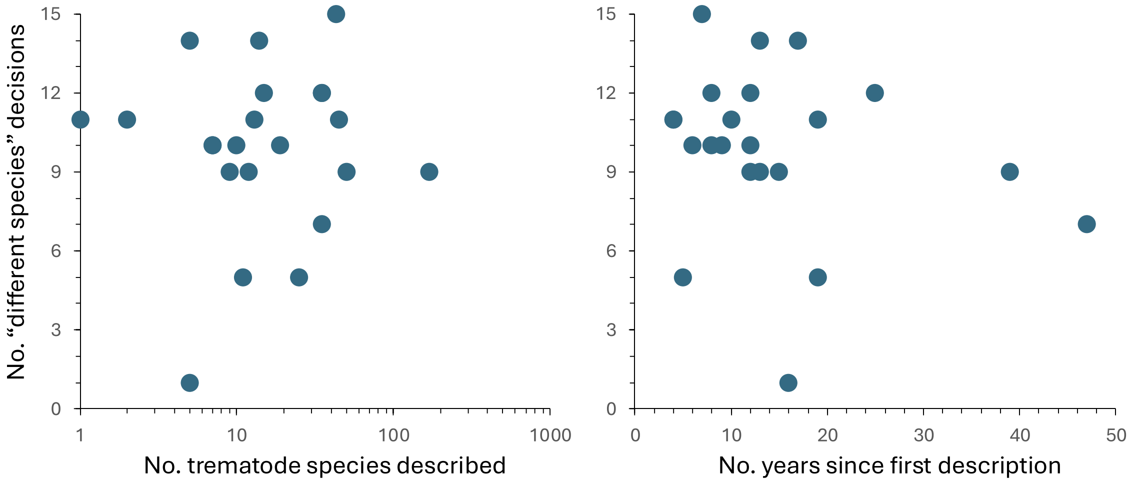


**Fig. 4.** Number of "different species" decisions rendered per taxonomist ($N = 22$) when considering 15 hypothetical sets of similar trematode specimens, as a function of their experience measured either as the total number of trematode species they have previously described (left) or the number of years since they published their first trematode species description (right). Note: the x-axis on the left is on a $\log_{10}$ scale to avoid clumping of data points.

experienced taxonomists toward a consensus *after* each expert has formed an independent opinion. In our study, respondents were instructed to decide on each case independently, i.e., without consulting other experts; undoubtedly, discussing these cases with others would have resulted in much less variation in their final decisions. Most published descriptions of new trematode (or other) species involve a single experienced taxonomist teaming up with junior researchers, such as graduate students; a second experienced opinion is missing in such cases. Peer review can provide that second opinion, however the unbalanced exchanges between reviewer and authors (with the anonymous reviewer having the upper hand and the authors sometimes forced to yield in order to publish) is not an ideal forum for debate.

We hesitate to advocate more extreme forms of noise-reduction that automate species delimitation by taking it out of the hands of human decision-makers. These can go beyond methods that process a large set of DNA sequences to identify likely species (e.g., Puillandre et al., 2012, 2021; Fujisawa and Barraclough, 2013; Carstens et al., 2013; Sukumaran et al., 2021). Indeed, they include algorithms using Bayesian inference that integrate genetic and phenotypic data to delimit species (e.g., Solis-Lemus et al., 2015) or approaches based on deep-learning algorithms and artificial intelligence (Wäldchen and Mäder, 2018; Karbstein et al., 2024). In principle, these are truly noise-free and bias-free, as long as they were designed to be free of bias by adhering to strict consensus rules established by a panel of experts, thus eliminating subjectivity and achieving repeatable decisions. Such approaches have been applied with some success to morphological data in taxa such as fungi where species determination is challenging (e.g., Zieliński et al., 2020; Bartlett et al., 2022). They have even been proposed for parasite taxonomy and applied to discriminate among capillarid nematode species based on the characteristics of their eggs (Borba et al., 2021). However, it may prove difficult to design algorithms for taxonomic decisions that will apply across all trematode (or any other group) taxa and be accepted by all experts. If deep-learning algorithms and AI are ever tailored to helminth taxonomic decisions, their main use might be for *post hoc* confirmation: if their output agrees with the expert's judgement, one can have greater confidence in the decision, but if not this may be a prompt for the expert to re-examine the evidence.

We therefore propose that a combination of the three noise reduction strategies above (establishment of clear guidelines, decomposition of the evidence, aggregation of independent judgements) would achieve greater consistency of species discrimination. To a large extent, distinguishing between closely-related species cannot be reduced to an inflexible algorithm; there will always be a need for expert judgment. Hard decision rules applied invariably across all cases are bound to produce errors, perhaps more than they are meant to avoid. Careful consideration of the particulars and intangibles of each case remains crucial; this is beyond what rigid decision rules and algorithms can achieve. Therefore, we feel that even the 'guidelines' suggested above should not be taken as hard rules, thus allowing experts to exert some discretion rather than merely follow a mechanical decision process.

The cost and consequences of eliminating all noise, i.e., eliminating all inconsistencies among taxonomists in decisions regarding species identification, may exceed the benefits. For example, if taxonomy moved to adopt the AI algorithm approach, by removing judgment from the hands of experts and delegating it to an automated process, the hard-earned expertise of taxonomists would be devalued. Even the three less extreme noise-reduction strategies suggested above will not appeal to everyone. We propose them here merely as possible steps toward decreasing the current level of judgment noise in parasite taxonomy, and not as ways to eliminate all noise. Human judgment and intuition must retain their place in species recognition, however setting them within a well-defined decision framework would limit the current high level of inter-individual differences of opinion regarding whether to lump similar specimens into one species or to split them into two species.

## CRediT authorship contribution statement

**Robert Poulin:** Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jerusha Bennett:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Bronwen Presswell:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization.

## Acknowledgements

We thank the many trematode taxonomists who have participated in our study and without whose time this study would not have been possible. We are grateful to two anonymous reviewers for constructive comments. J. Bennett was supported through the Cawthron Institute's Ministry of Business, Innovation and Employment, "Emerging Aquatic Diseases" Endeavour grant, award number CAWX2207.

## Data availability

All species discrimination data used in the present study are shown in Fig. 1. No further information can be published in order to protect the anonymity of respondents.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2025.06.003.

## References

Aiken, H.M., Bott, N.J., Mladineo, I., Montero, F.E., Nowak, B.F., Hayward, C.J., 2007. Molecular evidence for cosmopolitan distribution of platyhelminth parasites of tunas (*Thunnus* spp.). Fish Fisher. 8, 167–180.

Bartlett, P., Eberhardt, U., Schütz, N., Beker, H.J., 2022. Species determination using AI machine-learning algorithms: *Hebeloma* as a case study. IMA Fungus 13, 13.

Blasco-Costa, I., Cutmore, S.C., Miller, T.L., Nolan, M.J., 2016. Molecular approaches to trematode systematics: 'best practice' and implications for future study. Syst. Parasitol. 93, 295–306.

Borba, V.H., Martin, C., Machado-Silva, J.R., Xavier, S.C.C., de Mello, F.L., Iñiguez, A. M., 2021. Machine learning approach to support taxonomic species discrimination based on helminth collections data. Parasit. Vect. 14, 230.

Carlson, C.J., Dallas, T.A., Alexander, L.W., Phelan, A.L., Phillips, A.J., 2020. What would it take to describe the global diversity of parasites? Proc. R. Soc. B 287, 20201841.

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.

Conix, S., Cuypers, V., Zachos, F.E., De Block, A., 2023. Taxonomic disagreement about ranks in gray-area taxa: a vignette study. BioScience 73, 728–737.

Cribb, T.H., Bray, R.A., Justine, J.-L., Reimer, J., Sasal, P., Shirakashi, S., Cutmore, S.C., 2022. A world of taxonomic pain: cryptic species, inexplicable host-specificity, and host-induced morphological variation among species of *Bivesicula* Yamaguti, 1934 (Trematoda: Bivesiculidae) from indo-pacific holocentridae, muraenidae and serranidae. Parasitology 149, 831–853.

Cribb, T.H., Barton, D.P., Blair, D., Bott, N.J., Bray, R.A., Corner, R.D., Cutmore, S.C., De Silva, M.L.I., Duong, B., Faltýnková, A., Gonchar, A., Hechinger, R.F., Herrmann, K. K., Huston, D.C., Johnson, P.T.J., Kremnev, G., Kuchta, R., Louvard, C., Luus-Powell, W.J., Martin, S.B., Miller, T.L., Pérez-Ponce de León, G., Smit, N.J., Tkach, V.V., Truter, M., Waki, T., Vermaak, A., Wee, N.-Q.-X., Yong, R.-Q.-Y., Achatz, T.J., 2025. Challenges in the recognition of trematode species: consideration of hypotheses in an inexact context. J. Helminthol. 99, e54.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

Endersby, J., 2009. Lumpers and splitters: Darwin, Hooker, and the search for order. Science 326, 1496–1499.

Faurby, S., Eiserhardt, W.L., Svenning, J.-C., 2016. Strong effects of variation in taxonomic opinion on diversification analyses. Methods Ecol. Evol. 7, 4–13.

Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. Syst. Biol. 62, 707–724.

Hildebrand, J., Adamczyk, M., Laskowski, Z., Zalesny, G., 2015. Host-dependent morphology of *Isthmiophora melis* (Schrank, 1788) Lühe, 1909 (Digenea, Echinostomatinae): morphological variation versus molecular stability. Parasit. Vect. 8, 481.

Kahneman, D., Sibony, O., Sunstein, C.S., 2021. Noise: a flaw in human judgment. Harper Collins Publishers, London.

Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., Wagner, N.D., Emerson, B.C., Albach, D.C., Scheu, S., Bradler, S., de Vries, J., Irisarri, I., Li, H., Soltis, P., Mäder, P., Wäldchen, J., 2024. Species delimitation 4.0: integrative taxonomy meets artificial intelligence. Trends Ecol. Evol. 39, 771–784.

Kunz, W., 2002. When is a parasite species a species? Trends Parasitol. 18, 121–124.

Moss, W.W., 1971. Taxonomic repeatability: an experimental approach. Syst. Zool. 20, 309–330.

Pérez-Ponce de León, G., 1995. Host-induced morphological variability in adult *Posthodiplostomum minimum* (Digenea: Neodiplostomidae). J. Parasitol. 81, 818–820.

Pérez-Ponce de León, G., Garcia-Varela, M., Pinacho-Pinacho, C.D., Sereno-Uribe, A. L., Poulin, R., 2016. Species delimitation in trematodes using DNA sequences: middle-American *Clinostomum* as a case study. Parasitology 143, 1773–1789.

Pérez-Ponce de León, G., Poulin, R., 2018. An updated look at the uneven distribution of cryptic diversity among parasitic helminths. J. Helminthol. 92, 197–202.

Poulin, R., 2011. Uneven distribution of cryptic diversity among higher taxa of parasitic worms. Biol. Lett. 7, 241–244.

Poulin, R., Presswell, B., 2016. Taxonomic quality of species descriptions varies over time and with the number of authors, but unevenly among parasitic taxa. Syst. Biol. 65, 1107–1116.

Poulin, R., Presswell, B., 2022. Is parasite taxonomy really in trouble? A quantitative analysis. Int. J. Parasitol. 52, 469–474.

Poulin, R., Presswell, B., 2024. Nomenclatural stability and the longevity of helminth species names. Syst. Parasitol. 101, 34.

Presswell, B., Bennett, J., 2019. *Galactosomum otepotiense* n. sp. (Trematoda: Heterophyidae) infecting four different species of fish-eating birds in New Zealand: genetically identical but morphologically variable. J. Helminthol. 94, e86.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, automatic barcode gap discovery for primary species delimitation. Mol. Ecol. 21, 1864–1877.

Puillandre, N., Brouillet, S., Achaz, G., 2021. ASAP: assemble species by automatic partitioning. Mol. Ecol. Res. 21, 609–620.

Simpson, G.G., 1945. The principles of classification and a classification of mammals. Bull. Am. Mus. Nat. Hist. 85, 1–350.

Solis-Lemus, C., Knowles, L.L., Ané, C., 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69, 492–507.

Sukumaran, J., Holder, M.T., Knowles, L.L., 2021. Incorporating the speciation process into species delimitation. PLoS Comput. Biol. 17, e1008924.

Tibayrenc, M., 2006. The species concept in parasites and other pathogens: a pragmatic approach? Trends Parasitol. 22, 66–70.

Tobias, J.A., Seddon, N., Spottiswoode, C.N., Pilgrim, J.D., Fishpool, L.D.C., Collar, N.J., 2010. Quantitative criteria for species delimitation. Ibis 152, 724–746.

Wäldchen, J., Mäder, P., 2018. Machine learning for image-based species identification. Meth. Ecol. Evol. 9, 2216–2225.

Wee, N.-Q.-X., Cribb, T.H., Shirakashi, S., Cutmore, S.C., 2022. Three new species of *Helicometroides* Yamaguti, 1934 from Japan and Australia, with new molecular evidence of a widespread species. Parasitology 149, 622–639.

Zieliński, B., Sroka-Oleksiak, A., Rymarczyk, D., Piekarczyk, A., Brzychczy-Włoch, M., 2020. Deep learning approach to describe and classify fungi microscopic images. PLoS ONE 15, e0234806.