Biodiversity Journal, 2024,15 (2): 129–133 https://doi.org/10.31396/Biodiv.Jour.2024.15.2.129.133

Frequent misclassification by mtDNA barcoding as revealed by nuDNA and/or testable analysis of its expression products

Bernhard Seifert

Senckenberg Museum of Natural History Görlitz, D-02826 Görlitz, Am Museum 1, Germany; e-mail: bernhard.seifert@senkenberg.de

ABSTRACT Species identities are best indicated by analyses of nuclear DNA which is the material representing the working points of evolution. Additional good indicators of species identities are those expression products of nuclear DNA which are least modified by environmental influence and, as a consequence, show the highest correlation with the genetical core information. Among expression products such as morphology, products of metabolism and ethological or ecological traits, morphology is rated here as indicator with the highest correlation. The use of morphology as most important accessory indicator is furthermore favored by the leading position it played in species descriptions over 280 years of taxonomic research. Focusing on the example of ants, the paper considers 13 studies with parallel application of mtDNA barcoding, analysis of nuclear DNA and application of Numeric Morphology-Based Alpha-Taxonomy (NUMOBAT). Selected were only studies based on sufficiently high within-species numbers of samples. With nuclear DNA and NUMOBAT used as objective and testable control systems, the average classification error of mtDNA barcoding per sample or individual was 16.8% over 10 genera with 66 species with the extremes ranging from 0 to 32%. Ancient hybridization is considered a much more likely cause for mtDNA mismatches in ants than incomplete lineage sorting.

KEY WORDS Species classification; ants; mtDNA barcoding; nuclear DNA; morphology-based numeric alpha-taxonomy.

Received 12.02.2024; accepted 05.04.2024; published online 18.05.2024

INTRODUCTION

The year 2003 was the starting point for a determined campaign that was to have a strong and long-lasting impact on alpha-taxonomy. Hebert et al. (2003) presented the term "DNA barcoding" and praised it as a ready-to-use silver bullet for reliable species identification. They claimed that all the 10–25 million animal species on earth could be quickly recognized by large-scale screening of a mitochondrial DNA reference gene with comparably low costs. This idea of a turbo-taxonomic approach received an enormous echo ranging from top-ranking science journals such as Nature (Blaxter, 2003) to popular media such as The Times (Henderson, 2005). Massive counter-evidence for disagreement of mtDNA classifications with other indicators of species identity was presented by a meta-analysis of 584 studies of 526 eumetazoan genera already in the same year by Funk & Omland (2003) who detected mtDNA paraphyly or polyphyly in 23% of 2319 assayed species. This and numerous follow-up papers revealing mismatches between species

identities and mtDNA indication in the years since then could not stop the development of the global Consortium for the Barcoding of Life (CBOL) and its some 50 national offshoots. The wide application of mtDNA barcoding continued even after Ross (2014) presented another similarly broad meta-analysis confirming the conclusions of Funk & Omland (2003).

Caused by much improvement in analysis of nuclear DNA (nuDNA), both regarding methodology and costs, we currently observe an avertion of several biodiversity students from mtDNA barcoding but the general tenacity to adhere to this method remains astonishing. A frequent answer in personal talks with convinced barcoders was that, though not denying occasional occurrence of paraphylies in many species, they deemed the frequency of misclassification on individual level to be negligibly low.

Yet, are the frequencies really so low? Figuring out the true performance of Hebert's barcoding requires to fulfill three preconditions. Precondition 1 is that the alternative methods checking the classification by mtDNA are testable. Testability, or falsification and verification of classification hypotheses is only possible when they are based, in one or the other way, on numeric analyses but not on subjective idiosyncratic decision. Finding real frequencies furthermore requires sufficiently high within-species numbers of samples (precondition 2) and not juxtaposition of alternative classification systems in singletons as seen in the trees frequently published. Most important, or essential, for realistic assessment of barcoding performance is the that the controlling classification systems have the highest likelihood to indicate "true" species identities (precondition 3). In a paper introducing the Gene and Gene Expression (GAGE) species concept, Seifert (2020) wrote "*...Species are separable clusters that have passed a threshold of evolutionary divergence and are exclusively defined by nuclear DNA sequences and / or their expression products. Nuclear DNA sequences and their expression products are different character systems but have a highly correlated indicative function. Character systems with the least risk of epigenetic or ontogenetic modification have superior indicative value when conflicts between character systems of integrative studies arise...*" In other words, "true" species identities are best indicated when classification methods focus on the working points of evolution and this is nuDNA and those expression products of nuDNA least modified by environmental influence and thus most strongly correlating with nuDNA. These correlations are highest in protein sequences and morphology, considerable in behavioral traits and products of metabolism, and weak in ecological traits (Schlick-Steiner et al., 2010; Seifert, 2018). Another argument to favor morphology among the expression products is the leading role it played in the history of taxonomy since Linnaeus.

Condition 1 (testability) and condition 3 (indication close to "true" species identities) are best fulfilled by an approach or working philosophy named by Seifert (2009) Numeric Morphology-Based-Alpha-Taxonomy (NUMOBAT). Using the example of ants (Hymenoptera: Formicidae), I present here the results of 13 studies being in agreement with the three preconditions outlined above. Such hard checks, either nuDNA- or NUMOBAT-based (or both in combination) are still very rare and I have the impression that the advocates of mtDNA barcoding mentally displace such studies. Table 1 lists up the percentage of misclassifications by mtDNA barcoding per nest sample or individual.

Averaging these data, the result is sobering: the mean estimated misclassification per individual or nest sample within 13 studies over 10 genera and 66 species is 16.8%. Wrong indications of Hebert's barcoding in the same range are supposed by subjective assessment of morphology for the genera *Anochoetus* and *Odontomachus* (Fisher et al., 2008) and *Solenopsis* (Shoemaker et al., 2006). Furthermore there is introgression of mtDNA into nuDNA-defined lineages in socially hybridogenetic ants (e.g. Daras & Aron, 2015). In a very broad study in South Finland, Beresford et al. (2017) showed massive bidirectional introgression of heterospecific mtDNA into the populations of *Formica polyctena* and *aquilonia*.

A closer look at the data in Table 1 reveals a clear trend that the lowest classification errors by mtDNA barcoding occur in species with parapatric zoogeography (the *Plagiolepis* and *Temnothorax* cases) or reduced frequency of outcrossing due to high frequency of intranidal mating (the *Tapinoma nigerrimum* and *Cardiocondyla nuda* group examples). In contrast, higher errors are more frequent in species groups performing extranidal mating or normal nuptial flights and having at least partially

Misclassification	Species	Checking system	Reference
0%	*Plagiolepis taurica* and *P. pyrenaica*	nuDNA, NUMOBAT	Kirchner et al., 2023
1%	*Temnothorax nylanderi* and *T. crassispinus*	NUMOBAT, Geogr.	Pusch et al., 2006
6%	4 species of the *Tapinoma nigerrimum* group	NUMOBAT	Seifert et al., 2017a
7%	5 species of the *Cardiocondyla nuda* group	NUMOBAT	Seifert et al., 2017b
15%	*Formica pratensis* and *F. lugubris*	NUMOBAT	Seifert & Goropashnaya, 2004
16%	*Cardiocondyla latifrons* and *C. micropila*	NUMOBAT	Heinze pers. comm
17%	10 species of *Tetramorium*	NUMOBAT, nuDNA	Wagner et al., 2017
19%	17 species of Neotropical *Linepithema*	nuDNA	Wild, 2008
20%	8 species of *Serviformica*	nuDNA	Purcell pers. com
21%	6 species of the *Cataglyphis albicans* group	nuDNA	Eyer & Hefetz, 2018
23%	3 species of North African *Cataglyphis*	nuDNA, NUMOBAT	Knaden et al., 2005
24%	3 species of Tibetan *Myrmica*	nuDNA, NUMOBAT	Seifert et al., 2018
32%	2 species of *Colobopsis*	NUMOBAT, Geogr.	Schifani et al., 2021

Table 1. Percentage of misclassifications by mtDNA barcoding as revealed by checking systems which are either data of nuclear DNA, of Numeric Morphology-Based Alpha-Taxonomy (NUMOBAT) or of both. Note that zoogeography (Geogr.) was added as indicator in a case of parapatric distribution. For species names explicitly given in the table, the full names are *Plagiolepis taurica* Santschi, 1920, *Plagiolepis pyrenaica* Emery, 1921, *Temnothorax nylanderi* (Foerster, 1850), *Temnothorax crassispinus* (Karavajev, 1926), *Tapinoma nigerrimum* (Nylander, 1856), *Formica pratensis* Retzius, 1783, *Formica lugubris* Zetterstedt, 1838, *Cardiocondyla latifrons* Seifert, 2023, *Cardiocondyla micropila* Seifert, 2023 and *Cataglyphis albicans* (Roger, 1859).

sympatric geographic ranges. These data support the idea that ancient hybridization with subsequent introgression of misleading matrilines is the most frequent source for mtDNA barcoding errors in ants whereas incomplete lineage sorting during species splitting is rarer and, in the cases reported here, probably responsible for the situation in the parapatric *Colobopsis* species.

It has to be noted that the cases reported in Table 1, with mean error rates of 17%, refer to primary studies done by true experts of the species groups in question and that these researchers did not rely on data deposited in genetic online data banks such as GenBank or BOLD. A much worse situation may emerge in investigations in which three main sources of error come together and multiply to inflated figures. These error sources are (A) once again, the mtDNA mismatches caused by natural (evolutionary) reasons reported above, (B) careless processing of DNA samples by commercial companies and (C) the automated (=unscrupulous) comparison by these companies of their sequencing data with unreliable gene bank data that involve many species classifications not done by experts. Most illustrative is a study recently published by five experienced Thuringian entomologists (Förster et al., 2023). They collected insects of 579 species, identified them morphologically and combined these in two identical samples. The samples were evaluated by two commercial companies using Metabarcoding

(see Taberlet et al., 2012) – a methodology accepting the error sources A, B, and C. Förster et al. then compared the classifications of the companies with their own morphological determinations and found a mismatch of 65% in company 1 and of 44% in company 2. Even if supposing that an experienced team of five entomologists might have made some misidentifications, these figures show that COI-based Metabarcoding produces disastrously wrong species classifications. For which purpose may Metabarcoding then be of use? It may be applied only for a rough overall quantification – i.e. if a habitat A harbors more "species" or "biodiversity" than a habitat B.

REFERENCES

Beresford J., Elias M., Pluckrose L., Sundström L., Butlin R.K., Pamilo P. & Kulmuni J., 2017. Widespread hybridization within mound-building wood ants in Southern Finland results in cytonuclear mismatches and potential for sex-specific hybrid breakdown. Molecular Ecology, 26: 4013–4026. https://doi.org/10.1111/mec.14183

Blaxter M., 2003. Counting angels with DNA. Nature 421: 122–124.

Darras H. & Aron S., 2015: Introgression of mitochondrial DNA among lineages in a hybridogenetic ant. Biological Letters, 11: 20140971. http://dx.doi.org/10.1098/rsbl.2014.0971

Eyer P.A. & Hefetz A., 2018. Cytonuclear incongruences hamper species delimitation in the socially polymorphic desert ants of the *Cataglyphis albicans* group in Israel. Journal of Evolutionary Biology. https://doi.org/10.1111/jeb.13378

Förster T., Creutzburg F., Anton E., Weigel A. & Hartmann M., 2023. Metabarcoding versus morphologische Identifizierung: der Herausforderung gewachsen? Entomologische Zeitschrift, 133/2: 103–116.

Fisher B.L. & Smith M. A., 2008. A revision of Malagasy species of *Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae). Public Library of Science One 3(5), e1787.

Funk D.J. & Omland K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology and Evolutionary Systematics, 34: 397–423.

Hebert P.D.N., Cywinska A., Ball S.L. & deWaard J.R., 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B, 270: 313–321.

Hendersen M., 2005. Is it a bird? Is it a plant? Hang on, I'll check its barcode. The Times, February 10: 14.

Kirchner P., Seifert B., Kröll J., The STEPPE Consortium, Schlick-Steiner B.C. & Steiner F.M., 2023. Phylogenomic inference and demographic model selection suggest peripatric separation of the cryptic steppe ant species *P. taurica* and *P. pyrenaica* stat. rev. Molecular Ecology, 32: 1149–1168. https://doi.org/10.1111/mec.16828

Knaden M., Tinaut A., Cerdá X., Wehner S. & Wehner R., 2005. Phylogeny of three parapatric species of desert ants, *Cataglyphis bicolor*, *C. viaticus*, and *C. savignyi*: a comparison of mitochondrial DNA, nuclear DNA, and morphometric data. Zoology, 108: 169–177. https://doi.org/10.1016/j.zool.2005.03.005

Pusch K., Seifert B., Foitzik S. & Heinze J., 2006. Distribution and genetic divergence of two parapatric sibling ant species in Central Europe. Biological Journal of the Linnean Society, 88: 223–234.

Ross H.A., 2014. The incidence of species-level paraphyly in animals: a re-assessment. Molecular Phylogenetics and Evolution, 76: 10–17. https://doi.org/10.1016/j.ympev.2014.02.021

Schifani E., Gianetti D., Czösz S., Castellucci F., Luchetti A., Castracani C., Spotti F.A., Mori A. & Grasso D.A., 2021. Is mimicry a diversification-driver in ants? Biogeography, ecology, ethology, genetics and morphology define a second West-Palaearctic *Colobopsis* species (Hymenoptera: Formicidae). Zoological Journal of the Linnean Society, 20: 1–27.

Schlick-Steiner B.C., Steiner F.M., Seifert B., Stauffer C., Christian E. & Crozier R.H., 2010. Integrative Taxonomy: a multi-source approach to exploring biodiversity. Annual Review of Entomology, 55: 421–438.

Seifert B. & Goropashnaya A., 2004. Ideal phenotypes and mismatching haplotypes errors of mtDNA treeing in ants (Hymenoptera: Formicidae) detected by standardized morphometry. Organisms, Diversity & Evolution, 4/4: 295–305. https://doi.org/10.1016/j.ode.2004.04.005

Seifert B., D'Eustacchio D., Kaufmann B., Centorame M., Lorite P. & Modica M.V., 2017a. Four species within the supercolonial ants of the *Tapinoma nigerrimum* complex revealed by integrative taxonomy (Hymenoptera: Formicidae). Myrmecological News, 24: 123–124.

Seifert B., Okita I. & Heinze J., 2017b. A taxonomic revision of the *Cardiocondyla nuda* group (Hymenoptera: Formicidae). Zootaxa, 4290 (2): 324–356. https://doi.org/10.11646/zootaxa.4290.2.4

Seifert B., Schultz R., Ritz M.S. & Ritz C.M., 2018. Cryptic species of the *Myrmica tibetana* complex (Hymenoptera: Formicidae) revealed by integrative taxonomy. Myrmecological News, 27: 1–18.

Seifert B., 2020. The Gene and Gene Expression (GAGE) Species Concept: An Universal Approach for All Eukaryotic Organisms. Systematic Biology, 69: 1033–1038. http://dx.doi.org/10.1093/sysbio/syaa032

Shoemaker D., Ahrens M.E. & Ross K.G., 2006. Molecular phylogeny of fire ants of the *Solenopsis saevissima* species-group based on mtDNA sequences. Molecular Phylogenetics and Evolution, 38: 200–215. https://doi.org/10.1016/j.ympev.2005.07.014

Taberlet P., Coissac E., Pompanon F., Brochmann C. & Willerslev E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21: 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Wagner H.C., Arthofer W., Seifert B., Muster C., Steiner F.M. & Schlick-Steiner B.C., 2017. Light at the end of the tunnel: Integrative taxonomy delimits cryptic species in the *Tetramorium caespitum* complex (Hymenoptera: Formicidae). Myrmecological News, 25: 95–129.

Wild A.L., 2008. Evolution of the Neotropical ant genus *Linepithema*. Systematic Entomology, 34: 49–62. http://dx.doi.org/10.1111/j. 1365-3113. 2008. 00435. x