Documenting distributions of hyperdiverse groups (terrestrial invertebrates) through grid-based spatial sampling, high-throughput sequencing, and occupancy modeling

# Author Information

**Matthew L. Bowser,\* John M. Morton**

**M.L. Bowser** USFWS Kenai National Wildlife Refuge, Soldotna, Alaska 99669

**J.M. Morton** USFWS Kenai National Wildlife Refuge, Soldotna, Alaska 99669 (retired)

\*Corresponding author: matt\_bowser@fws.gov

# Introduction

Alaska National Wildlife Refuges have been given mandates to conserve natural biodiversity, habitats, biological integrity, and environmental health (96th Congress [1980](#ref-96th_Congress_1980); 105th Congress [1997](#ref-105th_Congress_1997)). Woodward and Beever ([2010](#ref-Woodward_Beever_2010)) noted that, under these broad purposes of National Wildlife Refuges in Alaska to protect natural landscapes and entire ecosystems, they must develop methods that can successfully monitor biodiversity.

Morton et al. ([2009](#ref-Morton_et_al_2009)) determined that the best way to measure biodiversity at the scale of Alaska Refuges was to document occurrences and co-occurrences of many species on a set of points distributed widely across the landscape. This kind of sampling scheme would yield species distributions and assemblages and, if repeated over time, could be used to assess whether Alaska Refuges were fulfilling their conservation purposes. Magness et al. ([2008](#ref-Magness_et_al_2008)) demonstrated the utility of this sampling design for generating species distribution data that could serve this purpose.

A key deficiency identified by Morton et al. ([2009](#ref-Morton_et_al_2009)) was an inability to feasibly monitor most of biodiversity due to the high cost and lengthy time required to obtain identifications for species in hyperdiverse groups, a well-known problem that has been referred to as the taxonomic impediment (Taylor [1983](#ref-Taylor_1983)). Since that time, high-throughput sequencing (HTS) methods have emerged as a means to obtain many identfications from environmental samples. Multiple studies (Gibson et al. [2015](#ref-Gibson_et_al_2015); Hajibabaei et al. [2016](#ref-Hajibabaei_et_al_2016); Bowser et al. [2017](#ref-Bowser_et_al_2017), [2020](#ref-Bowser_et_al_2020); Bush et al. [2019](#ref-Bush_2019)) have demonstrated that assemblages of diverse taxa can be efficiently identified using these methods.

The second deficiency recognized by Morton et al. ([2009](#ref-Morton_et_al_2009)) and Magness et al. ([2008](#ref-Magness_et_al_2008)) was the inability to account for imperfect detection. When a species is present but there is a chance that it will not be detected, this may lead to false negatives (i.e., recording an absence where a species is present), leading to bias in the resulting estimates of species distributions. Mackenzie et al. ([2002](#ref-MacKenzie_et_al_2002), [2006](#ref-MacKenzie_et_al_2006)) presented occupancy modeling methods that can yield unbiased estimates of species occurrence and distributions. These models explicitly account for imperfect detection through spatially or temporally repeated sampling.

The biological inventory work of Bowser et al. ([2020](#ref-Bowser_et_al_2020)) was designed in part to provide the kind of occurrence data required for monitoring distributions of multiple species (i.e., monitoring assemblages of species) and accounting for imperfect detection by spatial subsampling. Our intent in this paper is to test the usefulness of the methods of Bowser et al. ([2020](#ref-Bowser_et_al_2020)) for delivering unbiased estimates of species occurrence and distributions of hyperdiverse taxa.

# Study Site

The study site was described by Bowser et al. ([2020](#ref-Bowser_et_al_2020)). The study site was a portion of the Slikok Creek watershed on the Kenai National Wildlife Refuge (KNWR). The resulting 938 ha study area occupied a bounding box from 60.44° to 60.47° latitude and from -151.10° to -151.03° longitude.

Based on data from the National Land Cover Database (Homer et al. [2015](#ref-Homer_et_al_2015)), the land area of this study area was coverd mostly by mixed forest (29%), evergreen forest (22%), emergent herbaceous wetlands (13%), deciduous forest (12%), woody wetlands (11%), and open water (10%). Common trees in well-drained areas were white spruce (*Picea glauca* (Moench) Voss), Alaska birch (*Betula pendula* subsp. *mandshurica* (Regel) Ashburner & McAll.), quaking aspen (*Populus tremuloides* Michx.), and cottonwood (*Populus* × *hastata* Dode). Black spruce (*Picea mariana* Britton, Sterns & Poggenb.) was the dominant tree in forested wetlands. Herbaceous wetlands were characterized by mosses, sedges, and low shrubs. Shallow, eutrophic lakes (Headquarters Lake, Nordic Lake, and smaller, unnamed lakes) accounted for most of the open water in the study area.

# Methods

In the time since Bowser et al. ([2020](#ref-Bowser_et_al_2020)) was published, a handfull of identifications of the records on Arctos have been improved based on new data that became available in the reference databases (BOLD and Arctos).

A summary of all metabarcoding-based occurrences from sweep net samples obtained by Bowser et al. ([2020](#ref-Bowser_et_al_2020)) were downloaded from Arctos on 12 November 2020 (saved search URI: <https://arctos.database.museum/saved/2020-11-12-1157_Slikok_metabarcoding_taxa_list>), yielding a list of 975 unique identifications. For unidentified molecular operational taxonomic units (MOTUS) where 10 or more occurrences were observed, we attempted to improve identifications by submitting the corresponding sequences to BOLD’s Identification Engince and NCBI BLASTn. We obtained no new identifications.

The 2,375 occurrences were downloaded on 12 November 2020 (saved search URI: <https://arctos.database.museum/saved/2020-11-12-1400_Slikok_project_metabarcoding_occurrences>).

# References

105th Congress. 1997. 111 Stat. 1252 - National Wildlife Refuge System Improvement Act of 1997. Pages 1252–1260 in United States Statutes at Large, Volume 111, 105th Congress, 1st Session. Available from <https://www.govinfo.gov/app/details/STATUTE-111/STATUTE-111-Pg1252>.

96th Congress. 1980. Alaska National Interest Lands Conservation Act. in United States Statutes at Large, Volume 94, 96th Congress, 2nd Session. U.S. Government Printing Office. Available from <https://www.govinfo.gov/app/details/STATUTE-94/STATUTE-94-Pg2371>.

Bowser ML, Brassfield R, Dziergowski A, Eskelin T, Hester J, Magness DR, McInnis M, Melvin T, Morton JM, Stone J. 2020. Towards conserving natural diversity: A biotic inventory by observations, specimens, DNA barcoding and high-throughput sequencing methods. Biodiversity Data Journal **8**:e50124.

Bowser ML, Morton J, Hanson J, Magness D, Okuly M. 2017. Arthropod and oligochaete assemblages from grasslands of the southern Kenai Peninsula, Alaska. Biodiversity Data Journal **5**:e10792.

Bush A, Compson ZG, Monk WA, Porter TM, Steeves R, Emilson E, Gagne N, Hajibabaei M, Roy M, Baird DJ. 2019. Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. Frontiers in Ecology and Evolution **7**:434.

Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M. 2015. Large-scale biomonitoring of remote and threatened ecosystems via High-Throughput Sequencing. PLOS One **10**:e0138432.

Hajibabaei M, Baird DJ, Fahner NA, Beiko R, Golding GB. 2016. A new way to contemplate Darwin’s tangled bank: How DNA barcodes are reconnecting biodiversity science and biomonitoring. Philosophical Transactions of the Royal Society of London B: Biological Sciences **371**:20150330.

Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold ND, Wickham JD, Megown K. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing **81**:345–354. Available from <https://www.asprs.org/a/publications/pers/2015journals/PERS_May_2015/HTML/index.html#345>.

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology **83**:2248–2255.

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. 2006. Occupancy Estimation and Modeling. Elsevier, New York.

Magness DR, Huettmann F, Morton JM. 2008. Using Random Forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. Pages 209–229 in T. G. Smolinski, M. G. Milanova, and A.-E. Hassanien, editors. Applications of Computational Intelligence in Biology: Current Trends and Open Problems. Springer-Verlag, Berlin.

Morton JM, Bowser ML, Berg E, Magness D, Eskelin T. 2009. Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge, Alaska: An FIA adjunct inventory. Pages 1–17 in W. McWilliams, G. Moisen, and R. Czaplewski, editors. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado. Available from <http://www.treesearch.fs.fed.us/pubs/33332>.

Taylor RW. 1983. Descriptive taxonomy: Past, present, and future. Pages 93–134 in E. Highley and R. W. Taylor, editors. Australian Systematic Entomology: A Bicentenary Perspective. CSIRO, Canberra, Australia. Available from <http://hdl.handle.net/102.100.100/286670?index=1>.

Woodward A, Beever EA. 2010. Framework for ecological monitoring on lands of Alaska National Wildlife Refuges and their partners. Page 94. Report 2010-1300. Reston, Virginia. Available from <http://pubs.er.usgs.gov/publication/ofr20101300>.

# Figures and Figure Captions