

# De novo transcriptome assembly of the Baikal stone sculpin *Paracottus knerii* (Dybowski, 1874) from the Angara River as the first transcriptomic resource for Lake Baikal Cottidae

Vasiliy Pomazkin<sup>1</sup>, Polina Drozdova<sup>1,2</sup>, Ekaterina Borvinskaya<sup>1</sup>,  
Elizaveta Indosova<sup>1</sup>, Maxim Timofeyev<sup>1</sup>, Anton Gurkov<sup>1,2</sup>

**1** Institute of Biology, Irkutsk State University, 3 Lenin st., Irkutsk, 664025, Russia

**2** Baikal Research Centre, 5v Rabochaya st., Irkutsk, 664011, Russia

Corresponding author: Anton Gurkov (a.n.gurkov@gmail.com)

---

Academic editor: R. Yakovlev | Received 3 April 2025 | Accepted 10 April 2025 | Published 17 April 2025

---

<http://zoobank.org/3C097B75-D850-4DF7-B002-3A8C13712955>

---

**Citation:** Pomazkin V, Drozdova P, Borvinskaya E, Indosova E, Timofeyev M, Gurkov A (2025) *De novo* transcriptome assembly of the Baikal stone sculpin *Paracottus knerii* (Dybowski, 1874) from the Angara River as the first transcriptomic resource for Lake Baikal Cottidae. *Acta Biologica Sibirica* 11: 541–550. <https://doi.org/10.5281/zenodo.15208846>

---

## Abstract

Lake Baikal is the largest freshwater reservoir inhabited by a number of endemic species flocks belonging to different taxa. However, genomic resources for these groups of endemics are still relatively scarce, limiting understanding of the molecular mechanisms behind their physiology. One of these species flocks are Baikal sculpins of the family Cottidae. Here, we present the first transcriptome assembly for this group using stone sculpin *Paracottus knerii* (Dybowski, 1874) as a model. The transcriptome was obtained from the whole body of a *P. knerii* fry and contains a diverse array of immunity-related transcripts, paving the way for studies investigating the immune response of Lake Baikal sculpins.

## Keywords

Benthic fish, Baikal, Cottidae, immune system, immune response, transcriptome, sculpin

## Introduction

Ancient Lake Baikal, containing a fifth of the surface liquid freshwater on the planet, is known for extensive endemic biodiversity (Cristescu et al. 2010; Brown et al. 2021). The variety of animal species, which belong to a number of taxonomic groups, complicates the accumulation of the databases of protein- and rRNA-coding sequences for the endemics, which in turn hampers physiological studies, monitoring programs, and conservation efforts. Baikal sculpins (family Cottidae, Pisces) represent one of the adaptive radiations that have happened in the lake. These fishes diversified to the amazing variety of ~40 species (Goto et al. 2015; Bogdanov 2023) and inhabit all depths of the lake. Some species descend to the Angara river, the only outflow of the lake (Taliev 1955; Bogdanov 2023). However, the genetics of this group remains relatively less explored than morphology (Sandel et al. 2024).

For six species of Baikal Cottidae, mitochondrial genomes have been published (Sandel et al. 2017; Mugue et al. 2021), but while undoubtedly useful for species identification and population genomics, they do not offer much for researchers who wish to study the diversity and expression of protein-coding genes. A chromosome-level genome assembly for *Paracottus knerii* (Dybowsky 1874) was announced (Mugue et al. 2023), but has not been made publicly available yet.

To bridge this gap, we here aim to obtain exome-wide data for *P. knerii* as one of the most common species by sequencing and assembling a whole-body transcriptome. *P. knerii*, the so-called stone sculpin, inhabits the littoral zone of the lake, which is the part subjected to climate change and local coastal eutrophication (Brown et al. 2021). Furthermore, it has been used as a model species to study noise exposure as a component of human-induced rapid environmental change and for ecotoxicological research (Sapozhnikova et al. 2021; Sudakov et al. 2022). This species descends to the Angara and Yenisei rivers. Individuals from Angara are slightly different from those from Baikal in terms of morphology, but their morphometric results overlap (Bogdanov 2007).

## Materials and methods

### Animal sampling and maintenance

The studied species is neither endangered nor protected; the research was approved by the Animal Subjects Research Committee of the Institute of Biology at Irkutsk State University (Protocol #2024/13). A 15-mm-long *P. knerii* fry was sampled in the Angara River in Irkutsk (52.27769° N, 104.27568° E) in August 2024 and identified according to (Taliev, 1955). It was maintained for 40 days at 15 °C with nauplii of *Artemia salina* (Linnaeus, 1758) provided as feed, then euthanized (250 µl/l clove oil emulsion), dissected to remove the gut and preserved in RNAlater (Thermo Fisher Scientific) for transport.

## RNA sequencing

Total RNA was isolated with a MagMax kit (Thermo Fisher Scientific). After quality control (RNA concentration >100 ng/uL, RIN 7.6), 100 ng of RNA was processed with a TruSeq Stranded mRNA library preparation kit (Illumina), and the library was sequenced with a NovaSeq 6000 device (Illumina) (2 x 101 bp). Demultiplexing of the sequencing reads was performed with bcl2fastq v2.20 (Illumina). Adapters were trimmed with Skewer v0.2.2 (Jiang et al. 2014). RNA extraction, sequencing, and data analysis up to this point were performed by the CeGaT company (Germany). Read quality was analyzed with FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and allowed us to proceed directly with assembly.

## Transcriptome assembly

All analyses except indicated otherwise were performed using a small computing cluster (64 Gb RAM, 6 physical cores, 12 virtual cores). Reproducible code for all analyses is available from GitHub ([https://github.com/drozdovapb/Paracottus\\_knerii\\_transcriptome/](https://github.com/drozdovapb/Paracottus_knerii_transcriptome/)).

The main transcriptome assembly used in the subsequent analyses was performed with rnaSPAdes v3.13.1 (Bushmanova et al. 2019) using the option --ss-fr. Additionally, the Oyster River Protocol v2.3.3 was used to compare alternative assemblers (MacManes, 2018). Assembly quality was controlled with BUSCO v5.4.5 (Manni et al. 2021), which uses hmmsearch v3.3 (Eddy 2011) and Metaeuk v6.a5d39d9 (Karin et al. 2020), using the Actinopterygii database (actinopterygii\_odb10).

## Transcriptome annotation

The correspondence of the sample to the species *P. knerii* was confirmed by searching the transcriptome assembly for the reference mitochondrial genome sequence of this species (NCBI GenBank ID: MW732164; Mugue et al. 2021) with exonerate v2.4.0 (Slater & Birney 2005). Exonerate is part of Ensembl (Harrison et al. 2024).

The assembly was filtered in multiple steps to only retain the most reliable transcripts belonging to Chordata. First, kentUtils (<https://github.com/ENCODE-DCC/kentUtils>) were used to filter by length (>199 bp) and Ns (<9N). The resulting file passed the TRAPID threshold of at most 200,000 sequences and was taxonomically annotated with the TRAPID server (Bucchini et al. 2021), which uses kaiju v1.7.3 (Menzel et al. 2016) for classification. In parallel, protein prediction was performed with TransDecoder v5.7.0 (Haas et al. 2014) with the following non-default options: -m 50 (minimal length of 50 amino acids) and -single-best-only (only one best open reading frame predicted per transcript). The predicted proteome was mapped against the eggNOG v5.0 database (Huerta-Cepas et al. 2019) with the eggNOG-mapper web server v2.1.12 (Cantalapiedra et al. 2021) to obtain functional and ad-

ditional taxonomic annotation. The transcriptome assembly was filtered so that it contained only sequences identified as Chordata according to either TRAPID or eggNOG-mapper. The Euler diagram describing the intersection in their results was generated with the BioVenn package v1.1.3 (Hulsen 2021).

Next, the NCBI FCS pipeline v0.5.4 (Astashyn et al. 2024) was used to find and filter out potential contamination from the assembly through the public Galaxy server usegalaxy.org (The Galaxy Community 2024; Afgan et al. 2016). The lists were prepared with R v4.1.2 (R Core Team 2021). Filtering was done with seqkit v2.1.0 (Shen et al. 2024). Finally, annotation of the transcriptome assembly according to the best hits in the eggNOG database was prepared with R and appended to each sequence title with seqkit.

## Data availability

The final filtered assembly was submitted to NCBI TSA database under the accession number GLBE00000000 (<https://www.ncbi.nlm.nih.gov/nuccore/GLBE00000000.1>). Raw unfiltered assembly can be found via the GitHub repository ([https://github.com/drozdovapb/Paracottus\\_knerii\\_transcriptome/blob/main/02\\_assemblies/Pkn\\_rnaspades\\_ssrf.fasta.gz](https://github.com/drozdovapb/Paracottus_knerii_transcriptome/blob/main/02_assemblies/Pkn_rnaspades_ssrf.fasta.gz)). The reads were submitted to NCBI Bio-Project (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1200955>) and SRA (<https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR31806589>).

## Annotation to the KEGG pathway database

Involvement of transcripts in the immune response was assessed using the eggNOG-mapper annotation. The numbers of transcripts classified as belonging to the KEGG (Kanehisa et al. 2025) pathways related to the immune system (<https://www.genome.jp/kegg/pathway.html#organismal>) were calculated for the final filtered transcriptome assembly. We used the classification of the pathways according to KEGG release 85.0 in order to match the slightly outdated classification in the eggNOG v5.0 database. The number of transcripts annotated with each pathway of interest was counted with R and visualized with the ggplot2 v3.5.0 package (Wickham 2016).

## Results and discussion

### Read quality control and assembly

In total, 51M paired reads (10G bases) were obtained for *P. knerii* fry transcripts. Quality control with FastQC showed good quality (Q>28) and absence of remaining sequence adapters, thus the reads were directly used for transcriptome assembly. The rnaSPAdes assembly had the following BUSCO score: C:80.5%[S:64.2%,D:16.

3%], F:5.0%, M:14.5%, n:3640. It is not close to the ideally desired 100 % but comparable with other published fish transcriptome assemblies (e.g., Kokkonen et al., 2024) and is suitable for most downstream applications, such as primer design and as a database for mass spectrometric proteome analysis. As it is generally recommended to compare several alternative assemblers (Raghavan et al. 2022), we also performed this step with the Oyster River protocol, which compares four strategies and integrates them into one assembly. However, the best output of the Oyster River protocol was slightly worse (C:74.2% [S:52.2%, D:22.0%], F:6.7%, M:19.1%, n:364), so the rnaSPAdes assembly was used for the downstream analysis.

### Taxonomic annotation

To validate species identification, we performed a search for mitochondrial contigs in the transcriptome assembly. The three best hits (GLBE01000028.1, GLBE01000454.1, and GLBE01023288.1) covered >98% of the reference mitochondrial genome of *P. kneri* with >99% identity, undoubtedly confirming that the sample belonged to this species.

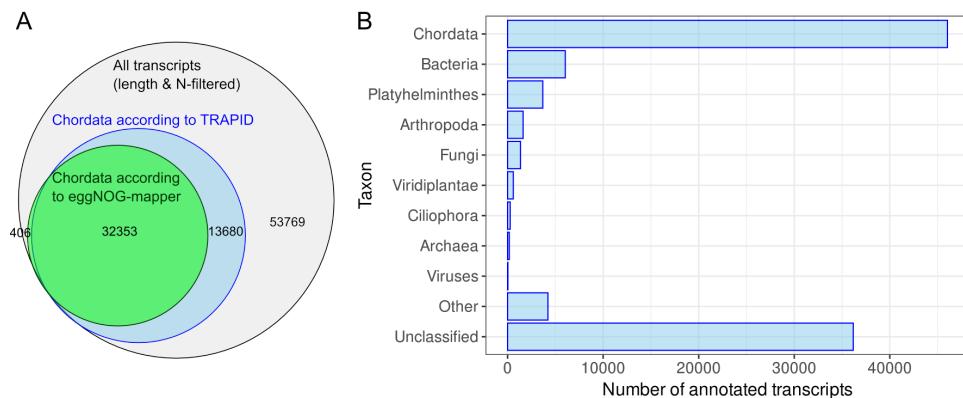
For taxonomic annotation of the assembly, we used TRAPID and eggNOG, and the transcripts classified as belonging to Chordata by these two tools largely coincided (Fig. 1A). TRAPID indicated a substantially higher number of Chordata transcripts, so we focused on this tool to assess the taxonomic annotation of the rest of transcripts (Fig. 1B). We found representatives of multiple taxonomic groups such as bacteria, flatworms, ciliates, viruses etc. in the unfiltered transcriptome assembly, which was expected given the high diversity of symbionts known for this species (Rusinek et al. 2024) and the fact that we used most of the body for RNA extraction. Although the fry gut was dissected to decrease contamination from the food, the relatively high amount of arthropod transcripts (Fig. 1B) indicated that contamination was still the case. Therefore, we cannot exclude the possibility that the identified transcripts of other taxonomic groups were sequenced from intestinal symbionts and not from those located in tissues or on the surface of the skin.

Importantly, 46,439 out of 100,208 transcripts were classified by either of the tools as coming from Chordata, that is, most probably from our object. This set was slightly reduced by the FCS pipeline to 46,413 transcripts due to possible contamination, and the resulting assembly was submitted to NCBI.

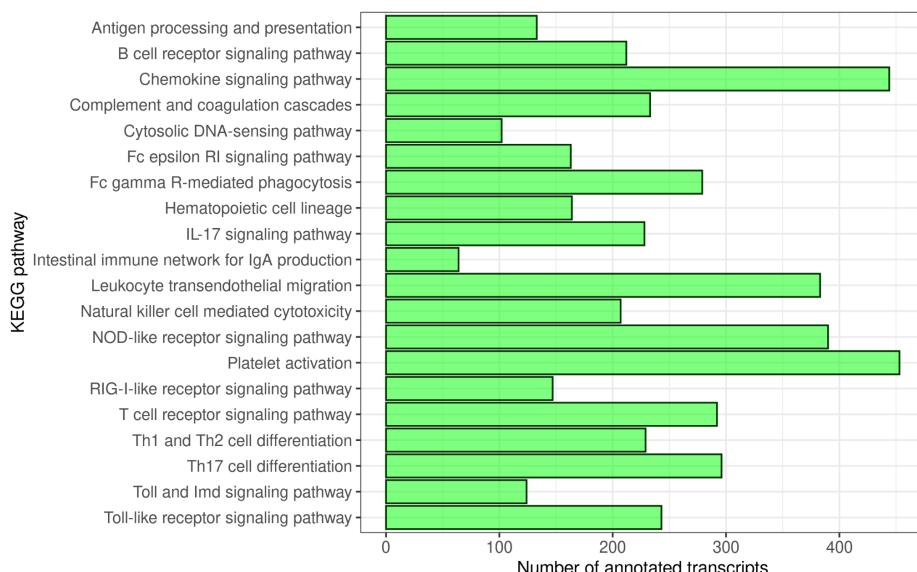
### The diversity of immunity-related transcripts

To evaluate the diversity of functional proteins in the filtered transcriptome assembly, we concentrated on the transcripts involved in the immune response. For this, we used annotation of all vertebrate transcripts according to eggNOG-mapper and calculated how many of them belong to KEGG pathways related to the immune system (Fig. 2). The results showed that the transcriptome of *P. kneri* includes from dozens to hundreds of transcripts related to each pathway. These numbers are com-

parable to the output of a similar analysis performed earlier for the *Betta splendens* transcriptome (Amparyup et al. 2020), and indicate that the obtained transcriptome assembly can be useful in search for immunity-related and probably other functional proteins of Lake Baikal sculpins.



**Figure 1.** Taxonomic annotation of the transcripts obtained. (A) Euler diagram showing intersection between the sets of transcripts annotated as Chordata by TRAPID and eggNOG-mapper. (B) The numbers of transcripts belonging to a manually selected set of taxonomic groups according to TRAPID.



**Figure 2.** Numbers of transcripts involved in immune system-related KEGG pathways according to the annotation by eggNOG-mapper. The analysis was applied to the final filtered transcriptome assembly, i.e. only to the sequences clearly belonging to Chordata.

## Conclusion

Here we presented the transcriptome assembly of a young stone sculpin *P. knerii* as the first exome-wide resource for searching functional proteins for the endemic diversity of Cottidae of Lake Baikal. The assembly published in NCBI was filtered to contain only transcripts most probably belonging to the host, but available raw reads can also be used for the search of sequences of the fish symbionts. Despite dissection to remove the gut, the reads probably contain RNA from the content of the digestive system, and the exact location of the symbionts in the fish body is unclear. Importantly, the obtained assembly contains a wide diversity of transcripts related to the immune system and can be used in physiological studies investigating the immune response of Lake Baikal sculpins.

## Acknowledgements

The study was supported by the Russian Science Foundation within project #24-74-00095 (<https://rscf.ru/en/project/24-74-00095/>).

## References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* 44: W3–W10. <https://doi.org/10.1093/nar/gkw343>

Amparyup P, Charoensapsri W, Samaluka N, Chumtong P, Yocawibun P, Imjongjirak C (2020) Transcriptome analysis identifies immune-related genes and antimicrobial peptides in Siamese fighting fish (*Betta splendens*). *Fish & Shellfish Immunology* 99: 403–413. <https://doi.org/10.1016/j.fsi.2020.02.030>

Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strope PK, Sylla PM, Wagner L, Bidwell SL, Brown LC, Clark K, Davis EW, Smith-White B, Hlavina W, Pruitt KD, Schneider VA, Murphy TD (2024) Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology* 25: 60. <https://doi.org/10.1186/s13059-024-03198-7>

Bogdanov BE (2007) Variation of stone sculpin *Paracottus knerii* (Cottidae, Scorpaeniformes) of Baikal and waters of Baikal region. *Journal of Ichthyology* 47: 162–174. <https://doi.org/10.1134/S003294520702004X>

Bogdanov BE (2023) The Sculpins (Perciformes: Cottidae) of Lake Baikal and Baikal region: updated checklist with the description of new tax. *Limnology and Freshwater Biology* 6(3): 63–95. <https://doi.org/10.31951/2658-3518-2023-A-3-63>

Brown KP, Gerber A, Bedulina D, Timofeyev MA (2021) Human impact and ecosystemic health at Lake Baikal. *WIREs Water* 8: e1528. <https://doi.org/10.1002/wat2.1528>

Bucchini F, Del Cortona A, Kreft L, Botzki A, Van Bel M, Vandepoele K (2021) TRAPID 2.0: a web application for taxonomic and functional analysis of *de novo* transcriptomes. *Nucleic Acids Research* 49: e101–e101. <https://doi.org/10.1093/nar/gkab565>

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8: giz100. <https://doi.org/10.1093/gigascience/giz100>

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38 (12): 5825–5829. <https://doi.org/10.1093/molbev/msab293>

Cristescu ME, Adamowicz SJ, Vaillant JJ, Haffner DG (2010) Ancient lakes revisited: from the ecology to the genetics of speciation. *Molecular Ecology* 19: 4837–4851. <https://doi.org/10.1111/j.1365-294X.2010.04832.x>

Eddy SR (2011) Accelerated Profile HMM Searches. *PLOS Computational Biology* 7: e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>

Goto A, Yokoyama R, Sideleva VG (2015) Evolutionary diversification in freshwater sculpins (Cottoidea): a review of two major adaptive radiations. *Environmental Biology of Fishes* 98: 307–335. <https://doi.org/10.1007/s10641-014-0262-7>

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512. <https://doi.org/10.1038/nprot.2013.084>

Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Becker A, Bennett R, Berry A, Bhai J, Bhurji SK, Boddu S, Branco Lins PR, Brooks L, Ramaraju SB, Campbell LI, Martinez MC, Charkhchi M, Chougule K, Cockburn A, Davidson C, De Silva NH, Dodiya K, Donaldson S, El Houdaigui B, Naboulsi TE, Fatima R, Giron CG, Genez T, Grigoriadis D, Ghattaoraya GS, Martinez JG, Gurbich TA, Hardy M, Hollis Z, Hourlier T, Hunt T, Kay M, Kaykala V, Le T, Lemos D, Lodha D, Marques-Coelho D, Maslen G, Merino GA, Mirabueno LP, Mushtaq A, Hossain SN, Ogeh DN, Sakthivel MP, Parker A, Perry M, Piližota I, Poppleton D, Prosovetskaia I, Raj S, Pérez-Silva JG, Salam AIA, Saraf S, Saraiva-Agostinho N, Sheppard D, Sinha S, Sipos B, Sitnik V, Stark W, Steed E, Suner M-M, Surapaneni L, Sutinen K, Tricomi FF, Urbina-Gómez D, Vedenberg A, Walsh TA, Ware D, Wass E, Willhoft NL, Allen J, Alvarez-Jarreta J, Chakachvili M, Flint B, Giorgetti S, Haggerty L, Ilsley GR, Keatley J, Loveland JE, Moore B, Mudge JM, Naamati G, Tate J, Trevanion SJ, Winterbottom A, Frankish A, Hunt SE, Cunningham F, Dyer S, Finn RD, Martin FJ, Yates AD (2024) Ensembl 2024. *Nucleic Acids Research* 52: D891–D899. <https://doi.org/10.1093/nar/gkad1049>

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019) eggNOG 5.0: a hi-

erarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47: D309–D314. <https://doi.org/10.1093/nar/gky1085>

Hulsen T (2021) BioVenn – an R and Python package for the comparison and visualization of biological lists using area-proportional Venn diagrams. *Data Science* 4: 51–61. <https://doi.org/10.3233/DS-210032>

Jiang H, Lei R, Ding S-W, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15: 182. <https://doi.org/10.1186/1471-2105-15-182>

Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M (2025) KEGG: biological systems database as a model of the real world. *Nucleic Acids Research* 53: D672–D677. <https://doi.org/10.1093/nar/gkae909>

Kokkonen AL, Searle PC, Shiozawa DK, Evans RP (2024) Using *de novo* transcriptomes to decipher the relationships in cutthroat trout subspecies (*Oncorhynchus clarkii*). *Evolutionary Applications* 17: e13735. <https://doi.org/10.1111/eva.13735>

Levy Karin E, Mirdita M, Söding J (2020) MetaEuk – sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* 8: 48. <https://doi.org/10.1186/s40168-020-00808-x>

MacManes MD (2018) The Oyster River Protocol: a multi-assembler and kmer approach for *de novo* transcriptome assembly. *PeerJ* 6: e5428. <https://doi.org/10.7717/peerj.5428>

Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021) BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols* 1: e323. <https://doi.org/10.1002/cpz1.323>

Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7: 11257. <https://doi.org/10.1038/ncomms11257>

Mugue N, Barmintseva A, Etingova A, Didorenko S, Selivanova M, Mugue L, Popov A, Bulyakov A, Kupchinskiy A (2021) Complete mitochondrial genomes of representatives of two endemic sculpin families (Perciformes: Cottoidei) from Baikal – the world's largest and deepest lake. *Mitochondrial DNA Part B* 6: 3190–3192. <https://doi.org/10.1080/23802359.2021.1989330>

Mugue NS, Selivanova MA, Etingova AA, Didorenko SI, Barmintseva AE, Mugue LN, Kupchinskiy AB (2023) Genomic research on endemic species complex of sculpins from Lake Baikal. In: Zaytseva EA (Ed.) *Life development under abiotic changes on Earth. Proceedings of the IV All-Russian conference, September 2023, Listvyanka (Russia)*. Publishing house of Irkutsk State University, Irkutsk, 80–84. <https://doi.org/10.24412/CL-34446-2023-4-80-84> [In Russian]

Raghavan V, Kraft L, Mesny F, Rigerte L (2022) A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics* 23: bbab563. <https://doi.org/10.1093/bib/bbab563>

Rusinek OT (2024) Parasite Fauna of Sand and Stone Sculpins (Scorpaeniformes, Cottidae) from Lake Baikal (Review of Results and Prospects of Investigations). *Biology Bulletin Reviews* 14: S91–S101. <https://doi.org/10.1134/S2079086424700105>

Sandel MW, Aguilar A, Neely D, Kirilchik S, Bogdanov B, Fast KM, Millwood J (2024) Sink or Swim: Phylogenomic Analysis of Baikal Sculpins Reveals Multiple Transformations to Pelagic, Bathybenthic, and Lotic Ecomorphologies. Available at SSRN: <https://doi.org/10.2139/ssrn.4888603>

Sandel MW, Aguilar A, Fast K, O'Brien S, Lapidus A, Allison DB, Teterina V, Kirilchik S (2017) Complete mitochondrial genomes of Baikal oilfishes (Perciformes: Cottoidei), Earth's deepest-swimming freshwater fishes. *Mitochondrial DNA Part B* 2: 773–775. <https://doi.org/10.1080/23802359.2017.1398603>

Sapozhnikova YP, Koroleva AG, Yakhnenko VM, Khanaev IV, Glyzina OYu, Avezova TN, Volkova AA, Mushinskaya AV, Tyagun ML, Shagun AN, Makarov MM, Kirilchik SV, Sudakov NP, Klimenkov IV, Sukhanova LV (2021) Sex Associated Effects of Noise Pollution in Stone Sculpin (*Paracottus kneri*) as a Model Object in the Context of Human-Induced Rapid Environmental Change. *Biology* 10: 1063. <https://doi.org/10.3390/biology10101063>

Shen W, Sipos B, Zhao L (2024) SeqKit2: A Swiss army knife for sequence and alignment processing. *iMeta* 3: e191. <https://doi.org/10.1002/imt2.191>

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31. <https://doi.org/10.1186/1471-2105-6-31>

Sudakov NP, Klimenkov IV, Bedoshvili YD, Arsent'ev KY, Gorshkov AG, Izosimova ON, Yakhnenko VM, Kupchinskii AB, Didorenko SI, Likhoshway YV (2022) Early structural and functional changes in Baikal Sculpin gills exposed to suspended soot micro-particles in experiment. *Chemosphere* 290: 133241. <https://doi.org/10.1016/j.chemosphere.2021.133241>

Taliev DN (1955) The Sculpins of Baikal (Cottoidei). Publishing house of the Academy of Sciences of USSR, Moscow, 604 pp. [In Russian]

The Galaxy Community (2024) The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research* 52: W83–W94. <https://doi.org/10.1093/nar/gkae410>

Wickham H (2016) Programming with ggplot2. In: Wickham H (Ed.) *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing, Cham, 241–253. [https://doi.org/10.1007/978-3-319-24277-4\\_12](https://doi.org/10.1007/978-3-319-24277-4_12)