

**PREPRINT**

*Author-formatted, not peer-reviewed document posted on 01/06/2026*

DOI: <https://doi.org/10.3897/arphapreprints.e201962>

---

# **Short-read genome skimming enables molecular barcoding of old myxomycete collections**

 Dmytro Leontyev, Martin Schnittler,  Oleg Shchepin

# Short-read genome skimming enables molecular barcoding of old myxomycete collections

Dmytro V. Leontyev<sup>1,2\*</sup>, Martin Schnittler<sup>2</sup>, Oleg N. Shchepin<sup>2</sup>

<sup>1</sup>H.S. Skovoroda Kharkiv National Pedagogical University, Alchevskikh 29, Kharkiv 61002 Ukraine

<sup>2</sup>Institute of Botany and Landscape Ecology, University Greifswald, Soldmannstr. 15, Greifswald D-17487 Germany

\*Corresponding author: [alwisiamorula@gmail.com](mailto:alwisiamorula@gmail.com)

## Abstract

This study evaluates the effectiveness of Illumina-based genome skimming for barcoding myxomycete herbarium collections ranging from 29 to 91 years in age. Despite significant DNA fragmentation, we successfully recovered partial sequences of the standard barcode marker (nucSSU) in all cases. For some collections, additional markers were recovered, including EF1a, COI, and mtSSU, contributing to a total of 28 recognized genes. In a 33-year-old specimen of *Lindbladia tubulina*, the assembly reached an N50 of 4.19 kb, enabling the recovery of extended functional loci. It was found that within the 29–91-year timeframe, specimen age does not strictly correlate with genomic informativeness, thus allowing us to conclude that even older collections can be successfully barcoded. The input genomic DNA quantity emerges as the primary determinant of sequencing success. Samples with high DNA yields provide representative amounts of contigs coming certainly (matching sequences in the NCBI nucleotide database) or presumably (no matches) from myxomycetes, regardless of specimen age. The taxonomic interpretation of the non-target fraction of contigs revealed distinct signals of both anthropogenic contamination (human DNA and skin microflora) and natural substrate inhabitants, including oribatid mites and bacteria from dead wood, soil, and grass litter. The results demonstrate that short-read genome skimming can provide a bridge to integrate historical type material of myxomycetes into contemporary phylogenetic research. This method can effectively overcome the length-dependent limitations of traditional Sanger sequencing, thus providing a roadmap for the future of museomics in myxomycetology.

**Keywords:** Amoebozoa; old DNA; bioinformatics; DNA degradation; herbarium genomics; metagenomics; next-generation sequencing; nucSSU.

## Introduction

Nomenclatural types serve as the foundation of biological systematics: every scientific name is ultimately anchored to a specific physical specimen preserved in a collection (Turland et al. 2018). Many discussions regarding the diagnostic features of a species, its phylogenetic relationships, circumscription, and boundaries can potentially be resolved by reference to the type. The splitting of a single polymorphic species into several new ones must as well rely on the type specimen; otherwise, it remains unclear which of the new species should retain the historical name (Singh 2019; Winston 1999). The loss or destruction of a type creates obstacles for any taxonomic revision, engendering nomenclatural uncertainty and forcing authors to interpret names arbitrarily.

As taxonomy has entered the molecular era, DNA barcoding has become a major tool for species identification and delimitation. However, alongside new opportunities, this methodology has revealed a serious limitation: type specimens collected in the distant past cannot be analysed using the same routine methods as fresh material. Specifically, for myxomycetes – a group of terrestrial protists with collectible fungus-like fructifications (Stephenson et al. 2008), the conventional threshold for specimen age is considered to be 10–15 (20) years (Leontyev et al. 2015, 2026; Janik et al. 2020; Schnittler et al. 2020, 2017). For herbarium collections exceeding this age, barcoding via standard Sanger sequencing mostly fails. This problem creates a fundamental gap between historical names, linked to non-barcoded types, and molecular data from contemporary collections. Therefore comparability is limited to morphological characters, which is often difficult due to the paucity or poor preservation of old collections (Leontyev et al. 2019a; Ronikier et al. 2022; Prikhodko et al. 2023). As a result, the large datasets of thoroughly studied contemporary specimens cannot be accurately matched with type material and, consequently, not identified correctly. For myxomycetes, this represents a major challenge: approximately 85% of known species in this group were described before the year 2000 (Schnittler et al. 2025), meaning their type specimens likely cannot be barcoded by standard methods.

However, advances in paleogenetics have demonstrated that DNA is a relatively stable molecule, capable of persisting for hundreds of thousands of years (Pääbo 1989; Kjær et al. 2022). In this context, the rapid loss of molecular accessibility in myxomycete specimens appears paradoxical. The explanation lies in the fact that the obstacle to barcoding old collections is not the total destruction of DNA, but its fragmentation following cell death (Brewer et al. 2019; Burrell et al. 2015; Pääbo et al. 2004). If the genome is fragmented into segments shorter than the target marker locus, its PCR amplification coupled with classical Sanger sequencing becomes technically impossible, even if the specimen still contains a significant amount of DNA. Based on this assumption, many authors have already turned to next-generation sequencing methods, specifically the Illumina platform, which produces millions to billions of short nucleotide sequence reads. The results have been impressive: it has been shown that even a specimen age of 150–200 years is not a limit for successful molecular barcoding of plants and fungi stored in herbaria (Bakker 2017; Shumskaya et al. 2023). Herbarium collections have proven to be a “repository of dead DNA”, a vast archive of genomic data that can be used for taxonomic, population, evolutionary, and ecological studies (Burrell et al. 2015; Bieker and Martin 2018). Consequently, the concept of *museomics* has emerged, encompassing the analysis of nucleic acid and protein molecules in museum collections (Guschanski et al. 2013; Raxworthy and Smith 2021).

Probably the most relevant and cost-effective approach for working with old specimens is genome skimming. This involves low-coverage random sequencing of the total DNA available in the material without attempting a full reconstruction of the nuclear genome. Its primary purpose is “skimming off” genes from high-copy regions, such as ribosomal and mitochondrial markers, from the total gene pool. Due to their multi-copy nature, these sequences are represented in libraries significantly more frequently than single-copy nuclear genes, allowing for reliable molecular barcoding even with extremely low DNA yields (Straub et al. 2012; Zeng et al. 2018).

In the present study, we attempted to evaluate the feasibility of barcoding myxomycete herbarium collections ranging from 29 to 91 years in age. We aimed to test the effectiveness of the genome skimming method for obtaining reliable genetic data from specimens with varying degrees of preservation and diverse DNA concentrations.

## Methods

*Material selection.* The study material consisted of seven myxomycete specimens from various taxonomic groups, collected by R. Hagelstein, T.E. Brooks, I. Ammirati, R. McHugh, S.L. Stephenson, Y.K. Novozhilov, and M. Schnittler (see Table 1) and currently housed in the herbarium of the Institute of Botany and Landscape Ecology at the University of Greifswald, Germany (Supplementary data 1, Sheet A). The specimens were selected to represent an age gradient, with collection dates ranging from 1935 to 1997. Initially, eight specimens were selected; however, one was found to be completely devoid of DNA for unclear reasons and was excluded from the study. Our analysis of this particular specimen is detailed in a separate publication (Leontyev and Schnittler 2025). The selected specimens differed significantly in size and number of fruiting bodies, but all contained friable, uniformly coloured spore masses and lacked visible contamination by fungi or arthropods. Around 1 mm<sup>3</sup> of spores were sampled from each specimen; the total spore count was not normalized.

*DNA library preparation and sequencing.* Following the recommendations of Shumskaya et al. (2023), we extracted DNA from the specimens using the filter-based QIAamp DNA Micro Kit (Qiagen, Germany), following the “Isolation of Genomic DNA from Tissues” protocol supplied by the manufacturer. A key feature of this kit is the presence of carrier RNA, i.e. polyA molecules that non-specifically bind short DNA fragments, preventing them from passing through the filter pores.

For each sample 50  $\mu\text{L}$  of DNA solution were obtained. The concentration of the resulting product was measured using a DeNovix DS-11 FX fluorometer (DeNovix Inc. USA) following the TN 143 Broad Range Assay Detailed Protocol (DeNovix Inc. USA). After concentration measurement, the samples were divided in two groups. The first group, where the extracted DNA amount reached 150 ng (sc11426 and sc2833), was directly used for sequencing. For the second group (the rest of the samples, with the lower amounts of DNA), a 2.5  $\mu\text{L}$  aliquot was taken for whole-genome amplification (WGA) using the 4BB TruePrime WGA Kit (4basebio, UK). The amplification process was conducted for 6 hours. Then, 35  $\mu\text{L}$  of genomic DNA solution (first group) or 72  $\mu\text{L}$  of WGA products (second group) with final amount of DNA ranging from 158 to 658 ng (see Table 1), were sent to Eurofins Genomics (Germany) for genome skimming on the Illumina NovaSeq (USA) platform, using the WGS Bacteria/Yeast/Fungi protocol with a sequencing depth of 5 million read pairs (2x 150 bp) per sample.

**Bioinformatic processing.** The demultiplexed, adapter-trimmed paired-end Illumina sequencing data were screened for quality and presence of remaining adapter sequences using FastQC v0.11.9 (Andrews 2019). To obtain a high sensitivity of detection of low-abundance genomic signal expected from old DNA samples, the reads were assembled de novo using a metagenomic assembler metaSPAdes v4.0.0 (Nurk 2017), which is equivalent to running SPAdes with the “--meta” flag. No prior quality-based read trimming or filtering was done before the assembly. The assemblies were run on the High-Performance Computing cluster of the University of Greifswald. For the sample *Lindbladia tubulina* sc2833, which had the highest input DNA amount and the largest sequencing data volume, this assembly method failed with an out-of-memory error. Therefore the assembly for this sample was made using the standard assembly mode of SPAdes v4.0.0 (Prjibelski et al. 2020).

For taxonomic profiling of contigs, each resulting assembly was searched against a locally installed NCBI nucleotide database (“core\_nt”, updated 31 Jan 2026) using BLASTn v. 2.5.0+ in megablast mode (Camacho et al. 2009). BLAST output was generated in tabular format (“outfmt 6”) including subject TaxIDs, and these TaxIDs were linked to the NCBI taxonomy using the NCBI “taxdump” files.

To estimate per-contig sequencing depth, the original reads of each sample were mapped back to the corresponding assembly using strobealign v. 0.17.0 (Sahlin 2022). Alignment output (SAM, Sequence Alignment Map) was converted to sorted and indexed BAM (Binary Alignment Map) files using samtools v. 1.12 (Danecek et al. 2021). Taxonomic assignments, GC content, contig length, and read depth were integrated and visualized with BlobToolKit v. 4.5.0 (Challis et al. 2020). For each sample, BlobTools2 was used to create a project from the assembly FASTA, BLAST hits, and BAM coverage file, and summary snail and blob plots were produced to assess the distribution of contigs and read support across inferred taxonomic groups at different taxonomic levels.

For every contig were the BLAST search over a local core\_nt database found a match to an organism belonging to Eumycetozoa, we additionally conducted a thorough online BLAST search over the NCBI GenBank Nucleotide database (see Supplementary file 1, Sheet B for a list of the best 10 matches with the smallest e-values and Sheet C for a list of genes found).

## Results

**Identification of the Target Genome Fragments.** For all seven samples, we obtained a substantial volume of raw genetic data. The total sequence length was relatively uniform, ranging from 1134 to 2354 Mbp (Table 1). However, the total length of the assembled contigs varied by two orders of magnitude, ranging from 1.7 Mbp in *L. epidendrum* sc25089 to 162.0 Mbp in *T. scabra* sc11426.

Compared to the genome size of *Badhamia polycephala* (Miller et al. 2021), this corresponds to an estimated genome coverage between 0.7% and 65%.

Visualization of the assemblies via blob plots (Supplementary file 4) demonstrates significant differences in the data structure between samples. In all samples, a varying number of contigs (1–19) were unambiguously classified as belonging to Eumycetozoa based on BLAST hits (Fig. 1, Tables 1, 2). However, in all specimens, most of the resulting assemblies belonged to the no-hit fraction—contigs with no matches in the NCBI database. We hypothesize (see Discussion) that the no-hit fraction is largely composed of myxomycete genetic material, which remains unidentified because genomic data for myxomycetes are still extremely sparse (Kang et al. 2017; Schaap et al. 2016; Takano et al. 2001). In the sample *Lindbladia tubulina* sc2833 (Fig. 1A, C), the no-hit fraction accounted for 82.3 Mbp and had a unimodal GC content distribution with highest contig density within a 35% to 45% range and a predominant base coverage of 20–40×, matching these of the contigs identified as fragments of myxomycete nuclear genome. In *Lycogala epidendrum* UARK25089 (Supplementary file 4), the no-hit fraction was smaller (1.4 Mbp) but maintained a characteristic distribution profile, allowing its identification as a residual signal of the nuclear genome. In *Trichia scabra* sc11426 (Fig. 1B, D), the no-hit contig fraction had a bimodal GC content distribution, with the second peak at around 68%, suggesting heavy contamination by prokaryotic DNA. In samples with low input DNA amount and high fragmentation, such as *Cribraria mirabilis* sc10428 and *Fuligo cinerea* sc11445 (Supplementary file 4), the data were represented by a sparse cloud of short contigs with a broad unimodal GC content distribution and the base coverage peaking around one.

In the combined no-hit & Eumycetozoa fraction of all samples except *L. tubulina* sc2833, the N50 values vary between 207–266 bp (Table 1, Fig. 2) and the statistical mode of read length distribution varies between 56 and 237 bp. This reflects the dominance of short unresolved fragments in the assembly graph that likely resulted from dead ends, short sequence repeats, rare variants, abundant contaminants, low-coverage sequences or sequencing errors. These short and low-coverage contigs are not filtered out from the results by metaSPAdes. The sample *L. tubulina* sc2833 stands out, with the N50 reaching 4.19 Kbp, which is likely explained by the higher input DNA amount and the different assembly mode applied to it (standard SPAdes mode – see Methods).

Specimen age shows no significant influence on genomic data quality. The two specimens with the highest DNA yields (*L. tubulina* sc2833 and *T. scabra* sc11426) differ substantially in age (33 and 64 years, respectively) and are not the youngest among the studied collections (Table 2). For the specimens treated with WGA, no correlation was found between specimen age and DNA amount, read length, or the number of Eumycetozoa contigs (Fig. 3). The relationship between input DNA amount and assembly statistics remains moderate; correlation coefficients are positive for all tested parameters but largely lack statistical significance.

**DNA Barcodes in Genomic Assemblies.** Although the detection of Eumycetozoa within the resulting data remains challenging due to the lack of reference genomes, we succeeded in identifying anchor contigs for all seven specimens that correspond to their preliminary morphological identification (Supplementary file 2). Despite the paucity of these fragments, their presence is critical for specimen barcoding – the primary objective of genome skimming for old herbarium collections. It is important to emphasize that for most specimens the identified Eumycetozoa markers are localized within the main no-hit cloud (Fig. 1; Supplementary file 4). The Sheet B in the Supplementary file 1 lists the ten best BLAST hits for each contig and the detected genes.

Nuclear small subunit ribosomal gene (nucSSU) sequences were identified for all specimens. Their length varied from 157 to 7755 bp, and base coverage ranged from 1 to 127 copies (Table 2). For

six out of seven specimens, the detected fragments of nucSSU at least partially covered the ca. 600 bp fragment conventionally used for DNA barcoding in myxomycetes (Supplementary file 3). This marker is established as the conventional molecular barcode for myxomycetes (Leontyev and Schnittler 2017).

The total number of recognized genes reached 28 (Supplementary file 1, Sheet C). Among them, the eukaryotic elongation factor 1-alpha (EF1a) gene was detected in four specimens, with the contig length exceeding 10 kbp in *L. tubulina* sc2833. For three specimens, contigs containing the mitochondrial cytochrome c oxidase subunit I (COI) gene were obtained, varying in length (up to 41 kb) and base coverage (14–1506). Sequences including the mitochondrial small subunit ribosomal gene (mtSSU) were recovered for only two specimens, which also differed in length (up to 2310 bp) and base coverage (13–330). Other identified genes like actin, alpha-tubulin, and several tRNAs have not been studied for bright-spored myxomycetes; consequently, the matches found did not allow for species-level identification, as the closest analogues were dark-spored myxomycetes and dictyostelids (Supplementary file 1, Sheet B).

*Analysis of contamination and associated biota.* Taxonomic interpretation of non-target contigs revealed a complex metagenomic background, the composition of which varies depending on specimen age, substrate conditions, and archival history (Supplementary file 4). The anthropogenic fraction, primarily represented by human DNA, was recorded in all examined specimens. The proportion of this fraction correlates with the total DNA yield: in libraries with low target material content (*C. mirabilis* sc10428, *L. epidendrum* sc25089), anthropogenic noise becomes dominant. In close association with human DNA, a skin microflora complex was identified, including the bacterium *Cutibacterium acnes*, representatives of the genus *Staphylococcus*, and lipophilic yeasts of the genus *Malassezia*. Domestic cattle DNA (*Bos taurus*) found in several samples likely also belongs to the anthropogenic fraction.

The second metagenomic fraction is conditionally natural, reflecting the substrate inhabitants and ecosystem relationships of the myxomycetes. Arthropod DNA was detected selectively; the most pronounced signal was recorded in specimen *L. tubulina* sc12075, where oribatid mites of the genus *Oppiella*, common inhabitants of decaying wood, were identified (Maraun and Scheu 2000). The bacterial component of this fraction is represented by organisms characteristic of forest ecosystems: the cyanobacterium *Nostoc* and the betaproteobacterium *Variovorax* (associates of rotting wood) were noted in specimen *T. scabra* sc11426, while nitrogen-fixing soil bacteria of the genus *Bradyrhizobium* dominate in *L. tubulina* sc2833. In the oldest specimen of the series, *F. cinerea* sc11445, a predominance of putrefactive bacteria of the genus *Bacillus* was observed. The archaeal component is limited to nitrifying soil forms (*Nitrososphaera*). The viral fraction is represented by fragments of bacteriophages.

## Discussion

The obtained data demonstrate that genome skimming is an effective method for the molecular identification of myxomycete herbarium specimens, even when the material is of significant age. We successfully extracted at least one of the four markers we searched for (nucSSU, mtSSU, EF1a, COI) from all specimens, including the oldest collection, *Fuligo cinerea* (1935). For six out of the seven specimens, more than one marker was identified. The lack of a direct correlation between specimen age and the quality of the resulting data leads to an important conclusion: within the 30–90-year timeframe, the informativeness of the assembly does not critically deteriorate with age. This suggests that even a specimen age exceeding 100 years is not an obstacle to genome skimming, as previously demonstrated for vascular plants and fungi (Bieker and Martin 2018; Shumskaya et al. 2023). At the same time, the starting DNA amount is crucial for obtaining high-

quality data. Large specimens with a high quantity of spores yielded good results regardless of their age.

The dominance of the no-hit fraction in all assemblies is likely explained by the critical lack of genomic data for myxomycetes, particularly for the bright-spored group, Lucisporomycetidae (Leontyev et al. 2019b), represented in our study by all specimens except for *F. cinerea*. The correspondence of the GC content and coverage profile of the no-hit fraction with the identified anchor nucSSU contigs allows us to preliminarily consider this fraction as largely belonging to the nuclear genome of the target organism. The consistency of the GC content within the no-hit fractions provides additional support for this assumption: since GC content is a group-specific genomic feature (Hildebrand et al. 2010), the observed clustering suggests a single taxonomic source. However, this assumption requires further evidence. Moreover, sequences identified by BLAST as belonging to Eumycetozoa may represent unstudied genes of the target organism, but as well the contamination by other myxomycete species. This methodological problem can only be resolved by creating high-quality *de novo* reference genomes using Hi-Fi PacBio and Hi-C technologies for all key groups of myxomycetes. Without such a roadmap, the potential of genome skimming for archival collections will remain limited to a narrow set of standard barcodes.

A considerable challenge is the high contamination load, which creates noise within the genomic data. Notably, specimens subjected to WGA generally exhibited higher contamination levels and lower yield of target DNA. This issue may be a direct consequence of the WGA process, which preferentially amplifies younger and less-degraded DNA of anthropogenic contaminants rather than the target DNA. Unfortunately, our data are insufficient to fully evaluate the impact of this factor, as the two specimens that did not undergo WGA and yielded more target DNA initially possessed the highest DNA concentrations. Consequently, the suitability of WGA for genome skimming remains an open question.

Genome skimming has provided insights not only into the target species but also into the ecological context of the specimen. The detection of mite DNA, lignotrophic betaproteobacteria (*Variovorax*), and cyanobacteria (*Nostoc*) allows for a partial reconstruction of the decaying wood communities at the time of collection. The components of the soil microbiome, such as nitrogen-fixing bacteria (*Bradyrhizobium*) and nitrifying archaea (*Nitrososphaera*), may as well represent the same ecosystem. A particularly intriguing case is the predominance of *Bacillus* spp. in the specimen *F. cinerea* sc11445. This myxomycete typically develops on dead herbaceous substrates, such as rotting straw, where mass development of *Bacillus* spp. is frequently observed (Mandic-Mulec et al. 2015). Thus, even in a 90-year-old specimen, metagenomic data still carry information regarding the substrate upon which the myxomycete developed. Undoubtedly, a similar study of fresh material could provide a wealth of data on the topical and trophic relationships of myxomycetes, including the composition of the microorganisms on which their plasmodia feed.

Even in the absence of comprehensive, annotated reference genomes for all myxomycete taxonomic groups, the approach demonstrated here still offers an effective solution for resolving taxonomic uncertainties. For instance, in cases where a nominal species, such as *Reticularia splendens*, *Trichia varia* or *Tubifera dimorphotheca* (Feng and Schnittler 2015; Leontyev et al. 2025, 2026), comprises a complex of a few semi-cryptic biological species, genome skimming of modern representatives allows for the creation of a robust reference database. By sequencing one modern voucher for each recognized biospecies, we can establish a reference set for the barcoding of old collections. Even if a type specimen yields only highly fragmented, low-quantity DNA, the Illumina-based reference library will provide the necessary data to accurately assign the type to its corresponding lineage.

NGS is not the only alternative for specimen barcoding of old collections with highly fragmented DNA. An alternative approach could be the design of primers specifically for the amplification of

short barcode fragments that fall within the length range of old, degraded DNA. The authors have achieved success through this method and will report on it in a follow-up publication.

## Conclusion

The proposed method opens new prospects for the revision of herbarium collections and the construction of a molecular bridge between historical types and modern species concepts. The ability to extract genomic information from almost century-old specimens allows for the integration of myxomycete nomenclatural types, facilitating the entry of myxomycetological research into the era of phylogenomics.

## Acknowledgements

This project (1208119) has received funding through the MSCA4Ukraine project, which is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, the European Research Executive Agency or the MSCA4Ukraine Consortium. Neither the European Union nor the European Research Executive Agency, nor the MSCA4Ukraine Consortium as a whole nor any individual member institutions of the MSCA4Ukraine Consortium can be held responsible for them.

## References

- Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bakker FT (2017) Herbarium genomics: skimming and plastomics from archival specimens. *Webbia* 72: 35–45. <https://doi.org/10.1080/00837792.2017.1313383>
- Bieker VC, Martin MD (2018) Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters* 165: 409–418. <https://doi.org/10.1080/23818107.2018.1458651>
- Brewer GE, Clarkson JJ, Maurin O, et al. (2019) Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science* 10: 1102. <https://doi.org/10.3389/fpls.2019.01102>
- Burrell AS, Disotell TR, Bergey CM (2015) The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44. <https://doi.org/10.1016/j.jhevol.2014.10.015>
- Camacho C, Coulouris G, Avagyan V, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421. <https://doi.org/10.1186/1471-2105-10-421>
- Challis R, Richards E, Rajan J, et al. (2020) BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics* 10: 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- Danecek P, Bonfield JK, Liddle J, et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008. <https://doi.org/10.1093/gigascience/giab008>
- Feng Y, Schnittler M (2015) Sex or no sex? Group I introns and independent marker genes reveal the existence of three sexual but reproductively isolated biospecies in *Trichia varia* (Myxomycetes). *Organisms Diversity & Evolution* 15: 631–650. <https://doi.org/10.1007/s13127-015-0230-x>
- Guschanski K, Krause J, Sawyer S, et al. (2013) Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* 62: 539–554. <https://doi.org/10.1093/sysbio/syt018>
- Hildebrand F, Meyer A, Eyre-Walker A (2010) Evidence of selection upon genomic GC-content in Bacteria. *PLoS Genetics* 6: e1001107. <https://doi.org/10.1371/journal.pgen.1001107>

- Janik P, Ronikier M, Ronikier A (2020) New protocol for successful isolation and amplification of DNA from exiguous fractions of specimens: a tool to overcome the basic obstacle in molecular analyses of myxomycetes. PeerJ 8: e8406. <https://doi.org/10.7717/peerj.8406>
- Kang S, Tice AK, Spiegel FW, et al. (2017) Between a pod and a hard test: the deep evolution of Amoebae. Molecular Biology and Evolution 34: 2258–2270. <https://doi.org/10.1093/molbev/msx162>
- Kjær KH, Winther Pedersen M, De Sanctis B, et al. (2022) A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature 612: 283–291. <https://doi.org/10.1038/s41586-022-05453-y>
- Leontyev DV, Schnittler M (2017) The phylogeny of Myxomycetes. In: Stephenson SL, Rojas C (Eds) Myxomycetes: biology, systematics, biogeography and ecology. Academic Press, London, 83–106. <https://doi.org/10.1016/B978-0-12-805089-7.00003-2>
- Leontyev DV, Schnittler M (2025) Not a *Lycogala*? Investigating the holotype of *Lycogala fuscoviolaceum*. European Journal of Protistology 99: 126151. <https://doi.org/10.1016/j.ejop.2025.126151>
- Leontyev DV, Schnittler M, Stephenson SL (2015) A critical revision of the *Tubifera ferruginosa* complex. Mycologia 107: 959–985. <https://doi.org/10.3852/14-271>
- Leontyev DV, Lloyd SJ, Moroz EL, Inoue M, Schnittler M (2025) Cryptic speciation in *Reticularia*: two hardly distinguishable new species from Japan and Tasmania. Fungal Systematics and Evolution 17.
- Leontyev DV, Schnittler M, Stephenson SL, Novozhilov YK (2019a) Systematic revision of the *Tubifera casparyi*-*T. dictyoderma* complex: resurrection of the genus *Siphoptychium* and introduction of the new genus *Thecotubifera*. Mycologia 111: 981–997. <https://doi.org/10.1080/00275514.2019.1660842>
- Leontyev DV, Kochergina AV, Inoue M, Lloyd SJ, Schnittler M (2026) A five-finger exercise: five overlooked non-European species of *Tubifera*. Fungal Systematics and Evolution 17: 110–123. <https://doi.org/10.15761/fuse.2026.17.07>
- Leontyev DV, Schnittler M, Stephenson SL, Novozhilov YK, Shchepin ON (2019b) Towards a phylogenetic classification of the Myxomycetes. Phytotaxa 399: 209. <https://doi.org/10.11646/phytotaxa.399.3.5>
- Mandic-Mulec I, Stefanic P, Van Elsas JD (2015) Ecology of Bacillaceae. Microbiology Spectrum 3: TBS-0017-2013. <https://doi.org/10.1128/microbiolspec.TBS-0017-2013>
- Maraun M, Scheu S (2000) The structure of oribatid mite communities (Acari, Oribatida): patterns, mechanisms and implications for future research. Ecography 23: 374–382. <https://doi.org/10.1111/j.1600-0587.2000.tb00294.x>
- Miller D, Padmanabhan R, Sarcar SN (2021) Genomics and gene expression in myxomycetes. In: Stephenson SL, Rojas C (Eds) Myxomycetes: Biology, systematics, biogeography and ecology. Academic Press, London, 377–416.
- Nurk S, Meleshko D, Korobeynikov A, et al. (2017) metaSPAdes: A new versatile metagenomic assembler. Genome Research 27: 824–834. <https://doi.org/10.1101/gr.213959.116>
- Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences USA 86: 1939–1943. <https://doi.org/10.1073/pnas.86.6.1939>
- Pääbo S, Poinar H, Serre D, et al. (2004) Genetic analyses from ancient DNA. Annual Review of Genetics 38: 645–679. <https://doi.org/10.1146/annurev.genet.37.110801.143214>
- Prikhodko IS, Shchepin ON, Bortnikova NA, et al. (2023) A three-gene phylogeny supports taxonomic rearrangements in the family Didymiaceae (Myxomycetes). Mycological Progress 22: 1–24. <https://doi.org/10.1007/S11557-022-01858-1>
- Prjibelski A, Antipov D, Meleshko D, et al. (2020) Using SPAdes de novo assembler. Current Protocols in Bioinformatics 70: e102. <https://doi.org/10.1002/cpbi.102>

- Raxworthy CJ, Smith BT (2021) Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution* 36: 1049–1060.  
<https://doi.org/10.1016/j.tree.2021.07.009>
- Ronikier A, Janik P, de Haan M, et al. (2022) Importance of type specimen study for understanding genus boundaries-taxonomic clarifications in *Lepidoderma* based on integrative taxonomy approach leading to resurrection of the old genus *Polyschismium*. *Mycologia* 114: 1008–1031.  
<https://doi.org/10.1080/00275514.2022.2109914>
- Sahlin K (2022) Strobealign: flexible seed size enables ultra-fast and accurate read alignment. *Genome Biology* 23: 260. <https://doi.org/10.1186/s13059-022-02831-7>
- Schaap P, Barrantes I, Minx P, et al. (2016) The *Physarum polycephalum* genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. *Genome Biology and Evolution* 8: 109–125. <https://doi.org/10.1093/gbe/evv237>
- Schnittler M, Dagamac NHA, Leontyev D, et al. (2020) Quick n' Cheap – a simplified workflow to barcode plasmodial slime molds (Myxomycetes). *Karstenia* 58: 385–392.  
<https://doi.org/10.29203/ka.2020.504>
- Schnittler M, Leontyev D, Yatsiuk I, et al. (2025) Species descriptions in myxomycetes – can we settle on rules for good taxonomic practice? *IMA Fungus* 16: e141199.  
<https://doi.org/10.3897/imafungus.16.141199>
- Schnittler M, Shchepin ON, Dagamac NHA, et al. (2017) Barcoding myxomycetes with molecular markers: challenges and opportunities. *Nova Hedwigia* 104: 323–341.  
[https://doi.org/10.1127/nova\\_hedwigia/2017/0397](https://doi.org/10.1127/nova_hedwigia/2017/0397)
- Shumskaya M, Mironov KS, Ballesteros JA, et al. (2023) DNA isolation and genome sequence of the 134-year-old holotype specimen of *Boletus subvelutipes* Peck. *Ecology and Evolution* 13: e10389. <https://doi.org/10.1002/ece3.10389>
- Singh G (2019) *Plant Systematics*. CRC Press, Boca Raton, 792 pp.  
<https://doi.org/10.1201/9780429289521>
- Stephenson SL, Schnittler M, Novozhilov YK (2008) Myxomycete diversity and distribution from the fossil record to the present. *Biodiversity and Conservation* 17: 285–301.  
[http://dx.doi.org/10.1007/978-90-481-2801-3\\_5](http://dx.doi.org/10.1007/978-90-481-2801-3_5)
- Straub SCK, Parks M, Weitemier K, et al. (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.  
<https://doi.org/10.3732/ajb.1100335>
- Takano H, Abe T, Sakurai R, et al. (2001) The complete DNA sequence of the mitochondrial genome of *Physarum polycephalum*. *Molecular and General Genetics* 264: 539–545.  
<https://doi.org/10.1007/s004380000357>
- Turland N, Wiersema J, Barrie F, et al. (2018) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Glashütten.
- Winston JE (1999) *Describing Species: Practical Taxonomic Procedure for Biologists*. Columbia University Press, New York, 518 pp.
- Zeng CX, Hollingsworth PM, Yang J, et al. (2018) Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* 14: 43. <https://doi.org/10.1186/s13007-018-0300-0>

Table 1. Sequencing data for the herbarium collections used in this study.

Species	<i>Fuligo cinerea</i>	<i>Trichia scabra</i>	<i>Lindbladia tubulina</i>	<i>Cribraria mirabilis</i>	<i>Lycogala epidendrum</i>	<i>Lindbladia tubulina</i>	<i>Lycogala epidendrum</i>
Species ID	sc11445	SC11426	sc12075	sc10428	UARK25089	sc2833	sc11299
Age of collection, years	91	64	51	47	41	33	29
DNA concentration before WGA, ng/ $\mu$ L	0.55	4.53	2.61	1.85	1.73	8.57	4.24
DNA concentration after WGA, ng/ $\mu$ L	5.07	–	9.14	4.51	2.76	–	4.94
DNA amount used for sequencing, ng	365	158	658	325	199	300	356
Total sequence length, Mbp	1532.1	1670.7	1133.9	1463.4	1425.4	2353.5	1516.7
Total assembly length, Mbp	11.1	162	9.1	2.6	1.7	84.1	16.8
No-hit fraction length, Mbp	5.1	108	4.7	2	1.4	82.3	16.6
Eumycetozoa fraction length, kbp	1.5	49.7	0.4	0.2	0.5	111	5.1
Eumycetozoa contigs count	6	17	2	1	2	19	11
No-hit & Eumycetozoa N50, bp	251	266	214	207	208	4190	243
No-hit & Eumycetozoa longest contig, kbp	8.2	41.6	4.1	1.81	1.63	114	4.7
No-hit & Eumycetozoa mean GC content, %	50.5	53.7	35.7	37.1	32.8	37.9	35.7
Interpretation of dominant contaminants	Decomposition microbiota: Saprophytic bacteria typical of long-term decay.	Forest substrate: Decaying wood microflora and photosynthetic cyanobacteria.	Mixed noise: indoor dust, mammalian DNA, and forest mesofauna (mites).	Indoor dust: fungal and bacterial microflora of residential areas.	Indoor dust: skin microflora and associated dermatophytic fungi.	Soil microbiota: predominantly nitrogen-fixing soil bacteria.	Anthropogenic noise: trace human DNA and typical skin microflora.



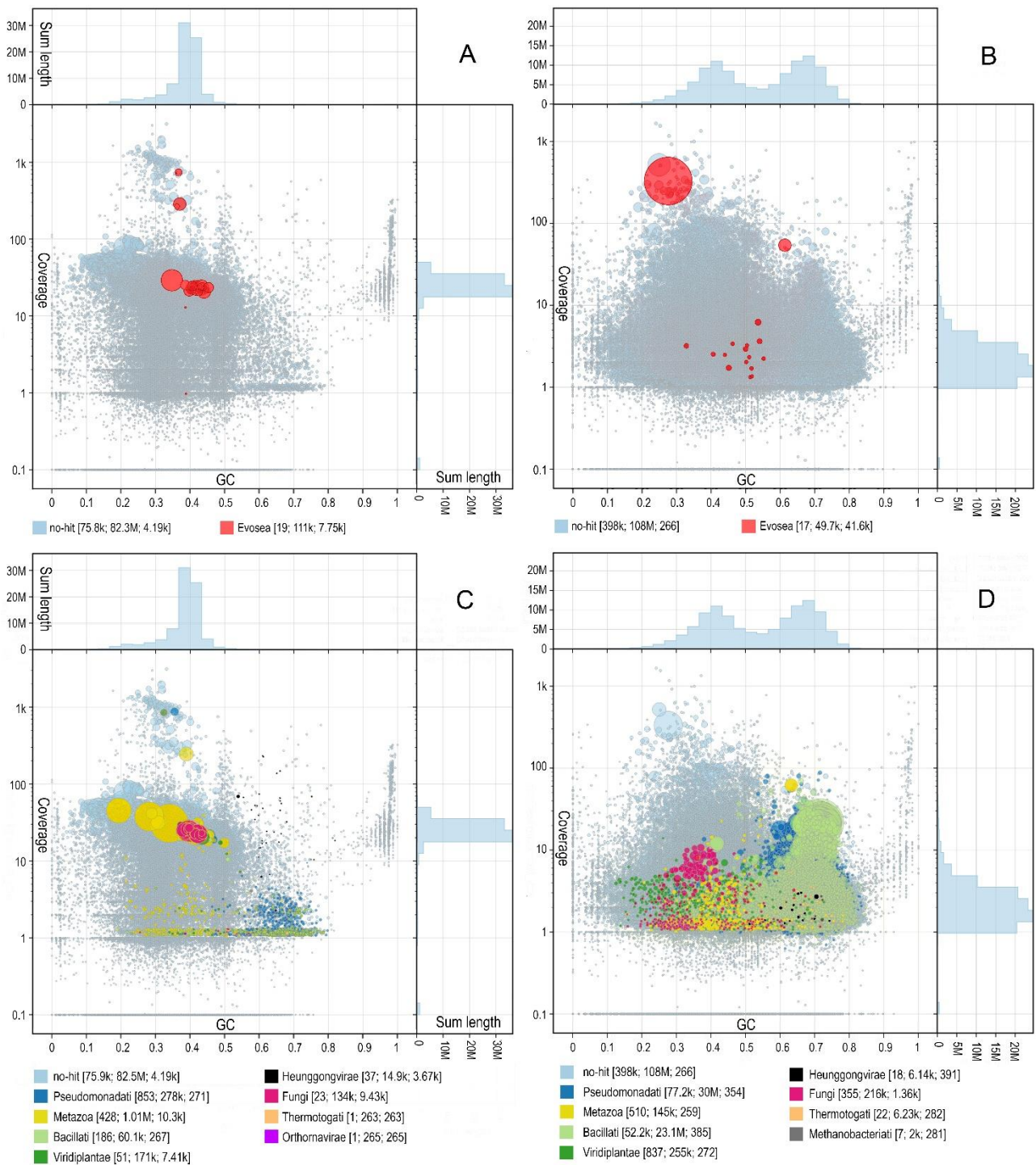


Fig. 1. Blob plot summary for assembled contigs of *Lindbladia tubulina* sc2833, 33 years old (A, C) and *Trichia scabra* sc11426, 64 years old (B, D). Axes: base coverage (logarithmic) against GC proportion. Circles are sized in proportion to contig length. Histograms show the distribution of contig length sum along each axis. A, B – Eumycetozoa contigs shown against no-hit fraction. C, D – leading groups of the sample metagenome at the level of kingdoms are shown.

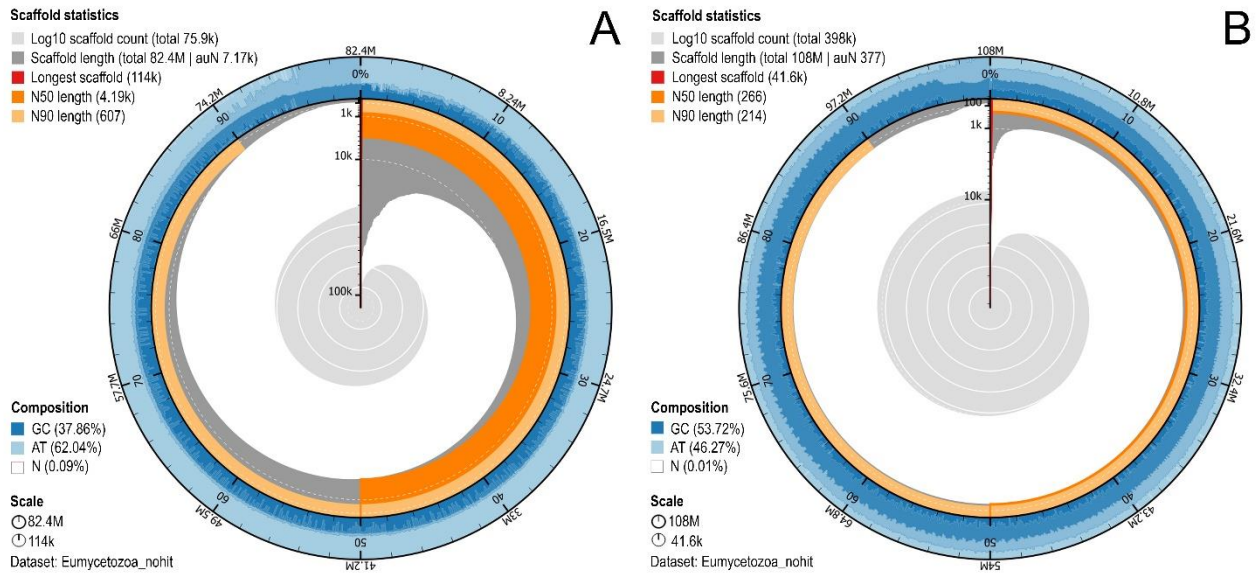


Fig. 2. Snail plot of assembly statistics for *Lindbladia tubulina* sc2833, 33 years old (A), and *Trichia scabra* sc11426, 64 years old (B). The dark grey area represents the sequence length distribution, with the radius scaled to the longest sequence (red line, 105k and 13.2k, respectively). Orange and pale-orange arcs indicate N50 and N90 lengths, respectively. The pale grey spiral shows the cumulative sequence count on a logarithmic scale (white lines show orders of magnitude). The outer bands visualize GC (blue), AT (pale-blue), and non-identified bases (white) content distributions, binned consistently with the inner sequence data.

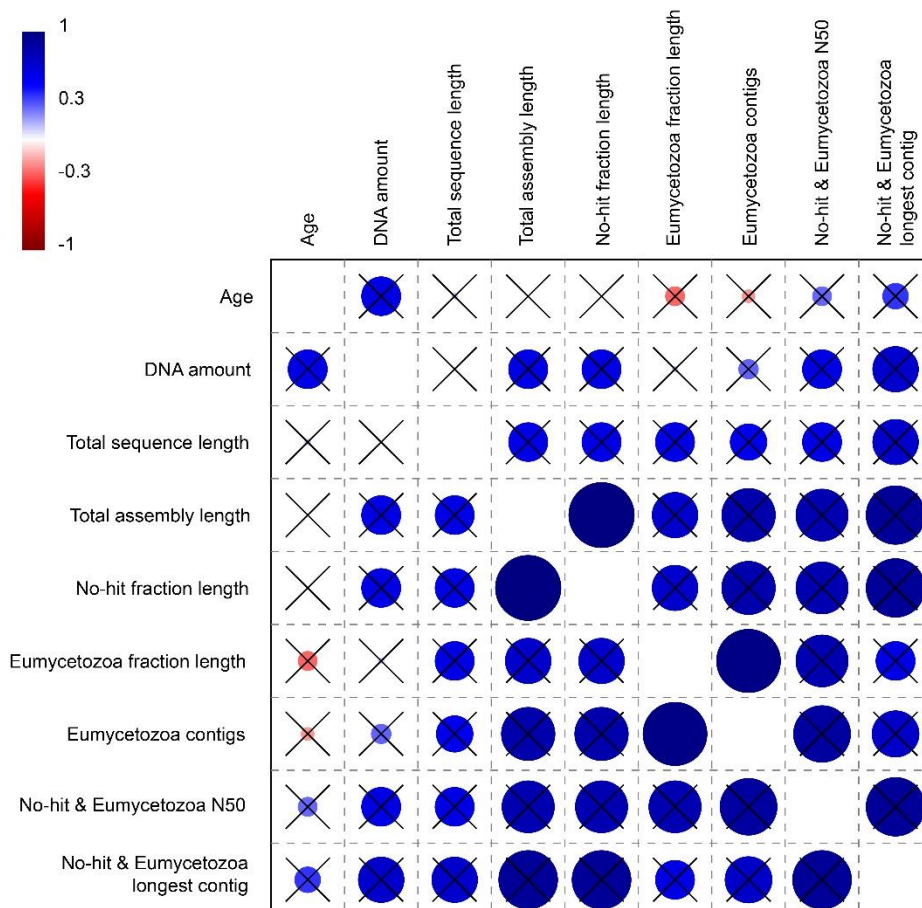


Fig. 3. Spearman correlation matrix ( $r_s$ ) between specimen age, DNA concentration, and bioinformatic characteristics of the sequencing data, calculated for five specimens sequenced after WGA. Blue and red colours represent a positive and negative correlation respectively. The size of the circles is proportional to the absolute value of the correlation coefficient  $r_s$ . Non-significant correlations ( $p > 0.05$ ) are marked with crosses. If the Bonferroni correction is applied, all values are insignificant. The diagram was generated using the PAST software.

## SUPPLEMENTS

**Supplementary file 1.** *Sheet A.* Collection data for the specimens used in the study. *Sheet B.* MultiBLAST for contigs where myxomycete genes could be identified. The established DNA barcode (nucSSU / 18S rRNA gene) is shown in red. *Sheet C.* Genes recognised in contigs identified as Eumycetozoa. Standard barcoding and phylogenetic markers are shown in red. The total length / base coverage are shown for each contig (*xlsx file*).

**Supplementary file 2.** Sequences obtained from each studied specimen by Illumina NovaSeq-based genome skimming and recognised as Eumycetozoa (*ZIP archive of FASTA files*).

**Supplementary file 3.** Alignment of the fragment of nucSSU (18S rRNA gene) used for species barcoding in our data and reference barcodes of the respective taxa obtained by Sanger sequencing (*FASTA file*).

**Supplementary file 4.** Blob-plots and snail-plots of genomic assemblies for all seven study specimens. Blob-plots display base coverage and fragment length against GC content with the identification of (A) all contigs and (B) Eumycetozoa contigs at all taxonomic levels, from species to kingdom. Snail-plots (C) visualize assembly statistics, including N50/N90 lengths, cumulative sequence count, and base composition (*ZIP archive of JPG files available via Zenodo: <https://doi.org/10.5281/zenodo.20467237>*).

## CONTRIBUTION

DL: Conceptualization, Investigation, Visualization, Writing – original draft, Funding acquisition

MS: Supervision, Writing – review & editing

OS: Supervision, Investigation, Data curation, Visualization, Writing – review & editing