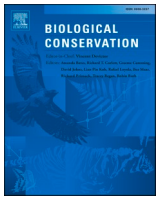
Biological Conservation 280 (2023) 109963

Contents lists available at ScienceDirect

Biological Conservation

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



Prioritizing taxa for genetic reference database development to advance inland water conservation

Marie-Eve Monchamp [a,b,*], Zofia E. Taranu [b,c], Rebecca E. Garner [b,d], Tessa Rehill [a], Olivier Morissette [b,e], Lars L. Iversen [a,b], Vincent Fugère [b,f], Joanne E. Littlefair [g], Naíla Barbosa da Costa [b,h], Jessica E. Desforges [i], Joe R. Sánchez Schacht [a,j], Alison M. Derry [b,k], Steven J. Cooke [i], Rowan D.H. Barrett [a,l], David A. Walsh [b,d], Jiannis Ragoussis [m], Monique Albert [n], Melania E. Cristescu [a,b], Irene Gregory-Eaves [a,b]

[a] Department of Biology, McGill University, Montreal, Québec, Canada
[b] Groupe de Recherche Interuniversitaire en Limnologie (GRIL), Montréal, Québec, Canada
[c] Aquatic Contaminants Research Division, Environment and Climate Change Canada, Montréal, Québec, Canada
[d] Department of Biology, Concordia University, Montreal, Québec, Canada
[e] Département des sciences fondamentales, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada
[f] Département des sciences de l'environnement, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada
[g] School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom
[h] Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada
[i] Department of Biology, Carleton University, Ottawa, Ontario, Canada
[j] Department of Natural Resource Sciences, McGill University, Sainte-Anne-de-Bellevue, Québec, Canada
[k] Département des sciences biologiques, Université du Québec à Montréal, Montréal, Québec, Canada
[l] Redpath Museum, McGill University, Montreal, Quebec, Canada
[m] McGill Genome Centre, Victor Philipp Dahdaleh Institute of Genomic Medecine, McGill University, Montréal, Québec, Canada
[n] University of Guelph, Centre for Biodiversity Genomics, Guelph, Ontario, Canada

ARTICLE INFO

Keywords:
Biodiversity
Freshwater
DNA barcodes
Genetic diversity
IUCN red list
COSEWIC
Genome database

ABSTRACT

Biodiversity loss has accelerated over the past century and freshwater species overall are among those experiencing greatest declines. Genetic resources have the potential to help evaluate the full magnitude of this loss and represent a key tool to effectively allocate conservation resources and monitor the success of restoration efforts. The full power of genetic resources will be realized when the daunting task of referencing all DNA sequences of freshwater organisms is complete. Here, we quantified the availability and distribution of barcode and genome data for freshwater macroscopic organisms in Canada, a country rich in inland water resources and thus particularly vulnerable to aquatic species losses. Impressively, most inland water species (86 %) were represented by barcodes recorded in the BOLD Systems database, while very few had full genomes available (<4 %) in the NCBI database. We identified barcode data deficiencies in northern regions and for taxa assessed as most at risk or without sufficient information for conservation status classification. As expected, the speciose insect group had a lower-than-average number of records per species and a high proportion of data deficient species without adequate barcode coverage. This study highlights where future sequencing resources should be prioritized within initiatives such as the Canada BioGenome Project and BIOSCAN Canada and provides a workflow that could be applied internationally to inform conservation management plans and to mitigate biodiversity loss.

1. Introduction

Freshwater ecosystems hold approximately 10 % of the world's described species. However, the deterioration of freshwater habitats by multiple stressors poses accelerating threats to freshwater biodiversity (Dudgeon et al., 2006; Dudgeon, 2019; Reid et al., 2019) and the

* Corresponding author at: Department of Biology, McGill University, Stewart Biology Building, 1205 Docteur Penfield, Room: N4/18, Montreal, Quebec H3A 1B1, Canada.
E-mail address: me.monchamp@gmail.com (M.-E. Monchamp).

https://doi.org/10.1016/j.biocon.2023.109963
Received 7 December 2022; Received in revised form 31 January 2023; Accepted 11 February 2023

irreplaceable services provided by these ecosystems (Albert et al., 2021). The highest rates of extinction and population decline among aquatic taxa are occurring in freshwater species (IPBES, 2019; Young et al., 2016) and global freshwater biodiversity loss has been characterized as a crisis (Harrison et al., 2018; Tickner et al., 2020). Improved monitoring methods were raised as a key priority by a panel of global freshwater biodiversity researchers to overcome current taxonomic limitations and inefficiencies in monitoring (Maasri et al., 2022). This is particularly important within Canada, a country which holds 20 % of the global freshwater volume (Messager et al., 2016). Notably, a study by Desforges et al. (2022) reporting the lack of knowledge to assess the state of freshwater biodiversity in Canada concluded that there were insufficient data to assess the conservation status for 38 % of species dependent on freshwater habitats for at least part of their lifecycle. Data deficiency was particularly high in invertebrate groups that serve as bioindicators of ecological integrity such as Trichoptera (caddisflies; 68 % of species are data deficient) and Ephemeroptera (mayflies; 78 % data deficient). Additionally, 12 % of freshwater species were assessed as'At Risk', with the greatest proportions of at-risk species appearing in reptiles (50 %), amphibians (31 %), and birds (20 %). A further 18 % of all species were listed as 'Special Concern' (Desforges et al., 2022).

In general, the extent of declines in freshwater biodiversity is not fully known as most ecosystems and taxa are poorly monitored (Díaz et al., 2019). Although open-access data repositories like the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) are improving our understanding of biodiversity distributions and dynamics, many gaps remain. Genetic resources are poised to address many conservation questions, but a comparison of the conservation status of a diversity of species with the availability of reference barcodes and genomes is lacking. Networks aimed at growing national and international initiatives to characterize biodiversity on Earth are contributing to a rapid and continuous expansion of reference genomes (e.g., the Earth Biogenome Project (Lewin et al., 2018), Genome 10K (Haussler et al., 2009)) and DNA barcodes (e.g., the Barcode of Life Data System (BOLD) (Ratnasingham and Hebert, 2007) and International Barcode of Life (IBOL) (https://ibol.org/)). Numerous independent barcoding campaigns are also targeting taxon- and region-specific flora and fauna (Weigand et al., 2019). These databases can provide more complete knowledge on species occurrences to help assess their status and work towards their management and restoration. For example, of 12.5 million individual specimens collected by the Canadian Aquatic Biomonitoring Network (CABIN), only 57 % could be identified to genus level based on morphology, but extensive sequencing efforts of these specimens are underway to help improve identifications (Curry et al., 2018).

Developments in applications of environmental (e)DNA (defined as DNA directly isolated from environmental samples) offer promising avenues for the biomonitoring of natural ecosystems to inform conservation (Thomsen and Willerslev, 2015). Meta-analyses have shown that eDNA approaches meet or exceed detection capabilities of conventional techniques for monitoring aquatic ecosystems (Keck et al., 2022; Mcelroy et al., 2020). eDNA can be used to survey a broad diversity of organisms, including elusive, rare, cryptic, or invasive species, which may evade conventional capture (Berry et al., 2021; Balasingham et al., 2018; Deiner et al., 2017; Jerde et al., 2011). Moreover, eDNA surveys can help to fill gaps in conservation status assessments by allowing researchers to track species distributions across bro. Finally, in places where conventional biomonitoring requires extensive resources, such as in freshwater-rich regions, eDNA surveys can be an option as they can be conducted across large spatial scales in a consistent way (Bush et al., 2019). While exhaustive and well curated reference sequence data are essential for biodiversity monitoring through eDNA, and existing databases have made massive progress since their inception a mere 25 years ago, there remain key taxonomic discrepancies to be addressed for conservation and management research (Hotaling et al., 2021). Given the urgency of the biodiversity crisis, it is prudent to identify and prioritize these gaps to help mitigate current and future biodiversity declines.

Our main study objectives were to 1) identify the availability of genetic records (i.e., annotated genomes or barcodes) for freshwater species in Canada and identify zones of record richness and deficiency; 2) determine the taxonomic groups exhibiting the greatest data deficiencies (i.e., taxa to focus future sequencing efforts on); and 3) evaluate whether sufficient genetic reference data exist for most 'At Risk' freshwater species in Canada (where At Risk is defined as taxa listed as 'Threatened', 'Endangered', or 'Extirpated'; see definitions in Desforges et al. (2022)). We compiled barcode and whole genome records from two major public repositories, the Barcode of Life Data System (BOLD, https://www.boldsystems.org/) for barcodes and the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) database for whole genome assemblies, and investigated the taxonomic distribution of DNA sequence records in relation to species conservation status, species ranges, and human footprint index. We first present the distribution of species barcodes and genome assemblies restricted to Canadian inland water taxa but mapped across a global footprint index. To do so, we used the list of freshwater-reliant taxa compiled by the Desforges et al. (2022) study, comprising six major taxonomic groups with Canadian species conservation status, and mapped the availability of barcodes and full genome assemblies for Canadian species at risk or with data deficiencies. We then provide a more in-depth study of Canadian/Canada-centric species in terms of their representation within the BOLD repository, the spatial extent of the genetic records relative to species ranges, and whether biases exist in terms of conservation status or geography (latitudinal and longitudinal gradients).

2. Material and methods

Our approach consisted of five main steps which span the compilation and analysis of genetic resources available for freshwater-reliant taxa in Canadian waters to identifying priority taxa for future sequencing efforts (see Fig. 1).

2.1. Canada's freshwater species list

Desforges et al. (2022) compiled a list of Canada's native freshwater-dependent taxa (defined loosely as habitats that spanned a gradient from fresh to brackish waters). Briefly, species data were collected for the following macroscopic taxonomic groups: plants, invertebrates, fishes, herpetofauna (amphibians and reptiles), birds, and mammals. Species were included if they accomplished all or part of their lifecycle in or on inland waters, or if the species showed dependency on inland waters for food or habitat. The species lists were compiled using data from the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), the Canadian Endangered Species Conservation Council (CESCC) 2015 Wildlife Species Report and the International Union for Conservation of Nature (IUCN) Red List. The final species list resulted in the identification of over 3000 unique taxa as species, subspecies, or geographically distinct populations (Desforges et al., 2022). COSEWIC was used as the primary source for conservation status data rather than global assessments due to the Canadian-specific assessment process. When species were not assessed by COSEWIC, CESCC data were used to determine their status (see Desforges et al. (2022) for details). These statuses were grouped in increasing degree of risk as follows: 'Not Available', 'Data Deficient', 'Not at Risk', 'Special Concern', 'Threatened', 'Endangered', 'Extirpated', and 'Extinct'; 'Not Available' was assigned to taxa that were not considered by COSEWIC or CESCC for conservation assessment and 'At Risk' was used to describe taxa with conservation statuses of 'Threatened' or worse.

For taxa that were assigned more than one conservation status in COSEWIC due to different management units, conservation statuses were scored (from 0-'Not Available' to 8-'Extinct') and the mean score was used to determine the status for the entire species. For taxa

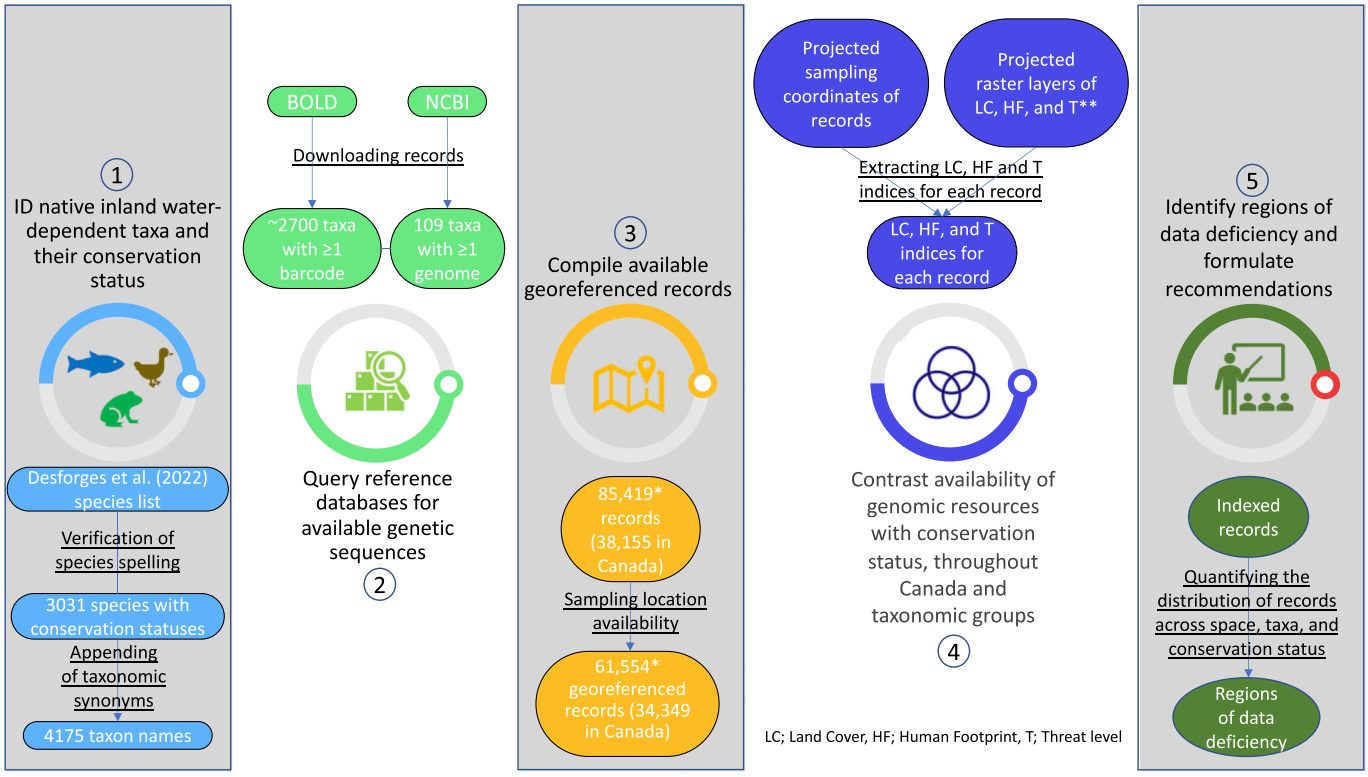


Fig. 1. General workflow of the approach used for the species list, the query of reference databases, the compilation and the projection of available georeferenced records over the global and Canadian human footprint index, and the identification of key regions of data deficiencies and taxonomic groups most underrepresented in BOLD and in the NCBI genome database. The colored ovals with white text represent data and the underlined black text represents data manipulations. The final data produced by each step were used as the initial input for the subsequent step. *Only georeferenced records with an assigned longitude are included. **LC = Land cover, HF = Human footprint, T = Threat level.

identified to the subspecies level, the conservation status reflected this subspecies only. For example, *Esox americanus vermiculatus* (Grass pickerel) was identified as being of 'Special Concern' whereas *Esox americanus americanus* (Redfin pickerel) is not at risk. These subspecies both occur in Canada, though they have diverging ranges, which has led to genetically distinct populations. In contrast, a mean conservation status was assigned for *Oncorhynchus nerka* (Sockeye salmon) since COSEWIC assesses different populations that are not sufficiently distinct to define specific subspecies. Overall, 12 % of all identified taxa were found to be 'At Risk', 18 % identified as 'Special Concern', and 38 % lacked sufficient data to enable their assessment (i.e., categorized 'Data deficient') (Desforges et al., 2022) (Online Appendix Table A1).

2.2. Harmonization of Canada's freshwater species list with reference databases

The list of taxa compiled by Desforges et al. (2022) was downloaded on March 30, 2022. In the case of duplicate entries with conflicting conservation status assessments (the case for only two taxa), the most critical status was selected to avoid potentially underestimating a species' conservation status. Following basic text cleaning of species names in R (i.e., removing non-ASCII characters and leading/trailing white-space), species spellings were verified using the gnr_resolve() function from the package taxize (Chamberlain and Szöcs, 2013) against taxon names indexed in the Integrated Taxonomic Information System (ITIS) and Global Biodiversity Information Facility (GBIF). Species flagged as misspelled were manually corrected based on the primary spelling recorded in ITIS or GBIF. To cast a wide net in our search for genetic/genomic records, we appended taxonomic synonyms where available using the ns_search_spp() function from the R package natserv (Chamberlain, 2020). Taxonomy upstream of genus was accessed from ITIS, or GBIF if valid taxonomy was not available in ITIS, using the classification () function in taxize; only family, order, and class assignments were retained in addition to genus, species and subspecies, where available.

2.3. Compilation of publicly available barcode and annotated genome records

Specimen and sequence records were downloaded from BOLD Systems using the bold_seqspec() function from the R package bold (Chamberlain, 2021) on June 18, 2022. In cases where subspecies were included by Desforges et al. (2022), records were queried at both the species and subspecies ranks. To determine whether taxa without programmatically accessed records were truly missing from BOLD Systems (as opposed to being the subject of program interruption or failure, for example), taxa without records were manually verified online at https://www.boldsystems.org/. In some cases (0.6 % of total records), we detected identical BOLD records for different species; sometimes this occurred when querying taxa with subspecies information or taxa with only genus-level records in BOLD. Redundant BOLD records were retained and factored into record counts for taxonomic groups, but unique record counts were also reported to assess frequencies of DNA markers and sequence lengths. Unique BOLD records were identified by their process identification codes (processid).

Basic genome records were downloaded from the NCBI Genome database on June 18, 2022 using the entrez_search() function, specifying only the organism search parameter, from the R package rentrez (Winter, 2017). Unique NCBI Genome records were filtered by their Entrez Unique Identifiers (UIDs).

2.4. Human footprint index

The human footprint index was compiled using eight anthropogenic pressures: (1) extent of built environments; (2) crop land; (3) pasture land; (4) human population density; (5) night-time lights; (6) railways; (7) roads; and, (8) navigable waterways. Following (Sanderson et al., 2002), each metric was scaled between 0 and 10, weighted to its pressure importance and summed. Canada's human footprint index was calculated following the same protocol, considering 12 pressures which included the global footprint and additional pressures: (9) presence of dams and reservoirs; (10) mining activity; (11) oil and gas; and, (12) forestry (Hirsh-Pearson et al., 2022).

2.5. Data analyses

BOLD records for Canadian inland water macroorganisms that had associated georeferenced data were projected (WGS 1984) over the global human footprint index 2009 (Venter et al., 2016) and the Canadian human footprint index (Hirsh-Pearson et al., 2022); note: not all BOLD records had georeferenced data. For every projected BOLD record, the value of the human footprint index was extracted using the sampling coordinates and the function extract() from the terra package in R (version 4.1.2; R Core Team, 2020) (Hijmans, 2023). Average values (±SD) of human footprint index were summarized by phylum (global mapping) and COSEWIC assessments (Canadian mapping). We tested the link between human footprint indexes and record-related factors (phylum, taxon, and conservation status) by applying linear mixed effect models (LMM) using sampled species as a random effect to account for the non-independence of same species observations.

For each species, the projected BOLD records were used to quantify the density of genetic records as a function of latitude. Using a resolution of 0.1-degree latitude, the number of records per latitudinal band was estimated as the (number of records)/ (maximum latitude - minimum latitude). For species with a single observation, this would be 1/0.1 (number of records/latitude band). Potential taxonomic differences in the range - BOLD density relationship were tested using a two-way ANOVA including observation density as the response variable and latitudinal band midpoint and phylum as explanatory variables, testing both their main effects and interaction.

To compare the density and distribution of taxa that had at least one georeferenced BOLD record (~1900 species) with all georeferenced observations present on GBIF in Canada, we applied a GBIF Occurrence Download from doi:10.15468/dl.a5mbxn on 15 June 2022. The distribution of species records was visualized at 100 km^2 resolution. Areas with a comparatively high and low number of records in BOLD and GBIF were visualized by clustering records into three groupings (low-medium-high) based on equal quantiles (0.33, 0.66, and 1) at a log(x + 1) scale.

3. Results

3.1. Compilation of records from BOLD (barcodes) and NCBI (annotated genomes)

Desforges et al. (2022) listed 3212 taxa, 17 of which were assigned to subspecies or geographically distinct populations. Given that many taxonomic synonyms were identified (1028 synonyms across 754 taxa), we queried a total of 4175 unique taxon names against the BOLD and NCBI Genome databases. Based on queries to BOLD, we identified 2751 taxa (86 %) that had at least one barcode. However, of the 85,419 BOLD records downloaded in total, 84,889 records were unique, indicating that 0.6 % of records (530) were redundant across taxa. Most of the unique records reported the DNA marker information (96 %), and a clear majority were for the 5′ end of the mitochondrial cytochrome *c* oxidase subunit I gene (COI-5P; total of 68,480 records). Other DNA markers were also represented, but at much lower frequencies including the internal transcribed spacer 2 (ITS2; >3 %), ribulose-bisphosphate carboxylase (rbcL; >3 %), and megakaryocyte-associated tyrosine kinase (matK; 3 %) genes. DNA sequences downloaded as BOLD metadata were on average 630 bp in length (range 75–1989 bp).

Most phyla were captured in BOLD, with fishes, plants, birds, amphibians, and reptiles having ≥96 % of their affiliated taxa represented by at least one barcode (Online Appendix Table A2). Taxa assessed as Not At Risk were the best represented group by BOLD records (97 %; 987 of 1020 taxa had at least one BOLD record). From there, taxa decreased in representation as their conservation status increased in severity: taxa of Special Concern (91 %; 513 of 562), Threatened (89 %; 153 of 172), Endangered (85 %; 142 of 168), Extirpated (80 %; 12 of 15), and Extinct (75 %; 3 of 4). Taxa assessed as Data Deficient had the lowest representation in BOLD (72 %; 762 of 1053 taxa had at least one BOLD record).

Arguably, a single barcode is not adequate to capture the potential intraspecific variability present. To explore which taxa had at least some redundancy in barcoding, we also identified which taxa had at least 5 barcodes per species (Fig. 2). Particularly vulnerable taxonomic groups containing high proportions of taxa at risk (or data deficiencies) and with few BOLD records were observed throughout the tree of life (Fig. 2). Podicipediformes (Aves), Squamata (Reptilia), Coleoptera (Arthropoda), Stylommatophora (Mollusca), Dioscoreales (Magnoliophyta), and Lycopodiales (Lycopodiophyta) were among the orders with the highest percentage of species at risk lacking representation in BOLD (Fig. 2; Online Appendix Table A3). As expected, the insects were among the most speciose groups and overall had a sizable number of records. However, even well-known aquatic bioindicator groups like the Ephemeroptera – Plecoptera – Trichoptera (EPT) had lower-than-average numbers of records per species and high proportions of data deficient species without adequate barcode coverage (Fig. 2; Online Appendix Table A3).

Based on queries to the NCBI Genome database, only 115 genomes corresponded to taxa in the Desforges list (<4 %), and of these, 109 records were unique (Fig. 2; Online Appendix A Table A4). Of the unique records, 63 were for complete genome assemblies and 46 for draft assemblies. Mammals were the group with the best representation in NCBI (5 of the 7 taxa listed by Desforges et al. (2022)), followed by reptiles (5 of 16). Fishes were the third best represented group (41 of 204 taxa in Desforges et al. (2022)) and also accounted for the most genome records in NCBI, with Salmoniformes and Perciformes having the most representative genomes. Invertebrates were the group represented by the fewest genome records (<1 %; 14 of 1959 taxa). Low numbers of Extinct and Extirpated taxa were reported in Desforges et al. (2022), but were disproportionately well represented by genomes (Extinct: 1 of 4 taxa; Extirpated: 3 of 15). Taxa designated Endangered, Threatened, of Special Concern, or with assessments Not Available were between 4 and 5 % represented by genomes in NCBI. Taxa designated Data Deficient were the least represented by genomes (0.7 %; 7 of 1053 taxa).

3.2. Sequence distributions in terms of geographic ranges and human footprint

BOLD records for Canadian inland water taxa were distributed throughout North America, Europe, Asia, South America, and Africa (Fig. 3a). The top 20 species found outside of Canada's borders were mostly restricted to insects from the orders Trichoptera and Ephemeroptera (Online Appendix Table A5). Overall, global human footprint index values varied significantly by phyla recorded from BOLD records (LMM, $F_{7,56,039} = 23.11$, $p < 0.001$), with Chordata records acquired from the most impacted regions (highest index). When grouping records by major taxonomic groups (at class level), fishes and reptiles were more commonly associated with a higher human footprint index compared to birds and invertebrates. Limiting the records to those acquired within Canada's borders showed that the COSEWIC assessment was significantly correlated to the Canadian human footprint index (LMM, $F_{7,1854}$

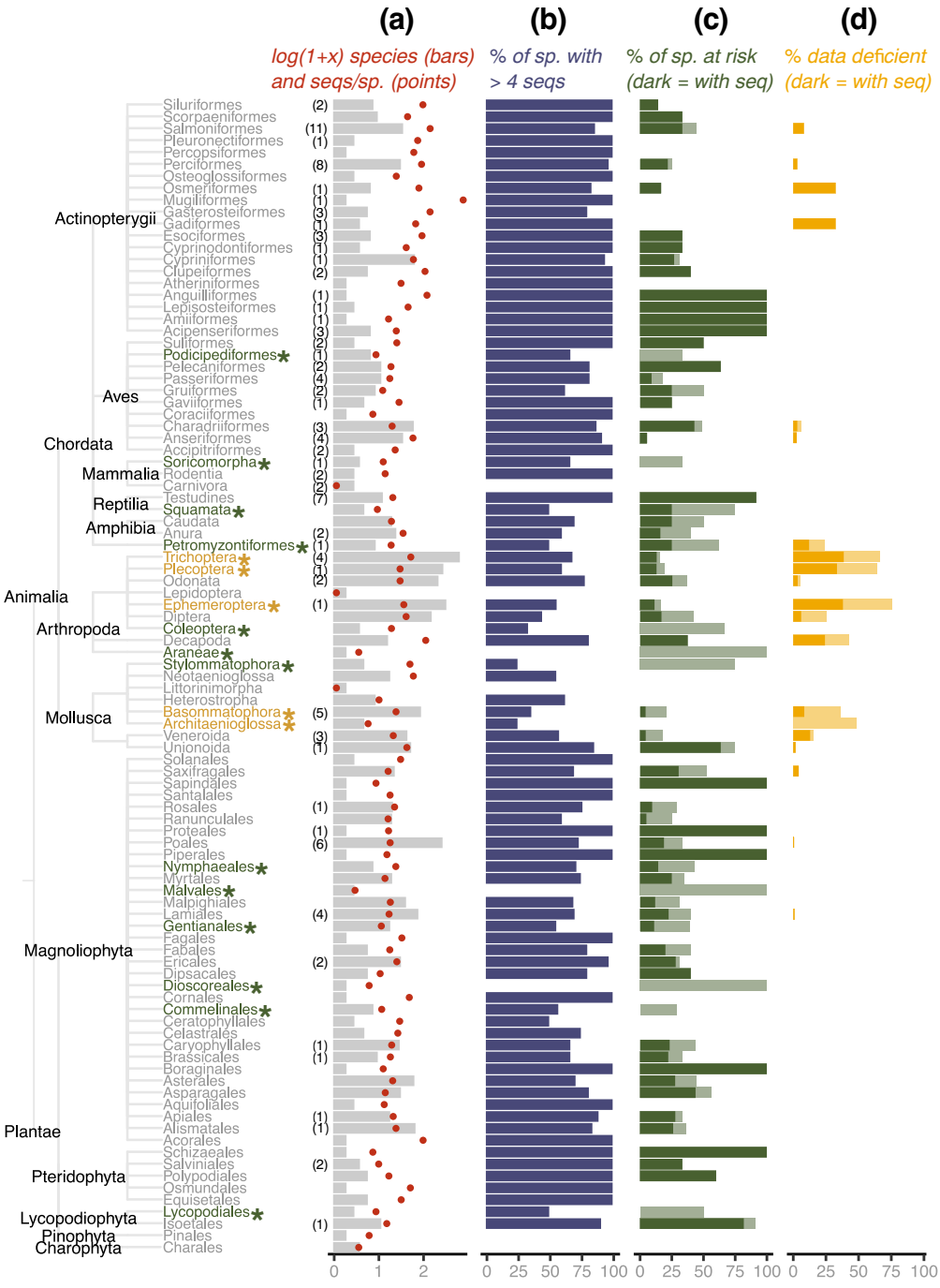


Fig. 2. Diversity, conservation status and availability of genetic and genomic records for various inland water-associated taxa in Canada. The numbers next to the order names refer to the number of whole genomes available in NCBI. (a) Log(1 + x)-transformed number of species (gray bars) and number of sequences per species available in BOLD (red points) for each taxonomic order. (b) % of species within each order that have ≥ 5 sequences in BOLD. (c) % of species at risk (assessed as Special Concern, Endangered, Threatened or Extirpated by COSEWIC) within each order, with darker bars indicating the proportion of those that have ≥5 sequences in BOLD. (d) % of data deficient species within each order with their sequence availability in BOLD. Taxonomic groups that have high proportions of at risk or data deficient taxa with few sequences are highlighted in green and yellow, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

= 8.73, $p < 0.001$), with records of species classified as Endangered and Threatened being associated with the highest footprint index (Fig. 3b). When considering taxa, the same pattern as observed in the global dataset emerged; Canadian fish records were associated with regions with the highest footprint index.

3.3. The spatial distribution of Canadian BOLD records

Species with a range centered in the Canadian North had a lower density of BOLD records compared to those with ranges in the south of the country (Fig. 4a). The decreasing relationship between density of BOLD records and the center of latitudinal range were similar across major taxonomic classes (Phylum × Range Center, $p = 0.50$) (Fig. 4b). Across Canada, there was a clear lack of BOLD data in the northern and central parts of the country. In comparison, areas with high concentrations of GBIF observations extended more to the north compared with the density of BOLD records (Fig. 5).

4. Discussion

Based on our analyses of two large sequence databases, we identified the distribution and taxonomic coverage of genetic resources for macroscopic species dependent on inland water habitats for at least part of their life cycle in a country endowed with vast aquatic resources. Although our findings showed that a large proportion of taxa were represented by at least one barcode, there were geographic gaps and taxonomic orders with fewer sequences per species (i.e., <5 sequences used here as cutoff). Furthermore, taxa whose conservation status could

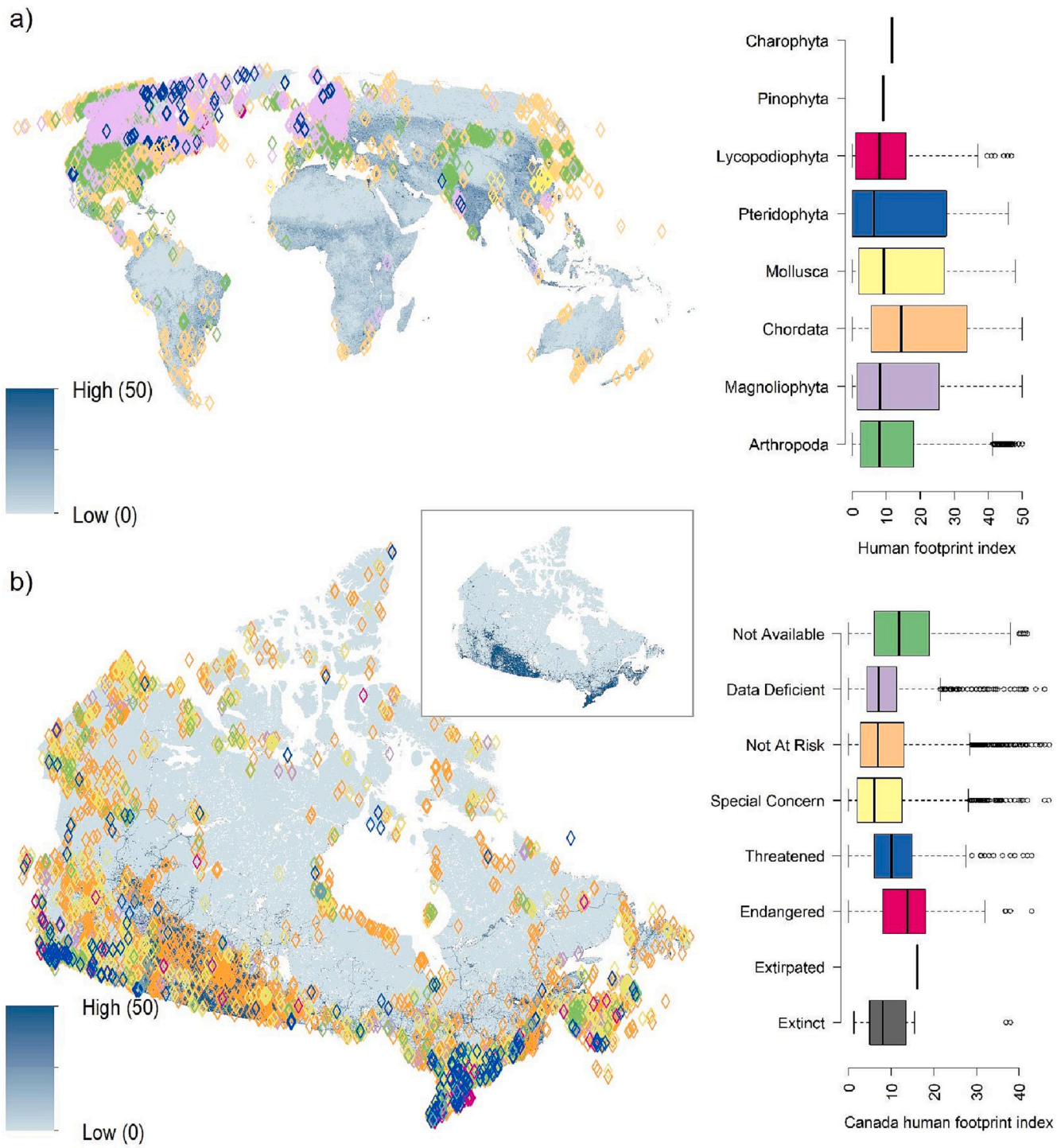


Fig. 3. Distribution of Canadian inland water taxa with BOLD records shown in the context of the human footprint (scale on the right, where a higher index value reflects more intense footprints). (a) Global map of distribution of human footprint index by phylum, with symbols colour coded by taxonomic groups. Inset plot on the right shows values of the global human footprint index across the different taxonomic groups. (b) Map of BOLD record distributions shown in the context of Canada's human footprint index, with symbols colour coded by COSEWIC assessment status. Inset plot on the right shows values of the Canada human footprint index across the conservation status bins. See Online Appendix Table A3 for a list of the top 20 Canadian taxa with barcodes found outside of Canada.

not be assessed due to data deficiency had the least number of genetic resources. Most studies use 5 to 10 individuals per species for the purpose of species delineations and/or to estimate genetic diversity within and between species using DNA barcoding; while these specimen sample sizes have been broadly applied, the required sample sizes for estimating intraspecific variation within species is highly taxon-specific (Phillips et al., 2019). With numerous programs underway to build comprehensive reference databases drawing from both barcoding and whole genome sequencing efforts, our workflow is a valuable output for triaging sequencing projects to prioritize taxa of unknown conservation status or at risk. We also consider the benefits of ensuring that multiple individuals are sequenced to assess intraspecific variability.

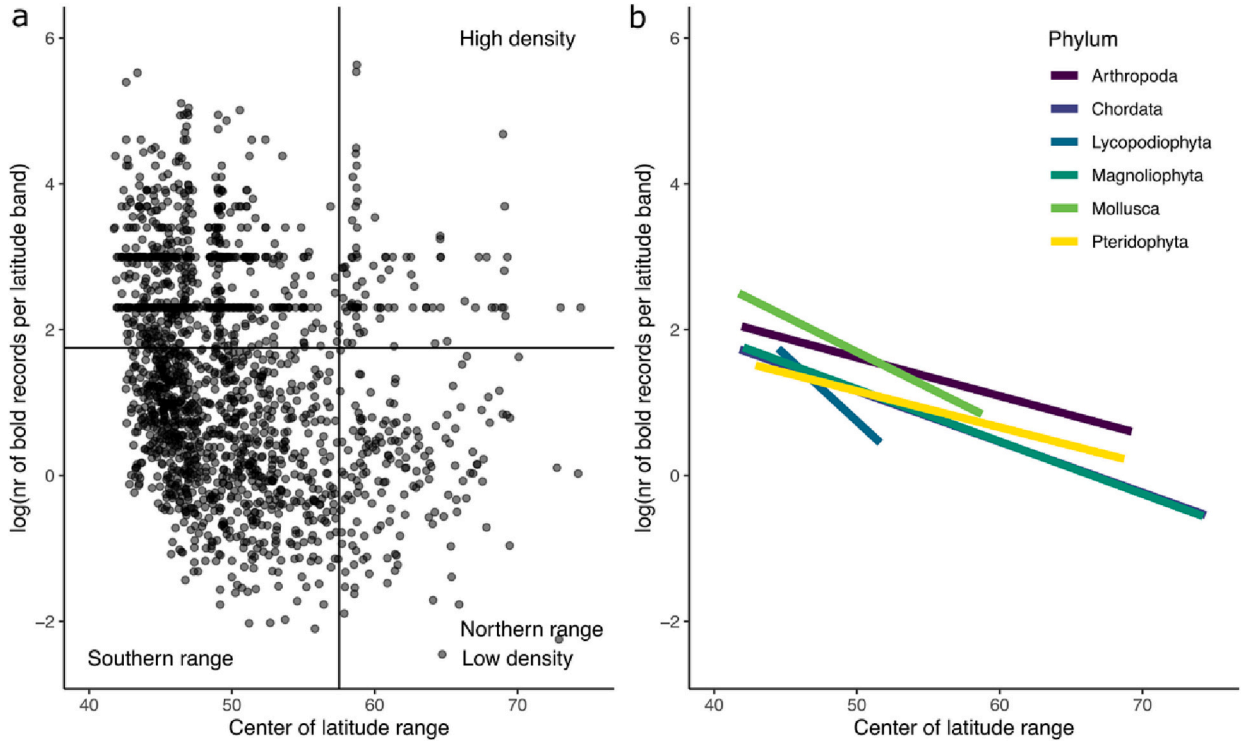


Fig. 4. The relationship between the spatial density of barcode sequences available in BOLD and the distribution of taxa (shown as the species latitudinal midpoint of BOLD records). Although many taxa had a relatively low number of records (located in bottom half of panel (a), northern taxa had disproportionately fewer BOLD records, with the numbers of BOLD records per latitude band decreasing as a function of the central position of a taxon's latitudinal distribution; a pattern evident across many phyla (b).

4.1. Biases in species representation

Charismatic taxa are most heavily sampled, resulting in an over-representation of some commonly studied organisms that are not necessarily the most important species from a functional or ecological perspective (i.e., charismatic may not equate to keystone). For instance, 28 % of arthropods considered were not listed in BOLD, yet this group is at the center of many food web interactions (Nakano and Murakami, 2001; Twining et al., 2019). Another bias is geographic coverage, whereby easily accessible sites (e.g., closer to roads and research centers) and those with a higher human footprint index (Hughes et al., 2021) are currently overrepresented. Both our worldwide and Canadian analyses have shown that records from fishes are typically acquired from more degraded habitats compared to birds, while invertebrates and amphibians are generally sampled in relatively intact regions (Fig. 3). This observation is concordant with previously identified sampling biases, notably showing that >80 % of GBIF records are acquired <2.5 km from a road (Hughes et al., 2021).

We observed that records from species with the highest conservation concerns (COSEWIC's assessments Endangered and Threatened) were from degraded regions (Fig. 3; Online Appendix Fig. A1). It is also worth noting that although the species list used as a starting point for this study (Desforges et al., 2022) is relatively exhaustive for macroscopic taxa, some groups (e.g., amphipods) as well as smaller organisms (e.g., crustacean zooplankton, rotifers, etc.) are not considered, and thus were not included in our data compilation. Other studies report that these taxonomic groups are currently poorly represented in barcode and genome databases (Weigand et al., 2019; Young and Hebert, 2022).

4.2. Challenges with taxonomic assignments and metadata

While DNA barcoding and metabarcoding are flexible and powerful biomonitoring approaches, there currently exist key limitations, particularly with respect to taxonomic identification. Notably, species identification requires both: 1) validated links between the DNA barcode and formal taxonomic identification (i.e., a voucher specimen, a Linnean name, and a valid taxonomy); and 2) a consistency of sequences within the species range. This latter point may not be equally feasible across different biomes and can thus require greater efforts in developing reference databases. Likewise, in megadiverse regions, such as the tropics, typically there are a large proportion of unassigned reads from species, because they have not yet been described or linked to a barcode. Here, barcodes without formal taxonomic identifications have been used in taxonomy-free monitoring approaches (Wilson et al., 2016). For barcodes matched to species by formal taxonomic identification, our study and others have highlighted that certain taxa are underrepresented in genetic databases, in terms of both the incidence and abundance of barcodes, in particular arthropods (Young and Hebert, 2022). Insects are indicative of freshwater ecosystem health and quality, forming part of the Water Framework Directive monitoring of ecological status in the European Union and of the Canadian Aquatic Biomonitoring Network (CABIN), and yet were found to be among the most data deficient group in our study which was based on species level assessments.

Finally, quality control of barcodes and their metadata, notably the geographic location, are important components of the taxonomic assignment protocol (Weigand et al., 2019). Approximately 90 % of barcodes in our study of Canadian taxa had sufficient geographic metadata to be included in our analysis of the spatial distribution of BOLD records, which was important to identify key areas where populations might have been under-sampled. Requirement standards vary between databases and ensuring that complete metadata are collected going forward will be valuable for conducting syntheses like this one, and can help address questions related to geographic variation in genetic diversity.

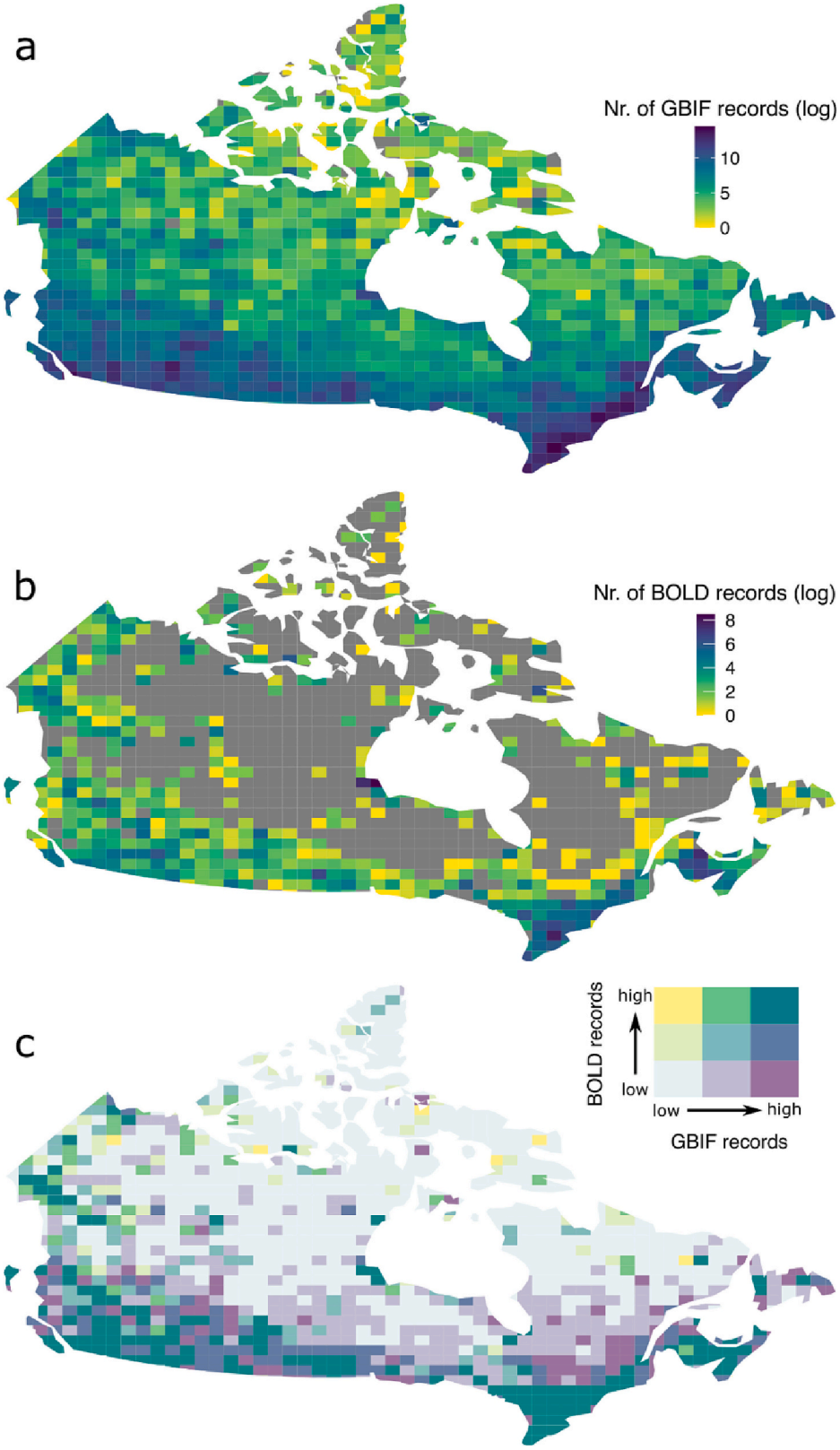


Fig. 5. Spatial overlap between georeferenced observations (GBIF) (a) and BOLD (b) records. Data include all inland water species listed by Desforges et al. (2022) with barcode sequences in BOLD. The gridded figure (c) is the high–low distribution of BOLD vs GBIF counts.

4.3. Reproducibility

The BOLD database, including barcode sequences and associated metadata, is easily accessible for download (e.g., via an R package) and is constantly updated, which means our study could be revisited to track the growth of species records. A challenge of this study and future endeavors towards the identification of data deficiencies in reference databases is the variability in taxonomy (including variable taxon name

spelling, synonyms), which in this study made it challenging to query databases when taxonomic names differed between the Desforges et al. (2022) and BOLD lists.

4.4. Opportunities for eDNA metabarcoding for biomonitoring and conservation

eDNA metabarcoding holds considerable potential as an ecological tool for inferring population- and community-level sorting of haplotypes and species (Overcast et al., 2021). Such an approach could be applied to understand drivers of biodiversity at different scales, especially in ecosystems responding to and recovering from anthropogenic disturbance. The majority of records in our BOLD search were for the mitochondrial cytochrome *c* oxidase subunit I gene (COI). DNA metabarcoding of COI is presently the best tool available to collect data on both species identification and intraspecific genetic diversity (Antil et al., 2022; Hebert et al., 2004).

COI barcodes contain important within-species variation for many different taxonomic groups (e.g., crustacean zooplankton: Derry et al., 2009; Frisch et al., 2021; Martin et al., 2021; rotifers: Derry et al., 2003; benthic macroinvertebrates: Ge et al., 2021). As such, COI barcoding is sometimes applied to track 'intraspecific diversity' in the form of cryptic speciation (e.g., De Luca et al., 2021). In addition, the genetic structure of metapopulations and metacommunities can be examined by tracking changes in COI haplotypes to reveal the relative importance of environmental selection (e.g., acidification) versus spatial and biogeographical landscape factors (Derry et al., 2009; Frisch et al., 2021; Martin et al., 2021). Combined with community metabarcoding of other markers (e.g., 18S rRNA gene or ITS), inferences of ecological responses to disturbance among taxa (COI + 18S rRNA gene: Astorg et al., 2022), and among genotypes or haplotypes within species (ITS2: Thibodeau et al., 2015; rbcL: Pérez-Burillo et al., 2021) can also be examined. Further, Petit-Marty et al. (2021) found reduced intraspecific nucleotide diversity in long-lived animal species assessed as threatened by the International Union of Conservation of Nature (IUCN); COI barcodes also hold potential as an intraspecific tool for early assessment of species conservation status. Taken together, COI data may be used to reveal differences in ecological tolerances among haplotypes, to identify neutral demographic processes linked with population bottlenecks resulting from environmental selection, or to analyze spatial distribution patterns resulting from a blend of historical and contemporary gene flow and colonization history (Frisch et al., 2021; Hebert et al., 2004; Orsini et al., 2013; Rubinoff and Holland, 2005).

Strategies to develop biodiversity monitoring priorities include focusing on at risk or data deficient species, selecting different phylogenetic groups to ensure a broad coverage across the tree of life, and increasing spatial coverage to improve our understanding of intraspecific species variability. Overall, genetic diversity in Canada's most pristine habitats is largely unknown (Fig. 5). For obvious logistical reasons, these habitats tend to be undersampled. However, new initiatives are aimed at improving the coverage of Canadian aquatic diversity and population health. In particular, Genome Canada's GEN-FISH initiative is currently attempting to sequence eDNA from representative fish species across Canada's six major drainage basins (e.g., Weigand et al., 2019). Funding for such monitoring programs is pertinent to both refining our understanding of Canada's genetic diversity and increasing the accuracy of public databases. Similar efforts in other jurisdictions are sorely needed.

BOLD provided an extensive coverage (86 %) of Canadian inland water species. The phylogenetic and spatial representation examined in our analyses helped identify where future resources should be focused to address deficiencies in the remaining 14 % of species, and to bolster sequencing efforts of taxa that have few barcodes to address intraspecific variability. Large-scale barcode sequencing initiatives such as iBOL BIOSCAN (https://ibol.org/programs/bioscan/) have targets to cover half of the world's ecoregions to support the iBOL Planetary Biodiversity Mission. These ambitious long-term initiatives will hopefully continue to increase the coverage of species by establishing a complete census of all multicellular species as well as a global biosurveillance program. However, with projected end dates of 2045, the timeline is likely to be too late to prevent the loss of significant biodiversity. Thus, in the near-term, targeted and collaborative initiatives such as the present study are essential to address taxa that are either At Risk or Data Deficient, so that resources grow together. One immediate way that genetic resources could prove useful is in identifying key biodiversity areas (KBAs). In Canada, efforts are underway by the Wildlife Conservation Society and partners to identify and prioritize KBAs (https://www.kbacanada.org/). For many species or areas, there has been the need to rely extensively on experts. Unfortunately, that expertise often fails to adequately consider the freshwater invertebrates discussed here. KBAs may be inappropriately delineated if the knowledge of species distributions is incomplete. As such, eDNA could serve as a rapid and robust means of ensuring that some of the most cryptic and understudied species are considered when identifying KBAs (Hunter et al., 2018).

Another major sequencing initiative is the Canada BioGenome Project, funded by Genome Canada and part of the International Earth BioGenome Project. This initiative aims to sequence species relevant to in situ and ex situ conservation planning, biodiversity, and the economy. Initial funding is targeting the generation of genome assemblies for approximately 400 species, which will be collaboratively selected with Indigenous peoples' organizations, as well as other stakeholders and end-users. Policy recommendations will also be provided on how to responsibly implement the use of genomic tools in wildlife conservation, co-management, ecosystem-based management, species restoration, and environmental monitoring. Whole genome sequencing, assembly and annotation in collaboration with the European Bioinformatics Institute (EBI) and in line with practices established as part of the Earth BioGenome Project (https://www.earthbiogenome.org) will be conducted for the chosen species. The resultant genomes will be made freely available to interested non-experts by the development of a user-friendly geospatial platform to store and distribute data. Additional genome sequences have been and/or are in the process of being released as part of the CanSeq150 project, a project established by CGEn, Canada's national platform for genome sequencing and analysis, to provide genomes for 150 Canadian species in support of conservation efforts. The list of species to be sequenced (or already sequenced) in this framework includes many taxa reliant on inland waters (Online Appendix Table A6; https://www.cgen.ca/canseq150-overview). Finally, portable long-read sequencing platforms (currently, Oxford Nanopore) offer opportunities to collect genetic data when shipping of specimens is difficult from remote sampling locations (Krehenwinkel et al., 2019).

5. Recommendations and conclusions

Conserving and managing biodiversity requires knowledge of the spatial distribution of species and how that distribution is changing over time. Such knowledge is fundamental to regional, national, and international (i.e., IUCN Red List) threat assessment. Here we argue that genetic resources could help address some of the data deficiencies that have been plaguing inland water (and other) ecosystems for decades. It is evident that many species are undergoing range shifts due to climate change, habitat destruction, species introductions, and species exploitation, and we are also losing species in some areas before we know they exist (Revenga et al., 2005). The use of eDNA to monitor biodiversity could allow us to detect and anticipate these urgent scenarios.

Using our analytical framework, we identified a priority list of 18 taxa that would benefit from immediate sequencing based on conservation risk and availability of barcodes (Online Appendix Table A3). To continue to grow biodiversity and conservation status assessments, consistent taxonomy must be used within and across databases to prevent redundancies and facilitate comparisons and synthesis. Overall, we have provided insights into the molecular resources available for taxa

reliant on inland waters in Canada, as well as highlighted general perspectives and recommendations that may contribute to future national and international biodiversity and conservation assessment based on eDNA approaches and associated molecular genetic resources.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2023.109963.

CRediT authorship contribution statement

Marie-Eve Monchamp: Conceptualization, Investigation, Writing – Original draft preparation, Visualization, Funding acquisition, Writing - Review & Editing **Zofia E. Taranu**: Writing – Original draft preparation, Investigation, Data curation, Visualization, Software, Writing - Review & Editing **Rebecca E. Garner**: Investigation, Writing – Original draft preparation, Data curation, Software, Writing - Review & Editing **Tessa Rehill**: Conceptualization, Investigation, Data curation, Writing - Review & Editing **Olivier Morissette**: Investigation, Visualization, Writing - Review & Editing, Software **Vincent Fugère**: Investigation, Visualization, Writing - Review & Editing, Software **Lars L. Iversen**: Investigation, Visualization, Writing - Review & Editing, Software **Joanne E. Littlefair**: Writing – Original draft preparation, Writing - Review & Editing **Alison M. Derry**: Writing – Original draft preparation, Writing - Review & Editing **Naíla Barbosa da Costa**: Investigation, Data curation, Writing - Review & Editing **Jessica E. Desforges**: Investigation, Data curation, Writing - Review & Editing **Joe R. Sánchez Schacht**: Investigation, Data curation, Writing - Review & Editing **Steven J. Cooke**: Investigation, Writing - Review & Editing **Rowan D.H. Barrett**: Writing – Original draft preparation, Writing - Review & Editing **David A. Walsh**: Writing - Review & Editing **Jiannis Ragoussis**: Investigation, Writing - Review & Editing **Monique Albert**: Writing - Review & Editing **Melania E. Cristescu**: Supervision, Writing - Review & Editing **Irene Gregory-Eaves**: Conceptualization, Writing – Original draft preparation, Funding acquisition, Supervision, Writing - Review & Editing.

Declaration of competing interest

The authors declare no competing interests.

Data availability

All data used in the present study are available in the paper or as supplementary materials. R scripts are available at https://github.com/monchama/FreshConservDNA.

Acknowledgements

We appreciate all of the hard work done by many scientists who uploaded their sequence data to BOLD and NCBI.

Funding

This work is the product of a workshop grant funded by the Groupe de Recherche Interuniversitaire en Limnologie (GRIL). Additional funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) to IGE, an NSERC postdoctoral fellowship to MEM, an FRQNT doctoral scholarship to REG, Genome Canada to JR, and a Canada Research Chair award to IGE.

References

Albert, J.S., Destouni, G., Duke-Sylvester, S.M., Magurran, A.E., Oberdorff, T., Reis, R.E., Winemiller, K.O., Ripple, W.J., 2021. Scientists' warning to humanity on the freshwater biodiversity crisis. Ambio 50 (1), 85–94. https://doi.org/10.1007/s13280-020-01318-8.

Antil, S., Abraham, J.S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., Bhagat, P., Gupta, R., Sood, U., Lal, R., Toteja, R., 2022. DNA barcoding, an effective tool for species identification: a review. Mol. Biol. Rep. 0123456789 https://doi.org/10.1007/s11033-022-08015-7.

Astorg, L., Gagnon, J.C., Lazar, C.S., Derry, A.M., 2022. Effects of freshwater salinization on a salt-naïve planktonic eukaryote community. Limnol. Oceanogr. Lett. https://doi.org/10.1002/lol2.10229.

Balasingham, K.D., Walter, R.P., Mandrak, N.E., Heath, D.D., 2018. Environmental DNA detection of rare and invasive fish species in two Great Lakes tributaries. Mol. Ecol. 27 (1), 112–127. https://doi.org/10.1111/mec.14395.

Berry, O., Jarman, S., Bissett, A., Hope, M., Paeper, C., Bessey, C., Schwartz, M.K., Hale, J., Bunce, M., 2021. Making environmental DNA (eDNA) biodiversity records globally accessible. Environ. DNA 3 (4), 699–705. https://doi.org/10.1002/edn3.173.

Bush, A., Compson, Z.G., Monk, W.A., Porter, T.M., Steeves, R., Emilson, E., Gagne, N., Hajibabaei, M., Roy, M., Baird, D.J., 2019. Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. Front. Ecol. Evol. 7 (November), 1–12. https://doi.org/10.3389/fevo.2019.00434.

Chamberlain, S., 2020. Natserv: 'NatureServe' Interface. R Package.

Chamberlain, S., 2021. bold: Interface to Bold Systems API. R Package.

Chamberlain, S.A., Szöcs, E., 2013. Taxize: taxonomic search and retrieval in R. F1000Research 2 (191), 191. https://doi.org/10.12688/f1000research.2-191.v2.

Curry, C.J., Gibson, J.F., Shokralla, S., Hajibabaei, M., Baird, D.J., 2018. Identifying north american freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? Freshwater Sci. 37 (1), 178–189. https://doi.org/10.1086/696613.

De Luca, D., Piredda, R., Sarno, D., Kooistra, W.H.C.F., 2021. Resolving cryptic species complexes in marine protists: phylogenetic haplotype networks meet global DNA metabarcoding datasets. ISME J. 15 (7), 1931–1942. https://doi.org/10.1038/s41396-021-00895-0.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Mol. Ecol. 26 (21), 5872–5895. https://doi.org/10.1111/mec.14350.

Derry, A.M., Prepas, E.E., Hebert, P.D.N., 2003. A comparison of zooplankton communities in saline lakewater with variable anion composition. Hydrobiologia 505 (1–3), 199–215. https://doi.org/10.1023/B:HYDR.0000007414.12566.19.

Derry, A.M., Arnott, S.E., Shead, J.A., Hebert, P.D.N., Boag, P.T., 2009. Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. Ecology 90 (8), 2275–2286. https://doi.org/10.1890/07-1852.1.

Desforges, J.E., Clarke, J., Harmsen, E.J., Jardine, A.M., Robichaud, J.A., Serré, S., Chakrabarty, P., Bennett, J.R., Hanna, D.E.L., Smol, J.P., Rytwinski, T., Taylor, J.J., Martel, A.L., Winegardner, A.K., Marty, J., Taylor, M.K., O'connor, C.M., Robinson, S.A., Reid, A.J., Cooke, S.J., 2022. The alarming state of freshwater biodiversity in Canada. Can. J. Fish. Aquat. Sci. 79 (2), 352–365. https://doi.org/10.1139/cjfas-2021-0073.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Lucas, A.G., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Zayas, C.N., 2019. Pervasive human-driven decline of life on earth points to the need for transformative change. Science 366 (6471). https://doi.org/10.1126/science.aax3100.

Dudgeon, D., 2019. Multiple threats imperil freshwater biodiversity in the anthropocene. Curr. Biol. 29, R960–R967. https://doi.org/10.1016/j.cub.2019.08.002.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. Camb. Philos. Soc. 81 (2), 163–182. https://doi.org/10.1017/S1464793105006950.

Frisch, D., Lejeusne, C., Hayashi, M., Bidwell, M.T., Sánchez-Fontenla, J., Green, A.J., 2021. Brine chemistry matters: isolation by environment and by distance explain population genetic structure of Artemia franciscana in saline lakes. Freshw. Biol. 66 (8), 1546–1559. https://doi.org/10.1111/fwb.13737.

Ge, Y., Xia, C., Wang, J., Zhang, X., Ma, X., Zhou, Q., 2021. The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. Ecol. Evol. 11 (10), 5669–5681. https://doi.org/10.1002/ece3.7470.

Harrison, I., Abell, R., Darwall, W., Thieme, M.L., Tickner, D., Timboe, I., 2018. The freshwater biodiversity crisis. Science 362 (6421), 1369. https://doi.org/10.1126/science.aav9242.

Haussler, D., O'Brien, S.J., Ryder, O.A., Keith Barker, F., Clamp, M., Crawford, A.J., Hanner, R., Hanotte, O., Johnson, W.E., McGuire, J.A., Miller, W., Murphy, R.W., Murphy, W.J., Sheldon, F.H., Sinervo, B., Venkatesh, B., Wiley, E.O., Allendorf, F.W., Amato, G., Turner, S., 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10000 vertebrate species. J. Hered. 100 (6), 659–674. https://doi.org/10.1093/jhered/esp086.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. 101 (41), 14812–14817. https://doi.org/10.1073/pnas.0406166101.

Hijmans, R.J., 2023. terra: Spatial Data Analysis. R package version 1.7-3. https://CRAN.R-project.org/package=terra.

Hirsh-Pearson, K., Johnson, C.J., Schuster, R., Wheate, R.D., Venter, O., 2022. Canada's human footprint reveals large intact areas juxtaposed against areas under immense anthropogenic pressure. Facets 7, 398–419. https://doi.org/10.1139/facets-2021-0063.

Hotaling, S., Kelley, J.L., Frandsen, P.B., 2021. Toward a genome sequence for every animal: where are we now? PNAS 118 (52), 1–8. https://doi.org/10.1073/pnas.2109019118/-/DCSupplemental.Published.

Hughes, A.C., Orr, M.C., Ma, K., Costello, M.J., Waller, J., Provoost, P., Yang, Q., Zhu, C., Qiao, H., 2021. Sampling biases shape our view of the natural world. Ecography 44, 1259–1269. https://doi.org/10.1111/ecog.05926.

Hunter, M.E., Hoban, S.M., Bruford, M.W., Segelbacher, G., Bernatchez, L., 2018. Next-generation conservation genetics and biodiversity monitoring. Evol. Appl. 11 (7), 1029–1034. https://doi.org/10.1111/eva.12661.

IPBES, 2019. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Zenodo. https://doi.org/10.5281/zenodo.6417333.

Jerde, C.L., Mahon, A.R., Chadderton, W.L., Lodge, D.M., 2011. 'Sight-unseen' detection of rare aquatic species using environmental DNA. Conserv. Lett. 4 (2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x.

Keck, F., Blackman, R.C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., Kirschner, D., Locher, N., Zhang, H., Altermatt, F., 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Mol. Ecol. 31 (6), 1820–1835. https://doi.org/10.1111/mec.16364.

Krehenwinkel, Pomerantz, Prost, 2019. Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. Genes 10 (11), 858. https://doi.org/10.3390/genes10110858.

Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I. V., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S., Castilla-Rubio, J.C., Zhang, G., 2018. Earth BioGenome project: sequencing life for the future of life. Proc. Natl. Acad. Sci. U. S. A. 115 (17), 4325–4333. https://doi.org/10.1073/pnas.1720115115.

Maasri, A., Jähnig, S.C., Adamescu, M.C., Adrian, R., Baigun, C., Baird, D.J., Batista-Morales, A., Bonada, N., Brown, L.E., Cai, Q., Campos-Silva, J.V., Clausnitzer, V., Contreras-MacBeath, T., Cooke, S.J., Datry, T., Delacámara, G., De Meester, L., Dijkstra, K.D.B., Do, V.T., Worischka, S., 2022. A global agenda for advancing freshwater biodiversity research. Ecol. Lett. 25 (2), 255–263. https://doi.org/10.1111/ele.13931.

Martin, G.K., Beisner, B.E., Chain, F.J.J., Cristescu, M.E., del Giorgio, P.A., Derry, A.M., 2021. Freshwater zooplankton metapopulations and metacommunities respond differently to environmental and spatial variation. Ecology 102 (1), 1–11. https://doi.org/10.1002/ecy.3224.

Mcelroy, M.E., Dressler, T.L., Titcomb, G.C., Wilson, E.A., Deiner, K., Dudley, T.L., Eliason, E.J., Evans, N.T., Gaines, S.D., Lafferty, K.D., Lamberti, G.A., Li, Y., Lodge, D.M., Love, M.S., Mahon, A.R., Pfrender, M.E., Renshaw, M.A., Selkoe, K.A., Allan, E.A., 2020. In: Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness, 8, pp. 1–12. https://doi.org/10.3389/fevo.2020.00276. August.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. 7, 1–11. https://doi.org/10.1038/ncomms13603.

Nakano, S., Murakami, M., 2001. Reciprocal subsidies: dynamic interdependence between terrestrial and aquatic food webs. Proc. Natl. Acad. Sci. U. S. A. 98 (1), 166–170. https://doi.org/10.1073/pnas.98.1.166.

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J.K., Pfrender, M.E., Weider, L.J., 2013. The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. Trends Ecol. Evol. 28 (5), 274–282. https://doi.org/10.1016/j.tree.2013.01.009.

Overcast, I., Ruffley, M., Rosindell, J., Harmon, L., Borges, P.A.V., Emerson, B.C., Etienne, R.S., Gillespie, R., Krehenwinkel, H., Mahler, D.L., Massol, F., Parent, C.E., Patiño, J., Peter, B., Week, B., Wagner, C., Hickerson, M.J., Rominger, A., 2021. A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. Mol. Ecol. Resour. 21 (8), 2782–2800. https://doi.org/10.1111/1755-0998.13514.

Pérez-Burillo, J., Trobajo, R., Leira, M., Keck, F., Rimet, F., Sigró, J., Mann, D.G., 2021. DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. Sci. Total Environ. 798, 149029 https://doi.org/10.1016/j.scitotenv.2021.149029.

Petit-Marty, N., Vázquez-Luis, M., Hendriks, I.E., 2021. Use of the nucleotide diversity in COI mitochondrial gene as an early diagnostic of conservation status of animal species. Conserv. Lett. 14 (1), 1–7. https://doi.org/10.1111/conl.12756.

Phillips, J.D., Gillis, D.J., Hanner, R.H., 2019. Incomplete estimates of genetic diversity within species: implications for DNA barcoding. Ecol. Evol. 9 (5), 2996–3010. https://doi.org/10.1002/ece3.4757.

R Core Team, 2020. R: A language and environment for statistical computing. In: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.r-project.org/.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system: barcoding. Mol. Ecol. Notes 7 (3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., Kidd, K.A., Maccormack, T.J., Olden, J.D., Ormerod, S.J., Smol, J.P., Taylor, W.W., Tockner, K., Vermaire, J.C., Dudgeon, D., Cooke, S.J., 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. 94, 849–873. https://doi.org/10.1111/brv.12480.

Revenga, C., Campbell, I., Abell, R., De Villiers, P., Bryer, M., 2005. Prospects for monitoring freshwater ecosystems towards the 2010 targets. Philos. Trans. R. Soc. B: Biol. Sci. 360 (1454), 397–413. https://doi.org/10.1098/rstb.2004.1595.

Rubinoff, D., Holland, B.S., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54 (6), 952–961. https://doi.org/10.1080/10635150500234674.

Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V., Woolmer, G., 2002. The human footprint and the last of the wild: the human footprint is a global map of human influence on the land surface, which suggests that human beings are stewards of nature, whether we like it or not. Bioscience 52 (10), 891–904. https://doi.org/10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2.

Thibodeau, G., Walsh, D.A., Beisner, B.E., 2015. Rapid eco-evolutionary responses in perturbed phytoplankton communities. Proc. R. Soc. B Biol. Sci. 282 (1814) https://doi.org/10.1098/rspb.2015.1215.

Thomsen, P.F., Willerslev, E., 2015. Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. 183, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

Tickner, D., Opperman, J.J., Abell, R., Acreman, M., Arthington, A.H., Bunn, S.E., Cooke, S.J., Dalton, J., Darwall, W., Edwards, G., Harrison, I., Hughes, K., Jones, T., Leclère, D., Lynch, A.J., Leonard, P., McClain, M.E., Muruven, D., Olden, J.D., Ormerod, S.J., Robinson, J., Tharme, R.E., Thieme, M., Tockner, K., Wright, M., Young, L., 2020. Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. Bioscience 70, 330–342. https://doi.org/10.1093/biosci/biaa002.

Twining, C.W., Brenna, J.T., Lawrence, P., Winkler, D.W., Flecker, A.S., Hairston, N.G., 2019. Aquatic and terrestrial resources are not nutritionally reciprocal for consumers. Funct. Ecol. 33 (10), 2042–2052. https://doi.org/10.1111/1365-2435.13401.

Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K.R., Possingham, H.P., Laurance, W.F., Wood, P., Fekete, B.M., Levy, M.A., Watson, J.E. M., 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nat. Commun. 7, 1–11. https://doi.org/10.1038/ncomms12558.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: gap-analysis and recommendations for future work. Sci. Total Environ. 678 (April), 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247.

Wilson, J.-J., Sing, K.-W., Lee, P.-S., Wee, A.K.S., 2016. Application of DNA barcodes in wildlife conservation in tropical East Asia. Conserv. Biol. 30 (5), 982–989. https://doi.org/10.1111/cobi.12787.

Winter, D.J., 2017. Rentrez: an R package for the NCBI eUtils API. R J. 9 (2), 520–526. https://doi.org/10.32614/rj-2017-058.

Young, M.R., Hebert, P.D.N., 2022. Unearthing soil arthropod diversity through DNA metabarcoding. PeerJ 10. https://doi.org/10.7717/peerj.12845.

Young, H.S., McCauley, D.J., Galetti, M., Dirzo, R., 2016. Patterns, causes, and consequences of anthropocene defaunation. Annu. Rev. Ecol. Evol. Syst. 47 (August), 333–358. https://doi.org/10.1146/annurev-ecolsys-112414-054142.