

Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs specimen age

Donald J. Baird¹

*Environment Canada, Canadian Rivers Institute, Department of Biology, University of New Brunswick,
P.O. Box 4400, Fredericton, New Brunswick, E3B 5A3, Canada*

Timothy J. Pascoe²

*Environment Canada, Canada Centre for Inland Waters, 867 Lakeshore Road, Burlington, Ontario,
L7R 4A6, Canada*

Xin Zhou³ AND Mehrdad Hajibabaei⁴

*Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, Ontario,
N1G 2W1, Canada*

Abstract. As part of its ongoing work in biomonitoring, Environment Canada's Canadian Aquatic Biomonitoring Network (CABIN) program has assembled an expert-verified reference collection of 3864 specimens of 604 species of Canadian freshwater macroinvertebrates. Such collections are a key resource for developing a deoxyribonucleic acid (DNA) barcode library to facilitate molecular identification of biomonitoring samples. We examined the problems encountered in using such legacy material to obtain reference barcodes. We focused on the influence of specimen age and preservation history. To supplement work on the reference collection, we determined the time-dependent effects of formalin preservation on DNA-barcode integrity in 4 common arthropod taxa by controlled exposure of fresh material obtained from laboratory cultures. Specimens in the reference collection were preserved with short-term fixation in formalin followed by prolonged preservation in 70% ethanol. Only 19 caddisfly larval specimens out of the total of 650 analyzed returned full-length sequences. In contrast, formalin preservation of freshly collected material for up to 20 d yielded good sequencing success and high-quality sequences. Freshly collected material clearly provides the best basis for the future development of DNA-barcode libraries, and formalin preservation should be avoided where possible to ensure that DNA integrity is maximized.

Key words: DNA barcode, identification, taxonomy, preservation methods, Trichoptera, reference collection.

Routine taxonomic processing of benthic samples is a significant bottleneck in the development and operation of large-scale biomonitoring programs. Capacity for taxonomic identification currently is limited and possibly declining (e.g., Hopkins and Freckleton 2002), and comprehensive keys permitting identification of all life-stages of species of interest are lacking. This bottleneck limits the potential in biomonitoring science for application of new methods for causal diagnosis (e.g., traits-based ecological risk

assessment, sensu Baird et al. 2008). The advent of molecular methods for specimen identification based on deoxyribonucleic acid (DNA) barcodes (Hebert et al. 2003) offers an opportunity to remove this bottleneck and the exciting prospect of being able to characterize the taxonomic composition of bulk samples of material collected from the field rapidly, accurately, and potentially at very low cost (Mardis 2008). However, before the potential of these new technologies can be realized, we need to begin the process of building a library of DNA-barcode information from taxonomically verified material. Once developed, this library will permit DNA-barcode sequences from field-collected benthic samples to be identified accurately by specimen-to-library-barcode

¹ E-mail addresses: djbaird@unb.ca

² timothy.pascoe@ec.gc.ca

³ xinzhou@uoguelph.ca

⁴ mhajibab@uoguelph.ca

matching through bioinformatics techniques (Singer and Hajibabaei 2009).

DNA-barcode libraries can be constructed from a variety of sources that range from purposeful field collection (Zhou et al. 2009) to mining of curated material from individual laboratory or museum reference collections. Working with existing identified material is preferable to field collection because it permits rapid population of DNA-sequence libraries from verified material. However, obtaining DNA information from aging, preserved material is challenging (Skage and Schander 2007, Espeland et al. 2010) because specimen DNA degrades unpredictably over time. In the case of laboratory reference collections of freshwater benthic invertebrates, reference specimens generally are maintained in ethanol, a DNA-friendly preservation medium, for long-term storage. However, benthic samples collected in the field often are preserved in formalin solution for periods ranging from days to weeks, depending on the nature and duration of the field work and individual laboratory practices. Formalin is considered a DNA-unfriendly medium by molecular biologists (Zimmermann et al. 2008), and its continued use presents a challenge to benthic scientists who want to use DNA-based identification methods. In contrast, many benthic scientists view formalin as a more fieldwork-friendly preservative than alcohol, which tends to be bulky to transport, because formalin can be carried in concentrated form and diluted for use while in the field. Moreover, samples preserved in alcohol are considered dangerous goods for shipping purposes because of their volatility and flashpoint risk. Thus, alcohol-preserved samples present health and safety issues during transport. However, formalin, a suspected carcinogen (e.g., Beane Freeman et al. 2009), also poses a potential health risk to users.

Our aims in this paper are 2-fold. First, we present a case study of mining an existing reference collection of taxonomically verified freshwater benthic macroinvertebrate specimens for DNA-barcode sequences with the purpose of constructing a library for future identification of field-collected samples. To achieve this goal, we sought to obtain DNA-barcode sequences of 658 base pairs (bp) in length (full-length barcodes). Identification can be obtained from smaller sequence lengths of ~130 bp—so called mini-barcodes (Hajibabaei et al. 2006, Meusnier et al. 2008)—but we focused on obtaining full-length barcode sequences where possible. Second, we investigated the influence of formalin preservation on the DNA integrity of macroinvertebrate species more directly by conducting an experiment to examine time-related

DNA degradation of specimens of 4 laboratory-cultured species.

Methods

Sourcing of reference-collection material

Environment Canada has been developing a national reference database of benthic invertebrates for Canada since the early 1990s. Established officially as the Canadian Aquatic Biomonitoring Network (CABIN) program in 2006, CABIN provides consistent, comparable, and scientifically defensible data through the use of nationally standardized protocols for the collection, identification, and analysis of benthic macroinvertebrates and associated water-quality and habitat information.

Most reference material maintained by CABIN personnel was obtained during research projects in the Laurentian Great Lakes, Ontario, and the Fraser River Valley, British Columbia, that formed the early basis for the national program. However, the collection spans the country and contains material from all major provinces and territories. The reference collection is housed at the Pacific Environmental Science Centre (PESC) in North Vancouver, British Columbia, and is curated by a full-time taxonomist. Specimens are stored at room temperature in 70% ethanol, sorted according to project, location of collection, and taxonomic group. At the time of our study, the reference collection contained 3864 specimens, identified and externally verified as 607 species, of which 424 were in the form of larval stages of aquatic insects. The specimens varied in time of preservation from <1 y to 23 y. Specimens originated from field biomonitoring samples, where the practice was to fix samples in a 10% solution of 1:3 buffered formalin in the field and to transfer to 70% alcohol for long-term storage in the laboratory. How long specimens had remained in formalin prior to transfer to alcohol could not be determined because this stage varied among field studies. In most cases it lasted from one to several days, depending on the remoteness of the sampling location.

We decided to focus on aquatic insects to examine the ability of the reference collection to provide a source of reference barcodes because they are used routinely for biomonitoring purposes and their taxonomy is relatively well-characterized. Almost ½ of the insect specimens were larval stages of Trichoptera. We limited our study to this group because the large sample size and range of specimen age permitted us to examine the influence of specimen age on DNA integrity while minimizing potentially confounding issues arising from differential affinity in primer DNA amplification among taxonomic groups.

DNA extraction and sequencing

DNA analyses were conducted at the Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Canada. Genomic DNA was extracted from a single leg or a piece of body tissue with NucleoSpin® 96 Tissue Kit (Macherey-Nagel, Düren, Germany) and default protocols. A total of 100 µL of genomic DNA was obtained, and 1 µL was used for amplification of the cytochrome c oxidase subunit I (COI) barcode region. Primer sets LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3')/HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al. 1994), and LepF1 (5'-ATTCAACCAATCA TAAAGATATTGG-3')/LepR1 (5'-TAAACTTCTGGAT GTCCAAAAATCA-3') (Hebert et al. 2004) were used to obtain the full-length barcodes (658 bp). Routine DNA-barcoding protocols were followed for polymerase chain reaction (PCR) amplification and sequencing (Ivanova et al. 2006). The final sequencing reactions were carried out bidirectionally using BigDye v3.1 on an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, California). If amplification and sequencing for formalin-preserved caddisfly reference samples failed in the first pass (full-length), Uni-MinibarF1 (5'-TCCACTAATCACAARGATATTGGTAC-3') and Uni-MinibarR1 (5'-GAAAATCATAATGAAGGCATGAGC-3') primers designed for a short fragment at the 5' terminus of the standard barcode region (130 bp mini-barcodes; Meusnier et al. 2008) were applied.

Sequencing trace files were edited using Sequencher (version 4.0.5; Gene Codes Corporation, Ann Arbor, Michigan). COI sequences were aligned using MEGA (version 4.0; Tamura et al. 2007) and integrated ClustalX using default settings. The alignment also was checked manually to avoid stop codons, indels, and amino-acid translation frame shift.

COI sequences and associated voucher information are publicly available projects in the Barcode of Life Data System (BOLD systems; <http://www.barcodinglife.org>): Barcoding formalin freshwater macroinvertebrates (CDINV) and Barcoding Trichoptera reference samples (EVCAD). COI sequences also have been published in GenBank under accession numbers: cultured macroinvertebrates: HM137887–HM138082, Trichoptera reference: HM137755–HM137884.

Formalin-preservation experiment

Cultures of benthic organisms have been established and maintained at the Canada Center for Inland Waters (CCIW; Burlington, Ontario) since the early 1990s (Borgmann et al. 1989, Reynoldson et al. 1991, 1998, Milani et al. 2003). Four culture organisms have been maintained for use in sediment toxicology

testing: the amphipod *Hyaella azteca* (Saussure), the midge *Chironomus riparius* (Meigen) (larval stage), the mayfly *Hexagenia limbata* (Serville) (larval stage), and the oligochaete worm *Tubifex tubifex* (Müller).

We used individuals obtained from these laboratory cultures to study the effect of formalin exposure on DNA extraction and sequencing. For each species, we allocated a total of 30 specimens in groups of 5 replicate individuals to 6 formalin-preservation treatments. Formalin treatments consisted of 6 timed exposures (0 [control], 2, 5, 10, 15, or 20 d) to a 1:3 solution of 10% buffered neutral formalin and water. We used certified 10% phosphate-buffered formalin for exposure (Fisher Scientific, Fair Lawn, New Jersey) and City of Burlington tap water (from Lake Ontario) for culturing and testing purposes. This water is aerated and C-filtered prior to use and is analyzed monthly for nutrients, major ions, and water hardness/alkalinity to ensure it is within acceptable standards (Reynoldson et al. 1998).

We carried out a 2nd set of 6 treatments under identical conditions, but with the addition of sediment to the formalin mixture to simulate inclusion of organic material typically associated with field-collected benthic samples. We used a quality assurance/quality control sediment used for toxicology research at the National Water Research Institute. The sediment is collected on an annual basis from the United Nations Educational, Scientific, and Cultural Organization (UNESCO) world biosphere reserve at Long Point, Lake Erie, and has a high organic C content and appropriate particle size for all species (Reynoldson et al. 1998). We used 5 replicate individuals in each treatment, with the exception of *Hyaella azteca* in the 10-d formalin exposure, in which we exposed 4 individuals. We subsequently transferred individual organisms to ~1 mL of 70% ethanol solution and held them in latch racks of 96 Matrix 2D barcoded screw-top storage tubes (Catalog number 3744; Thermo Scientific, Nepean, Canada) and processed them for DNA analysis. We used DNA extraction methods identical to those described above for sequencing the reference collection.

Results

DNA analysis of the CABIN reference-collection material

The reference voucher caddisfly specimens ranged in age (since date of collection) from <1 to 23 y. The overall sequencing success rate was low (Fig. 1). Of the 650 specimens sequenced, 82 produced usable sequences, but only 19 (2.9%) yielded full-length barcodes. An additional 115 mini-barcodes (17.5%) were acquired using a 2nd round of Uni-MinibarF1

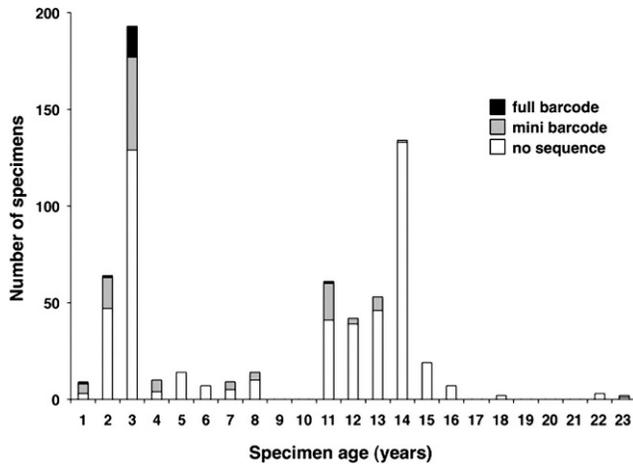


FIG. 1. Sequencing success of 650 larval caddisfly specimens of varying age obtained from Environment Canada’s Canadian Aquatic Biomonitoring Network (CABIN) reference collection. Sequencing success is reported in terms of full-length barcode reads, mini-barcode reads, or failures.

and Uni-MinibarR1 primers from the samples that failed the 1st round of Folmer primer amplification. The remaining 78.4% of samples did not return usable sequence data.

DNA analysis of the formalin-exposure experiment

In a single run, 196 of 219 specimens (90%) were sequenced successfully with standard Folmer et al. (1994) primers (Appendix; available online from: <http://dx.doi.org/10.1899/10-013.1.s1>). DNA quality was generally high (full-length and low noise) when formalin exposure was <15 d (Table 1). Sequence quality began to decrease after 20 d of exposure (although many 20-d-exposure samples did yield high-quality sequences). Deterioration was reflected in the increased number of ambiguous base calls and low sequencing success in *Chironomus riparius* and *Hexagenia limbata*. DNA quality did not seem to be adversely influenced by the presence of sediments (Table 1).

TABLE 1. Sequencing success in 4 cultured macroinvertebrate species exposed to formalin for 0, 2, 5, 10, 15, or 20 d for samples exposed without (-) and with (+) sediments. One set of control specimens (0 d) was used for both treatments. Table contents indicate number of successful sequences obtained for each set of *n* = 4 or *n* = 5 replicate specimens.

Taxon	0 d		2 d		5 d		10 d		15 d		20 d	
	-	+	-	+	-	+	-	+	-	+	-	+
<i>Chironomus riparius</i>	5/5	n/a	4/5	5/5	4/5	5/5	5/5	5/5	5/5	5/5	1/5	5/5
<i>Hexagenia limbata</i>	5/5	n/a	5/5	5/5	5/5	4/5	4/5	5/5	5/5	4/5	5/5	1/5
<i>Hyalella azteca</i>	2/5	n/a	4/5	4/5	5/5	5/5	4/4	5/5	5/5	5/5	5/5	5/5
<i>Tubifex tubifex</i>	5/5	n/a	4/5	5/5	4/5	5/5	4/5	4/5	5/5	5/5	5/5	4/5

Discussion

The poor sequencing success obtained from the reference-collection material indicates that the typical pattern of fixing benthic samples in formalin in the field coupled with long-term storage in 70% ethanol had a strongly negative influence on DNA integrity. We found some evidence that more recently collected material could produce successful sequences, but most specimens failed to yield sequences of sufficient quality to be useful for creating a barcode reference library consistent with the global consensus for standard DNA-barcode reference sequences, i.e., sequence length >500 bp, with <1% ambiguous base calls (Consortium for the Barcode of Life; www.barcoding.si.edu).

On the other hand, we had better results when obtaining mini-barcode sequences. The ability to obtain mini-barcodes from legacy specimen libraries is a significant result because this ability could be critical for verifying specimens of rare species. Mini-barcodes appear to be effective for resolving ~91% of species (Hajibabaei et al. 2006, Meusnier et al. 2008). They can be used to link older reference material (e.g., type specimens) to newly collected material of the same species. Thus, the full-length barcode of the newly collected species can be linked to material that has been used for developing taxonomic keys and for species description, and degraded legacy material can be linked to material from the present-day (an example of this situation in Lepidoptera is presented in Meusnier et al. 2008).

In contrast to the poor results for the reference-collection material, success in sequencing formalin-preserved fresh material from 4 laboratory-cultured freshwater macroinvertebrate species was remarkably high, even after 20 d of exposure to formalin. No evidence was found that inclusion of sediment material adversely affected sequence quality. The success rate for PCR amplification was slightly lower for the mayfly larva, *H. limbata*. This result probably was caused by an issue with primer compatibility. Thus, the overall performances of PCR and sequencing might be further improved with additional efforts with various primer sets.

Preservation in DNA-friendly media, such as 95% ethanol, is preferable when samples are to be sequenced for DNA, but preservation in ethanol is not practical in many situations. The results of our pilot experiment indicate that aquatic macroinvertebrate specimens collected by routine methods involving short exposures to formalin in bioinventory and biomonitoring programs can provide good source material for DNA barcoding without a requirement for significant modification of the existing biomonitoring protocol. However, for voucher specimens, where full-length barcode sequences are required, preservation in ethanol is essential.

Our results are far from a comprehensive demonstration that formalin is not the culprit in degradation of DNA in legacy material, but they certainly point in that direction. However, maintenance of specimens in 70% ethanol, even in an otherwise well-managed collection, will not ensure DNA preservation. We recommend preserving larval material in $\geq 95\%$ ethanol, preferably in a -10°C freezer.

Acknowledgements

Our study was supported by funds from Genome Canada through the Ontario Genomics Institute. The authors thank Environment Canada for providing access to the CABIN reference collection, Corinne Siefker and Heather McDermott for pulling specimens and providing reference-collection statistics, and Danielle Milani, Sherri Thompson, and Jennifer Webber for help in accessing invertebrate culture material.

Literature Cited

- BAIRD, D. J., M. N. RUBACH, AND P. J. VAN DEN BRINK. 2008. Trait-based ecological risk assessment: the new frontier? *Integrated Environmental Assessment and Management* 4:2–3.
- BEANE FREEMAN, L. E., A. BLAIR, J. H. LUBIN, P. A. STEWART, R. B. HAYES, R. N. HOOVER, AND M. HAUPTMANN. 2009. Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute cohort. *Journal of the National Cancer Institute* 101:751–761.
- BORGMANN, U., K. M. RALPH, AND W. P. NORWOOD. 1989. Toxicity test procedures for *Hyalella azteca* and chronic toxicity of cadmium and pentachlorophenol to *H. azteca*, *Gammarus fasciatus*, and *Daphnia magna*. *Archives of Environmental Contamination and Toxicology* 18:756–764.
- ESPELAND, M., M. IRESTEDT, K. A. JOHANSON, M. ÅKERLUND, J.-E. BERGH, AND M. KÄLLERSJÖ. 2010. Dichlorvos exposure impedes extraction and amplification of DNA from insects in museum collections. *Frontiers in Zoology* 2010:7. doi:10.1186/1742-9994-7-2
- FOLMER, O., M. BLACK, W. HOEH, R. LUTZ, AND R. VRIJENHOEK. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3:294–299.
- HAJIBABAEI, M., M. A. SMITH, D. H. JANZEN, J. J. RODRIGUEZ, J. B. WHITFIELD, AND P. D. N. HEBERT. 2006. A minimalist barcode can identify as specimen whose DNA is degraded. *Molecular Ecology Notes* 6:959–964.
- HEBERT, P. D. N., A. CYWINSKA, S. L. BALL, AND J. R. DEWAARD. 2003. Biological identification through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* 270:313–321.
- HEBERT, P. D. N., E. H. PENTON, J. M. BURNS, D. H. JANZEN, AND W. HALLWACHS. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101:14812–14817.
- HOPKINS, G. W., AND R. P. FRECKLETON. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* 5: 245–249.
- IVANOVA, N. V., J. R. DEWAARD, AND P. D. N. HEBERT. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* 6: 998–1002.
- MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24: 133–141.
- MEUSNIER, I., G. A. C. SINGER, J. F. LANDRY, D. A. HICKEY, P. D. N. HEBERT, AND M. HAJIBABAEI. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9:214. doi:10.1186/1471-2164-9-214
- MILANI, D., T. B. REYNOLDS, U. BORGMANN, AND J. KOLASA. 2003. The relative sensitivity of four benthic invertebrates to metals in spiked-sediment exposures and application to contaminated field sediment. *Environmental Toxicology and Chemistry* 22:845–854.
- REYNOLDS, T. B., C. LOGAN, D. MILANI, T. PASCOE, AND S. P. THOMPSON. 1998. Methods manual IV: sediment toxicity testing, field and laboratory methods and data management. NWRI Report No. 99-12. Environment Canada, Burlington, Ontario, Canada.
- REYNOLDS, T. B., S. P. THOMPSON, AND J. L. RAMSEY. 1991. A sediment bioassay using the tubificid oligochaete worm *Tubifex tubifex*. *Environmental Toxicology and Chemistry* 10:1061–1072.
- SINGER, G. A. C., AND M. HAJIBABAEI. 2009. iBarcode.org: web-based molecular biodiversity analysis. *BMC Bioinformatics* 10(Suppl 6):S14.
- SKAGE, M., AND C. SCHANDER. 2007. DNA from formalin-fixed tissue: extraction or repair? That is the question. *Marine Biology Research* 3:289–295.
- TAMURA, K., J. DUDLEY, M. NEI, AND S. KUMAR. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software. Version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.

ZHOU, X., S. J. ADAMOWICZ, L. M. JACOBUS, R. E. DEWALT, AND P. D. N. HEBERT. 2009. Towards a comprehensive barcode library for arctic life—Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology* 6:30.

ZIMMERMANN, J., M. HAJIBABAEI, D. C. BLACKBURN, J. HANKEN, E. CANTIN, J. POSFAI, AND T. C. EVANS. 2008. DNA damage in

preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology* 5:18.

Received: 7 February 2010

Accepted: 25 October 2010