

SCOR Working Group proposal (2019-2022)

MetaZooGene

**Toward a new global view of marine zooplankton biodiversity
based on DNA metabarcoding and reference DNA sequence databases**

Submitted by:

Ann Bucklin (Proposed Chair)
Professor, Department of Marine Sciences
University of Connecticut - Avery Point
1080 Shennecossett Road
Groton, CT 06340 USA
Email: ann.bucklin@uconn.edu

K.T.C.A. (Katja) Peijnenburg (Proposed Vice-Chair)
Senior Researcher - Marine Biodiversity
Associate Professor at the University of Amsterdam
Institute for Biodiversity and Ecosystem Dynamics (IBED)
Naturalis Biodiversity Center
Leiden, The Netherlands
Email: katja.peijnenburg@naturalis.nl

Ksenia Kosobokova (Proposed Vice-Chair)
Leading Scientist, Professor
Institute of Oceanology, Russian Academy of Sciences
36 Nakhimova Avenue
117997 Moscow, Russia
Email: xkosobokova@ocean.ru

April 16, 2018

MetaZooGene: Toward a new global view of marine zooplankton biodiversity based on DNA metabarcoding and reference DNA sequence databases

Abstract

Marine zooplankton biodiversity remains a significant unknown throughout the global ocean. Molecular approaches, including DNA barcoding (use of short DNA sequences for species recognition and discrimination) and metabarcoding (large-scale taxonomic identification of complex samples via analysis of one or few orthologous DNA barcode regions), are expected to significantly revise global estimates of zooplankton diversity. Primary obstacles to broader use of metabarcoding for fundamental research on the biodiversity, biogeography, and trophic dynamics of marine zooplankton include lack of best practices for barcoding and metabarcoding analysis, and lack of taxonomically-comprehensive global-scale DNA sequence reference databases (libraries of DNA sequences determined for identified specimens) for the selected barcode gene regions. The MetaZooGene SCOR Working Group will develop a global vision for the continuing development of integrative molecular – morphological taxonomic analysis of marine zooplankton biodiversity through international communication, cooperation, and collaboration. Terms of Reference include: 1) create an open-access web portal for DNA barcodes for marine zooplankton; 2) design an optimal DNA barcoding pipeline for marine zooplankton; and 3) develop best practices for DNA metabarcoding of marine zooplankton biodiversity. Deliverables include a project website, web portal, and database; dissemination of best practices for DNA barcoding and metabarcoding of marine zooplankton; and publications and review papers in open access peer-reviewed journals. Workshops to build capacity for the MetaZooGene mission will be designed for early career scientists and those from developing countries, and will be held in association with international conferences.

Scientific Background and Rationale

Marine zooplankton are key players in pelagic food webs, central links in ecosystem function, and useful indicators as rapid responders to environmental variation and climate change (Beaugrand et al., 2010). Characterization of biodiversity of the marine zooplankton assemblage is complicated by many factors, including the systematic complexity of the assemblage; presence of cryptic, rare, and novel species; and high local-to-global ratios of species diversity (Snelgrove et al., 2016). Molecular approaches, including DNA barcoding (use of short DNA sequences for species recognition and discrimination) and metabarcoding (large-scale taxonomic identification of complex samples via analysis of one or few orthologous DNA barcode regions), are providing important new insights into the ‘hidden diversity’ of marine zooplankton (Bucklin et al., 2011, 2016; Lindeque et al. 2013). A compelling question driving development of new genetic and genomic tools for characterization of zooplankton biodiversity is that the number of species occurring in the pelagic realm remains unknown (Mora et al., 2013; Chust et al., 2017). Metabarcoding is expected to increase estimates marine zooplankton biodiversity dramatically, although recent studies differ in these estimations by at least several orders of magnitude (e.g., Brown et al., 2015; de Vargas et al., 2015; Sommer et al., 2017).

The MetaZooGene SCOR Working Group will seek to develop a global vision for the continuing development of integrative molecular – morphological taxonomic analysis of marine zooplankton biodiversity through international communication, cooperation, and collaboration. The overarching goal will be to promote and facilitate development and use of best-practices for molecular, statistical and bioinformatics approaches for DNA barcoding and metabarcoding, and thereby to accelerate progress in characterizing local-to-global scale patterns of biodiversity and biogeographic distributions of marine zooplankton, and contribute to better understanding of the functioning of pelagic ecosystems.

Metabarcoding relies on high-throughput DNA sequencing (HTS) technologies, which yield millions of DNA sequences in parallel and allow large-scale analysis of environmental samples. A number of different gene regions are used for metabarcoding, including several hypervariable regions of nuclear

small- (18S) and large-subunit (28S) rRNA (Lindeque et al., 2013; De Vargas et al., 2015; Hirai et al., 2015). Metabarcoding studies have also used mitochondrial cytochrome oxidase I (LeRay and Knowlton, 2016), which can detect species-level diversity, but may yield inconsistent results. In addition to fundamental research on pelagic biodiversity and ecosystem functioning, promising applications of metabarcoding of marine zooplankton include rapid detection of impacts of climate change, monitoring and assessment of ecosystem health, characterization of food webs, and detection of introduced and non-indigenous species (Aylagas et al., 2014; Bucklin et al. 2016; Deagle et al., 2017; Goodwin et al., 2017).

A primary obstacle to the widespread use of metabarcoding for fundamental research on the biodiversity, biogeography, and trophic dynamics of marine zooplankton is the lack of taxonomically-comprehensive global-scale DNA sequence reference databases (libraries of DNA sequences determined for specimens identified to species by morphological taxonomic characters) for the selected barcode gene regions. Reference DNA databases have been shown to markedly improve the accuracy and taxonomic resolution of metabarcoding data for zooplankton diversity (Hirai et al, 2015; Machida et al., 2017; Yang et al., 2017). Among the challenges remaining for reliable and routine application of metabarcoding for analysis of zooplankton diversity are evaluation and inter-comparison of results using different barcode gene regions; development of best practices to ensure accurate identification, discrimination, and detection of taxa; and continued development of taxonomically comprehensive reference databases for all barcode gene regions.

Metabarcoding is revolutionizing the analysis of marine biodiversity and is expected to significantly revise global estimates of zooplankton diversity. Yet the remarkable promise of DNA barcoding and metabarcoding will only be realized with global-scale conversation, cooperation, and collaboration among scientists devoted to the integration of morphological and molecular taxonomic approaches. The proposed MetaZooGene WG members share a dedication to this shared goal, and will work together to promote and facilitate the broad application of DNA barcoding and metabarcoding for analysis of zooplankton diversity, to allow more rapid detection and description of the impacts of climate change, and provide a new foundation for future research, monitoring and management of the pelagic realm.

Terms of Reference

1) Create an open-access web portal for DNA barcodes for marine zooplankton

An online open-access database will be designed and populated with species and specimen metadata, photographs, and deep links to sequence data deposited in major repositories. All molecular data, metadata, and protocol information will be publicly available via one or more of the existing open-access repositories. The portal will facilitate and expedite searches for data for marine zooplankton species, which are difficult to discriminate and retrieve via usual keyword searches of online repositories.

2) Design an optimal DNA barcoding pipeline for marine zooplankton

MetaZooGene WG members will develop and recommend best practices for DNA barcoding, including evaluation of all steps necessary for the bug-to-sequence pipeline. The WG will coordinate and accelerate the augmentation of taxonomically-comprehensive DNA barcode datasets for the global ocean by setting priorities and identifying gaps.

3) Develop best practices for DNA metabarcoding of marine zooplankton biodiversity

The MetaZooGene WG will evaluate and compare the many diverse molecular, analytical, and bioinformatics approaches now used for metabarcoding. A globally-integrated metabarcoding dataset will be developed to allow comparison of results from local, regional, and global scale efforts by all participating investigators. This shared resource and the ensuing discussions and deliberations will provide the basis for high-level synthetic and review papers that recommend and promote best practices for metabarcoding of marine zooplankton biodiversity.

Working Plan

The MetaZooGene web portal and database (ToR #1) will be designed and populated with data and information (including unique identifiers) for samples, specimens and vouchers. The database will include deep links via accession numbers or other unique identifiers for DNA sequence data available on established public open access repositories. Only repositories that assure fully open and public access to all data by any user – without password control, registration, or other impediments to open access – will be included in the MetaZooGene database and web portal.

The MetaZooGene WG members will develop and recommend a sequence and set of procedures for barcoding pipelines (ToR #2), designed to guide related efforts in laboratories or facilities throughout the world. Description of best practices for DNA barcoding will include morphological identification of species by taxonomic experts, photographs of specimens, archives of specimen and DNA vouchers, DNA sequencing primers and protocols for all agreed-upon barcode gene regions, inclusion of all specimen data and metadata in open access websites and portals, and submission of all DNA sequence data to existing open access data repositories. Information and advice will be sought from several WG members who are currently engaged in bug-to-database multi-gene semi-automated DNA barcoding pipelines at their home universities and institutes.

MetaZooGene WG members will develop recommendations for best practices for metabarcoding of marine zooplankton biodiversity (ToR #3), based on examination, review, comparison and evaluation of key issues, including selection of marker gene regions, design of PCR primers and sequencing protocols, comparisons of analytical and bioinformatics parameters and approaches, and benefits of taxonomically comprehensive global-scale reference DNA sequence databases. Particular attention will be given to ongoing programs that have implemented metabarcoding for routine local-to-regional monitoring and assessment of zooplankton biodiversity.

The MetaZooGene WG will meet three times over the 4 years in association with international scientific conferences (see *Timeline*). Four MetaZooGene workshops will be organized (one each year) to provide in-depth training for graduate students and early-career scientists, especially from developing countries (see *Capacity Building*).

Timeline

Year 1 (January 1 – December 31, 2019)

- Create MetaZooGene database and web portal; populate the database with all available specimen data and metadata (including unique identifiers), and deep links to DNA sequences in existing open-access repositories.
- Develop and describe best practices for multi-gene DNA barcoding “pipelines” from bug-to-database, including sample collection, preservation and archiving; specimen identification and vouchering; DNA extraction, PCR and sequencing primers and protocols; submission of data and metadata to open-access repositories.
- Convene first MetaZooGene Working Group meeting following the ICES 2019 Annual Science Conference; September 9-12, 2019; Gothenburg, Sweden.
- Organize a MetaZooGene Workshop to review and compare present approaches and best practices for multi-gene DNA barcoding pipelines from bug-to-database. The workshop will be held following the International Council for the Exploration of the Sea (ICES) 2019 Annual Science Conference; September 9-12, 2019; Gothenburg, Sweden.

Year 2 (January 1 – December 31, 2020)

- Conduct comparative examination of metabarcoding approaches and protocols for analysis of marine zooplankton biodiversity.
- Convene the second MetaZooGene Working Group meeting following AGU/ASLO/TOS Ocean Sciences Meeting; February 16-21, 2020; San Diego, California, USA.
- Organize a workshop to evaluate and compare use of DNA sequence databases for different barcode gene regions for analysis of zooplankton diversity based on DNA barcoding and metabarcoding. The workshop will be held following the AGU/ASLO/TOS Ocean Sciences Meeting; February 16-21, 2020; San Diego, California, USA
- Publish comprehensive review paper in open access peer-reviewed journal summarizing and assessing the current view of marine zooplankton biodiversity based on metabarcoding.

Year 3 (January 1 – December 31, 2021)

- Organize an intensive hands-on ‘DNA-to-data’ training workshop in DNA barcoding and metabarcoding targeted for graduate students and early-career scientists. The workshop will be held at Academia Sinica (Taipei, Taiwan) and hosted by WG member Ryuji Machida.
- Publish comprehensive review paper in open access peer-reviewed journal summarizing current status of multi-gene reference DNA sequence database, with demonstration of the usefulness of reference databases for improved identification of taxa from metabarcoding analysis of zooplankton diversity.
- Fully populate the MetaZooGene database, with specimen data and metadata for described species of marine holozooplankton, including DNA sequences for multiple barcode gene regions and multiple specimens throughout each species’ geographic range.

Year 4 (January 1 – December 31, 2022)

- Convene the third MetaZooGene Working Group meeting following the AGU/ASLO/TOS Ocean Sciences Meeting; February 27-March 4, 2020; Honolulu, Hawaii, USA
- Organize a workshop to educate and encourage use of best practices, including use of taxonomically-comprehensive global-scale reference DNA sequence databases, for metabarcoding analysis of marine zooplankton biodiversity. The workshop will be scheduled to follow the AGU/ASLO/TOS Ocean Sciences Meeting; February 27-March 4, 2020; Honolulu, Hawaii, USA
- Publish and distribute a comprehensive recommendation for best practices for metabarcoding of marine zooplankton biodiversity, including use of taxonomically comprehensive global-scale reference DNA sequence databases.

Deliverables

Project website, web portal, and database: The MetaZooGene WG will have a dedicated website and web portal for display and distribution of project updates and results, as well as a login section for exchange of information among WG Full and Associate members. Database design, development and maintenance will be the responsibility of WG member Todd O’Brian, with costs provided by the NOAA National Marine Fisheries Service. All data and metadata associated with the MetaZooGene WG efforts will be publicly available and open access, without password control or required registration.

Best practices for DNA barcoding and metabarcoding: The MetaZooGene WG members will produce recommendations for best practices for both DNA barcoding pipelines and metabarcoding analysis of marine zooplankton diversity.

Publications in peer-reviewed journals: MetaZooGene WG members will publish and acknowledge WG contributions for papers in peer-reviewed open access journals on topics relevant and related to the WG mission. The anticipated goal is a total of 25 publications over the active years of WG effort.

Review papers: Three comprehensive review papers will be co-authored by WG members and published in peer-reviewed open access journals. The reviews will provide summary evaluation and analysis consistent with the MetaZoogene primary goals, including: 1) overview of available DNA sequence reference DNA databases for marine zooplankton; 2) evaluation of best practices for “bug-to-sequence” pipelines for production of DNA sequences for barcode marker genes; and 3) evaluation and recommendation of best practices for metabarcoding analysis of marine zooplankton biodiversity.

Capacity Building

Four workshops are planned to address the key themes central to the MetaZooGene mission. Workshops will be designed for early career scientists and those from developing countries, and will be held in association with international oceanographic conferences. Announcements and invitations will seek to recruit participation by the target groups. The workshops are:

- Best practices for multi-gene DNA barcoding pipelines from bug-to-database. Will be held following ICES 2019 Annual Science Conference; September 9-12, 2019; Gothenburg, Sweden
- Use of DNA reference sequence databases for analysis of zooplankton diversity based on DNA barcoding and metabarcoding. Will be held following the AGU/ASLO/TOS Ocean Sciences Meeting; February 16-21, 2020; San Diego, California, USA.
- Hands-on ‘DNA-to-data’ training workshop in DNA barcoding and metabarcoding. Will be held during Fall, 2021 at Academia Sinica (Taipei, Taiwan) and hosted by WG member Ryuji Machida.
- Best practices for metabarcoding analysis of marine zooplankton biodiversity. Will be held following the AGU/ASLO/TOS Ocean Sciences Meeting; February 27-March 4, 2020; Honolulu, Hawaii, USA

The MetaZooGene workshop participants will have top priority for invitations for collaborative research visits to WG members’ laboratories. A mailing list will be maintained and used for frequent announcements and updates on the activities of the WG. The workshop participants will be used for invitations to special sessions at international conferences and symposia on relevant research topics. The goal will be to ensure significant and continued influence and impact on this group of early career scientists.

Key References

Aylagas, E, Borja A, Rodríguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). *PLoS One* 9:e90529.

Beaugrand G, Edwards M, Legendre L (2010) Marine biodiversity, ecosystem functioning, and carbon cycles. *Proc. Nat. Acad. Sci.* 107:10120-10127.

Brown EA, Chain FJJ, Crease TJ, MacIsaac HJ, Cristescu ME (2015) Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecol. Evol.* 5:2234–2251.

- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of marine Metazoa. *Annu. Rev. Mar. Sci.* 3:471-508.
- Bucklin A, PK Lindeque, N Rodriguez-Ezpeleta, A Albaina, M Lehtiniemi (2016) Metabarcoding of marine zooplankton: Progress, prospects and pitfalls. *J. Plankton Res.* 38:393-400.
[doi:10.1093/plankt/fbw023](https://doi.org/10.1093/plankt/fbw023)
- Bucklin A, K DiVito, I Smolina, M Choquet, JM Questel, et al. (2018) Population Genomics of Marine Zooplankton. In: *Population Genomics: Marine Organisms*, OP Rajora and M Oleksiak (Eds.), Springer Cham. DOI https://doi.org/10.1007/13836_2017_9
- Chust G, Vogt M, Benedetti F, Nakov T, Villéger S, et al. (2017) Mare incognitum: a glimpse into future plankton diversity and ecology research. *Front. Mar. Sci.* 4:68. doi: 10.3389/fmars.2017.00068
- Deagle BE, LJ Clarke, JA Kitchener, AM Polanowski, AT Davidson (2017) Genetic monitoring of open ocean biodiversity: an evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources* 2017:1-16
- DeVargas C, Audic S, Henry N, Decelle J, Mahé F, et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605.
- GIGA (2014) The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrate Genomes. *J Hered* 105:1–18. doi: 10.1093/jhered/est084
- Goodwin KD, LR Thompson, B Duarte, T Kahlke, AR Thompson, et al. (2017) DNA sequencing as a tool to monitor marine ecological status. *Front. Mar. Sci.* 4:107. doi: 10.3389/fmars.2017.00107
- Hirai J, Tsuda A (2015) Metagenetic community analysis of epipelagic planktonic copepods in the tropical and subtropical Pacific. *Mar. Ecol. Progr. Ser.* 534:65–78.
- Leray M, Knowlton N (2016) Censusing marine eukaryotic diversity in the twenty-first century. *Phil. Trans. R. Soc. B* 371:20150331. <http://dx.doi.org/10.1098/rstb.2015.0331>
- Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next Generation Sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS One* 8:e81327.
- Machida RJ, Leray M, Ho SL, Knowlton N (2017) Data descriptor: Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Sci. Data* 4:170027.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? *PLOS Biology* 9:e1001127. <https://doi.org/10.1371/journal.pbio.1001127>
- Snelgrove P, E Vanden Berghe, P Miloslavich, N Bailly, A Brandt, A Bucklin, et al. (2016) Chapter 34. Global Patterns in Marine Biodiversity. In: *United Nations World Ocean Assessment I*. http://www.un.org/depts/los/global_reporting/WOA_RegProcess.htm
- Sommer SA, L Van Woudenberg, PH Lenz, G Cepeda, E Goetze (2017) Vertical gradients in species richness and community composition across the twilight zone in the North Pacific Subtropical Gyre. *Molecular Ecology* 2017:1-21.
- Yang J, Zhang X, Zhang W, Sun J, Xie Y, et al. (2017) Indigenous species barcode database improves the identification of zooplankton. *PLoS ONE* 12:e0185697. <https://doi.org/10.1371/journal.pone.0185697>