

Towards conserving natural diversity: A biotic inventory by observations, specimens, DNA barcoding and high-throughput sequencing methods

Matthew Lewis Bowser[‡], Rebekah Brassfield[§], Annie Dziergowski[|], Todd Eskelin[‡], Jennifer Hester[¶], Dawn Robin Magness[‡], Mariah McInnis[#], Tracy Melvin[»], John M. Morton[«], Joel Stone[»]

[‡] U.S. Fish & Wildlife Service, Kenai National Wildlife Refuge, Soldotna, Alaska, United States of America

[§] Salish Kootenai College, Pablo, Montana, United States of America

[|] U.S. Fish & Wildlife Service, North Florida Ecological Services Office, Jacksonville, Florida, United States of America

[¶] City of Soldotna, Planning and Zoning Commission, Soldotna, Alaska, United States of America

[#] Auburn University, School of Forestry & Wildlife Sciences, Auburn, Alabama, United States of America

[»] Michigan State University, College of Agriculture & Natural Resources, Department of Fisheries and Wildlife, East Lansing, Michigan, United States of America

[«] U.S. Fish & Wildlife Service (retired), Soldotna, Alaska, United States of America

[»] University of Alaska Fairbanks, Fairbanks, Alaska, United States of America

Corresponding author: Matthew Lewis Bowser (matt_bowser@fws.gov)

Academic editor: John-James Wilson

Abstract

The Kenai National Wildlife Refuge has been given a broad conservation mandate to conserve natural diversity. A prerequisite for fulfilling this purpose is to be able to identify the species and communities that make up that biodiversity.

We tested a set of varied methods for inventory and monitoring of plants, birds and terrestrial invertebrates on a grid of 40 sites in a 938 ha study area in the Slikok Creek watershed, Kenai Peninsula, Alaska. We sampled plants and lichens through observation and specimen-based methods. We surveyed birds using bird call surveys on variable circular plots. We sampled terrestrial arthropods by sweep net sampling, processing samples with High Throughput Sequencing methods. We surveyed for earthworms, using the hot mustard extraction method and identified worm specimens by morphology and DNA barcoding. We examined community membership using clustering methods and Nonmetric Multidimensional Scaling.

We documented a total of 4,764 occurrences of 984 species and molecular operational taxonomic units: 87 vascular plants, 51 mosses, 12 liverworts, 111 lichens, 43 vertebrates, 663 arthropods, 9 molluscs and 8 annelid worms. Amongst these records, 102 of the arthropod species appeared to be new records for Alaska. We found three non-native species: *Deroceras agreste* (Linnaeus, 1758) (Stylommatophora: Agriolimacidae),

Dendrobaena octaedra (Savigny, 1826) (Crassiditellata: Lumbricidae) and *Heterarthrus nemoratus* (Fallén, 1808) (Hymenoptera: Tenthredinidae). Both *D. octaedra* and *H. nemoratus* were found at sites distant from obvious human disturbance. The 40 sites were grouped into five community groups: upland mixed forest, black spruce forest, open deciduous forest, shrub-sedge bog and willow.

We demonstrated that, at least for a subset of species that could be detected using these methods, we were able to document current species distributions and assemblages in a way that could be efficiently repeated for the purposes of biomonitoring. While our methods could be improved and additional methods and groups could be added, our combination of techniques yielded a substantial portion of the data necessary for fulfilling Kenai National Wildlife Refuge's broad conservation purposes.

Keywords

biomonitoring, metabarcoding, vegetation, birds, terrestrial invertebrates, earthworms

Introduction

In order to conserve global biodiversity, given current and expected realities of species distribution shifts, novel assemblages and potential extinctions, we must be able to routinely document species distributions and assemblages. Historically, this has not been feasible except for a small set of large, easily-recognised species because of the taxonomic impediment—the difficulty, time and expense of identifying species from hyperdiverse groups (see Goldstein and DeSalle 2010).

High-Throughput Sequencing (HTS) methods have been advocated as a means of overcoming the taxonomic impediment, enabling identifications of species from diverse groups in mixed environmental samples (Hajibabaei et al. 2011, Baird and Hajibabaei 2012, Yu et al. 2012, Aylagas et al. 2016, Lobo et al. 2017, Porter and Hajibabaei 2018b, Bush et al. 2019, Watts et al. 2019). These HTS methods have recently been put into practice for biomonitoring of species assemblages of non-marine invertebrates (Gibson et al. 2015, Hajibabaei et al. 2016, Bush et al. 2019).

In this study, we tested the ability of several methods to rapidly determine species assemblages on a portion of a National Wildlife Refuge in Southcentral Alaska for fulfilling the broad conservation purposes of that refuge.

The United States Congress mandated that all Alaska National Wildlife Refuges must, "conserve fish and wildlife populations and habitats in their natural diversity," (96th Congress 1980); "provide for the conservation of fish, wildlife, and plants, and their habitats...", (105th Congress 1997); and, "ensure that ... biological integrity, diversity, and environmental health ... are maintained for the benefit of present and future generations..." (105th Congress 1997). Woodward and Beever (2010) noted that, under these broad purposes of National Wildlife Refuges in Alaska to protect natural landscapes

and entire ecosystems, they "must develop a monitoring program to assess whether protective management is successfully conserving something as complex and open-ended as biodiversity."

Alaska Maritime National Wildlife Refuge and Kenai National Wildlife Refuge were given an additional purpose of providing opportunities for scientific research (96th Congress 1980).

Morton et al. (2009) recognised that a prerequisite for fulfilling the broad conservation mandates of Alaska National Wildlife Refuges was to document species distributions and assemblages. Given the remoteness of most of the land area of Alaskan Refuges, generally requiring access by aircraft and a need to obtain sample sizes large enough to make meaningful inferences about species distributions and communities over these land areas, they selected methods that would enable sampling of a large number of sites over a short time with minimal travel cost. Ensuring that all sampling methods could be executed in under 1 hr per visit enabled up to six remote sites to be sampled per crew per day.

These methods delivered meaningful metrics useful for accomplishing the Alaska National Wildlife Refuges' conservation and research purposes, but two key deficiencies were identified. First, the lack of spatially or temporally repeated sampling led to an inability to account for imperfect detection probabilities (as defined by MacKenzie et al. 2002, MacKenzie et al. 2005). Second, it was clear from the effort and time required to identify invertebrates using morphological methods that it would not be practical to repeatedly sample using these methods. In order to enable identifications of species by DNA barcoding, the U.S. Fish & Wildlife Service, Alaska Region supported development of a DNA barcode library of Alaskan terrestrial arthropods (Sikes et al. 2017a), setting the stage for biomonitoring using HTS methods.

In the current effort, we tested biomonitoring methods similar to those used by Morton et al. (2009), with the goal of maintaining efficient inventory and monitoring techniques for species assemblages, while addressing previous methodological shortcomings. We subsampled spatially to account for imperfect detection and we identified invertebrates using HTS methods. In this article, we describe our methods and the resulting biodiversity data. We intend to assess our ability to account for imperfect detection in a subsequent paper.

Materials and methods

Study area

Our study area comprised a section of the Slikok Creek watershed, a well-defined geographic region representative of the lowlands of the western Kenai Peninsula and the watershed. We determined the Slikok Creek watershed boundary (HUC12 code: 190203021804) using the national Watershed Boundary Dataset (U.S. Geological Survey 2016).

The 5,917 ha Slikok Creek watershed originates on the Kenai National Wildlife Refuge (KNWR) with headwaters in the wetlands and hills around Headquarters Lake, Nordic Lake and Slikok Lake. Streams from these lakes and wetlands coalesce into Slikok Creek, which then flows through a mosaic of public and private lands before joining the Kenai River.

We limited our study area to the part of the watershed within the KNWR. We also restricted our study area to the part of the watershed north of 60.44° latitude to eliminate long walking distances required to access the more remote southern portion of the watershed. This yielded a study area with boundaries not more than roughly 3 km from established roads and trails. 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 (Fig. 1, Suppl. material 1).

Sampling design

For an initial field inventory, a 500 m grid was chosen by using the coordinates of the centroids of the 250 m pixels from the Alaska eMODIS product (Jenkerson et al. 2010), choosing every other centroid to make a grid of sites having 500 m spacing. The resulting sample frame consisted of 42 sites (Fig. 1, Suppl. material 1). This sample frame was representative of the study area in terms of proportions of cover classes (Fig. 2), based on cover classes from the National Land Cover Database (NLCD, Homer et al. 2015). For the present study, we excluded the two aquatic sites, leaving 40 terrestrial sites.

Field methods

As in Morton et al. (2009), our interest was in methods that could be executed rapidly, requiring all sampling to take less than one hour per visit. This precluded consideration of any methods that would have required passive sampling (e.g. sound recording devices, camera traps, malaise traps, pitfall traps and pan traps).

Plot establishment

Sampling sites were marked by driving 122 cm long, 13 mm diameter SunGUARD Smart Stake™ fibreglass rods into the ground, then labelling them with aluminium tags (Fig. 3). During the survey period, sites were also temporarily marked with high-visibility forestry flagging tape.

Plants and lichens

Vegetation was sampled from 18 July to 10 August 2016. We recorded presence within the 5.64 m radius circular plot and species identity of all vascular plants, bryophytes and lichens that could be identified in the field. In some cases, plants were collected from outside the plot to be identified in the lab. In addition, representatives of all bryophyte and lichen species present on the plots were collected from outside of the plot for subsequent identification. We collected from outside of plots to maintain integrity of the plots for the purpose of potential long-term monitoring.

Bird call surveys

Breeding bird calls were sampled from 14 June to 17 June 2016. We sampled bird abundance and occurrence using variable circular plot methods adapted from the Alaska Landbird Monitoring System (ALMS) protocol (Handel and Cady 2004, Morton et al. 2009), identifying birds in the field. Horizontal distances to each bird were estimated at 1 min increments during a 10 min sampling interval using auditory or visual cues. Surveys were conducted 30 min after sunrise during the first 4-5 h of the morning.

Earthworms

Earthworms were sampled concurrently with vegetation sampling, using the hot mustard extraction method (Lawrence and Bowers 2002). At each site, a quadrat was selected from the 5.64 m radius circular plot, but within about 20 m from the circular plot. Where available, we selected sites where ample hardwood leaf litter was available to increase the probability of detecting earthworms. A 0.5 m square (0.25 m²), metal frame was used as a quadrat. Within this quadrat, litter was first removed carefully by hand and all worms found were collected. Next, about 2 l of mustard powder mixture (30 g yellow mustard powder and 3.8 l of water) were poured evenly over the quadrat. Worms were collected for the next 5 min as they emerged from the soil, then the hot mustard extraction was repeated using the remaining mustard powder mixture.

Sweep net samples of terrestrial arthropods

Sweep net samples of terrestrial arthropods were collected concurrently with the bird surveys from 14 June to 17 June 2016. A second set of sweep net samples was collected when the plots were revisited to sample vegetation on 18 July to 9 August 2016. A total of 160 sweep net samples were collected (40 plots × 2 samples/plot × 2 visits/plot).

Arthropods were sampled within a 100 m², 5.64 m radius, circular plot using the centre stake as plot centre. To enable comparison with the previous work of Morton et al. (2009), we used the same methods except that we subsampled spatially. We split the plot into two subplots, dividing along the north-south axis. Each semicircular subplot was independently sweep-netted, such that the entire area was swept from the ground surface up to a height of roughly 2 m. No defined pattern of sweeping was enforced, but we ensured that all substrates and macrohabitats within reach were swept over once within a time limit of 5 min per sample. We used a BioQuip™ model 7112CP 30.5 cm diameter net with a BioQuip™ model 7312AA 30.5 cm extension handle and a BioQuip™ model 7112CPA net bag with a mesh size of approximately 8 × 9 meshes/mm.

All specimens were collected into a single Nalgene® model 2104-0008 wide-mouth 250 ml bottle containing UniGard -100 propylene glycol antifreeze. Even though we could have used ethanol as a preservative, we chose propylene glycol because its non-flammability makes it much safer than ethanol for helicopter operations, which would be required for biological inventories over much of Alaska National Wildlife Refuges. We had tested the

use of propylene glycol for samples, intended to be processed by HTS methods, in previous studies (Bowser et al. 2017, Bowser et al. 2019).

Laboratory methods

Observation data and specimens were processed using methods that varied for each taxonomic group and sample type. A graphical summary of all methods used is provided in Fig. 4.

Plants and lichens

Vascular plants, that could not be identified in the field, were identified in the lab using pertinent keys (Hultén 1968, Welsh 1974, Tande and Lipkin 2003). Vascular plant specimens were not retained.

Lichen and bryophyte samples were sent to Trevor Goward (Enrichened Consulting Ltd., Clearwater, British Columbia) for identification. Specimens were identified by Trevor Goward and Curtis Björk (Suppl. material 3). Most specimens were discarded, but some lichen and bryophyte specimens were retained by Trevor Goward.

Plant and lichen data were entered into Arctos (<https://arctosdb.org/>) as observation data.

Bird call survey data

Bird call survey data were entered into Arctos as observation records.

Earthworms

Worm specimens were deposited in the Kenai National Wildlife Refuge's entomology collection*¹, where data are managed and published through Arctos. Worm specimens were sorted to morphospecies. Adult earthworms (Lumbricidae) were identified morphologically using the key of Reynolds (1977). Representatives of each perceived morphospecies of enchytraeids and immature lumbricids in each sample were mailed to the Center for Biodiversity Genomics (Guelph, Ontario) for COI sequencing using LifeScanner kits (<http://lifescanner.net>).

Sweep net samples of terrestrial arthropods

Arthropods and any other invertebrates in the sweep net samples were separated from debris by hand under a stereomicroscope. All fragments of invertebrates were retained. Samples were stored in a -23°C freezer until they were shipped out for sequencing.

Due to budget limitations, we processed 125 of the 160 sweep net samples. We selected all 80 samples taken from the east side of each plot (40 plots × 1 sample/plot × 2 visits/plot). To choose 45 samples from the remaining 80, we selected plots spatially. First, we chose 20 samples from plots at 1 km spacing (10 plots × 2 visits/plot), then we chose 25 of

26 samples from another 13 plots that were maximally distant from these 10 plots (13 plots × 2 visits/plot). These 45 samples from west plot halves were intended to be used for estimating occupancy metrics.

Sweep net samples were shipped to RTL Genomics (<http://rtlgenomics.com>) for extraction and sequencing steps. DNA extraction methods are included in Suppl. material 2.

Sequencing was performed on an Illumina MiSeq platform and reads were processed using RTL Genomics' standard methods with the *mCOLintF/HCO2198* primer set of Leray et al. (2013), yielding a 313 bp region of the COI gene. We selected this primer set because it has been shown to amplify well across a broad set of arthropod groups (Brandon-Mong et al. 2015, Hajibabaei et al. 2019).

Samples were amplified for sequencing in a two-step process. The forward primer was constructed (5'-3') with the forward Illumina overhang adapter (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) added to the *mCOLintF* primer (GGWACWGGWTGAACWGTWTAYCCYCC). The reverse primer was constructed (5'-3') with the reverse Illumina overhang adapter (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) added to the *HCO2198* primer (TAACTTCAGGGTGACCAAAAAATCA). Amplifications were performed in 25 µl reactions with Qiagen HotStar Taq master mix (Qiagen Inc, Valencia, California), 1 µl of each 5 µM primer and 1 µl of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosystems, Carlsbad, California) under the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 s, 54°C for 40 s, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold.

Products from the first stage amplification were added to a second PCR, based on qualitatively determined concentrations. Primers for the second PCR were designed, based on the Illumina Nextera PCR primers as follows: Forward - AATGATACGGCGACCACCGAGATCTACAC[i5index]TCGTCGGCAGCGTC and Reverse - CAAGCAGAAGACGGCATACGAGAT[i7index]GTCTCGTGGGCTCGG. The second stage amplification was run the same as the first stage except for 10 cycles. Amplification products were visualised with eGels (Life Technologies, Grand Island, New York). Products were then pooled equimolarly and each pool was size-selected in two rounds using SPRIselect Reagent (BeckmanCoulter, Indianapolis, Indiana) in a 0.75 ratio for both rounds. Size-selected pools were then quantified using the Qubit 4 Fluorometer (Life Technologies) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, California) 2 × 300 flow cell at 10 pM.

Our metagenomic analysis was carried out on the Yeti Supercomputer (USGS Advanced Research Computing 2019). We used the SCVUC COI metabarcoding pipeline (https://github.com/EcoBiomics-Zoobiome/SCVUC_COI_metabarcoding_pipeline) except that we used neither the RDP Classifier (Wang et al. 2007) nor the CO1 classifier (Porter and Hajibabaei 2018a) for taxonomic assignments. We intentionally selected a metagenomics pipeline that preserved all Amplicon Sequence Variants (ASVs) for maximum reusability and comprehensiveness of the data derived from this project (see Callahan et al. 2017).

Forward and reverse reads were paired with SeqPrep (<https://github.com/jstjohn/SeqPrep>) using the default settings of a minimum quality of Phred of 20 and an overlap of at least 25 bp. We removed forward and reverse primers with cutadapt v2.3 (Martin 2011, <http://cutadapt.readthedocs.io/en/stable/index.html>), accepting default settings but requiring a minimum length after trimming of at least 150 bp, minimum read quality of Phred 20 at the ends of the sequences and allowing a maximum of 3 Ns. We de-replicated FASTA files using VSEARCH 2.4.3 (Rognes et al. 2016). We de-noised reads using USEARCH v11 (Edgar 2010) with the UNOISE3 algorithm (Edgar 2016) specifying a minimum abundance of 3. The resulting ASV table and ASV sequences are provided in Suppl. material 4 and Suppl. material 5.

We made initial taxonomic assignments to all Amplicon Sequence Variants (ASVs) using the `bold_identify` command of the `bold` package version 0.8.6 (Chamberlain 2018) in R version 3.5.1 (R Core Team 2018), using the COX1 reference dataset of BOLD (Ratnasingham and Hebert 2007). In all cases of potential new distribution records, we manually scrutinised the results of BOLD Identification Engine (http://boldsystems.org/index.php/IDS_OpenIdEngine) and NCBI Nucleotide BLAST (Altschul et al. 1990) searches.

For consistency, clarity and transparency in the use of provisional names, we followed the standards of Open Nomenclature (Sigovini et al. 2016) in assigning identifications. Amplicon Sequence Variants that could not be confidently assigned to described species were assigned to BOLD Barcode Index Numbers (BINs, Ratnasingham and Hebert 2013). Amplicon Sequence Variants that could be assigned to neither species nor BINs were given provisional names including ASV labels, e.g. "*Liriomyza* sp. SlikokOtu253".

In order to exclude potential false positive detections as defined by MacKenzie et al. (2002) and MacKenzie et al. (2005) due to demultiplexing errors, we conservatively removed from the ASV table all occurrences that represented less than 0.05% of the total number of reads for any ASV, based on assuming a 0.01% to 0.03% rate of mis-assignment of reads (Deiner et al. 2017). We also removed all occurrences represented by only one or two reads.

We removed sequences from fungi, bacteria, red algae and humans by first constructing a phylogenetic tree using qiime phylogeny align-to-tree-mafft-fasttree (Lane 1991, Price et al. 2010, Katoh and Standley 2013, Bolyen et al. 2018), then pruning off branches of non-target groups on iTOL (Letunic and Bork 2019). An interactive version of this tree is available on iTOL at <https://itol.embl.de/tree/1641591522370151564097588>. Finally, we filtered the ASV table, based on the ASVs retained in the pruned tree.

Data analysis

For the purposes of analyses, we considered BIN identifications to be species-resolution identifications. We also removed occurrence records greater than 200 m from plot centres, consistent with the 200 m cut-off used by Morton et al. (2009) in the analysis dataset. To

make sampling effort consistent across all plots, we removed all sweep-net samples taken from the west side of plots.

Analyses and plotting were performed under R, version 3.5.1 (R Core Team 2018) using the packages DiagrammeR, version 1.0.5 (Iannone 2020); GISTools, version 0.7-4 (Brunsdon and Chen 2014); maptools, version 0.9-4 (Bivand and Lewin-Koh 2018); phytools, version 0.6-99 (Revell 2012); raster, version 2.8-4 (Hijmans 2018); recluster, version 2.8 (Dapporto et al. 2015); reshape2, version 1.4.3 (Wickham 2007); rgdal, version 1.3-6 (Bivand et al. 2018); rgeos, version 0.4-2 (Bivand and Rundel 2018); vegan, version 2.5-3 (Oksanen et al. 2018); and VennDiagram, version 1.6.20 (Chen 2018).

To compare the distribution of non-native species detected in this study to the previously known distributions of non-native species in our study area, we generated a map of non-native species records. We downloaded non-native plant occurrences from the Alaska Exotic Plants Information Clearinghouse (AKEPIC 2016). We also downloaded all occurrence records for the study area from the Global Biodiversity Information Facility (GBIF 2016) and selected names that were recognised as non-native in Alaska by Simpson et al. (2019).

The total numbers of species in each phylum were estimated using the Chao estimator (Chao 1987, Chiu et al. 2014) implemented by the specpool function of the vegan package. We generated species accumulation curves using the specaccum function from the vegan package.

To classify observed communities, we first removed all rare species that were detected on less than 5% of plots, yielding an observation matrix of 40 sites and 415 species. From this, we generated a UPGMA (unweighted pair group method with arithmetic mean, Sokal and Michener 1958) consensus tree using the recluster.cons command of the recluster package, accepting default values, except that the number of trees was set to 1,000 and the Jaccard index (Jaccard 1901) was used for calculating distances. We obtained bootstrap values for nodes of the tree with the recluster.boot command of the recluster package, generating 1,000 trees with the same parameters as the original tree.

To examine community relationships, we used Nonmetric Multidimensional Scaling (NMDS) by running the same observation matrix of 40 sites and 415 species through the metaMDS function of the vegan package. As in the clustering analysis, we used the Jaccard index as the distance measure. We set the number of dimensions to two because adding more dimensions decreased the stress only slightly. To determine community membership, all species detected at 25% or more of sites in a community were assigned to that community.

Data publishing

We sought to follow the guidelines of Penev et al. (2017) for publication of biodiversity data. Field notebooks, field data sheets, laboratory notebooks and occurrence data have been made available via Arctos (<https://arctosdb.org/>) and are associated together via an

Arctos project (<http://arctos.database.museum/project/10002227>). These occurrence data on Arctos have been published to GBIF via the VertNet IPT (<http://ipt.vertnet.org/>). The occurrence data used in this analysis are provided in Suppl. material 6 and Suppl. material 9.

Sequence data from worm specimens sequenced using LifeScanner kits have been made publicly available through BOLD (<http://boldsystems.org/>). Raw sequence data from the sweep net samples of terrestrial arthropods have been published via Zenodo (Bowser 2019) and to GenBank's Sequence Read Archive in accessions [SRR10454582–SRR10454706](#) under BioProject [PRJNA427721](#).

Results

Occurrences

Collectively, 4,764 catalogued occurrence records were generated (Suppl. material 6), of which 4,703 were within 200 m from plot centres. A total of 710 formally described species were documented. An additional 274 ASVs were identified with BIN identifications, making a total of 984 species or BIN identifications (Fig. 5). From this point on, all numbers reported and figures include BIN identifications as "species". We documented 49 to 137 species per site (mean = 88, Fig. 6).

Of the 397 described arthropod species documented, 102 (26% of the described arthropod species found) appear to be newly reported from Alaska (Suppl. material 7). The new records included 65 Diptera, 14 Hymenoptera, 11 Hemiptera, 8 Lepidoptera, 2 Neuroptera and 1 species of Psocodea and Araneae. Five species (*Allodia czernyi* (Landrock, 1912) (Diptera: Mycetophilidae); *Exechia parva* Lundström, 1909 (Diptera: Mycetophilidae); *Idiocerus elegans* Flor, 1861 (Hemiptera: Cicadellidae); *Rymosia pinnata* Ostroverkhova, 1979 (Diptera: Mycetophilidae); and *Zygoneura sciarina* Meigen, 1830 (Diptera: Sciaridae)) appear to be new records for North America.

We detected three non-native species: *Deroceras agreste* (Linnaeus, 1758) (Stylommatophora: Agriolimacidae), *Dendrobaena octaedra* (Savigny, 1826) (Crassieclitellata: Lumbricidae) and *Heterarthrus nemoratus* (Fallén, 1808) (Hymenoptera: Tenthredinidae). *Deroceras agreste* was found at one site less than 100 m from a road, *Dendrobaena octaedra* was widespread over the study area and *Heterarthrus nemoratus* was found at one site more than 3 km from human development (Fig. 7).

Amongst the birds observed were three species of special interest. We documented *Sitta canadensis* Linnaeus, 1766 (Passeriformes: Sittidae) at four sites. *Regulus satrapa*, Lichtenstein, 1823 (Passeriformes: Regulidae) was detected at six sites. *Contopus cooperi* Nuttall, 1831 (Passeriformes: Tyrannidae) was documented at ten sites.

One species of potential conservation concern, *Lathrapanteles heleios* Williams, 1985 (Hymenoptera: Braconidae) was detected on two separate occasions at a single site.

The COI sequence which we obtained was 98.53% similar (p-dist) to a specimen with processid [JSHYO264-11](#), identified as *Lathrapanteles heleios* and it was placed within a clade of sequences of this species (Suppl. material 8).

Species diversity

The analysis dataset (Suppl. material 9) included 3,090 occurrence records of 849 species. For all taxa except arthropods, the methods used captured more than half of the estimated numbers of species that could be detected using our methods (Table 1, Fig. 8); for arthropods, we detected only 42% of the estimated number of species. Based on species accumulation curves, adding more sites would contribute relatively few species per plot for all taxa except for arthropods (Fig. 9). The slope of the species accumulation curve at the 39th site ranged from 0.05 species added per additional site for segmented worms (Annelida) and liverworts (Marchantiophyta) to 8 for arthropods (Table 1).

Communities

The UPGMA tree grouped the 40 sites into five community groups: 22 in upland mixed forest, 11 in black spruce forest, 3 in open deciduous forest, 3 in shrub-sedge bog and 1 in willow. This grouping remained consistent even when different clustering methods were used and when rare species were included. These community groupings also loosely corresponded to the NLCD classification of these sites (Fig. 10).

The NMDS analysis including two dimensions resulted in a stress value of 0.13, a "satisfactory" stress value according to the guidelines of McCune and Grace (2002). The five community groups from the cluster analysis were also separated along the NMDS axes (Fig. 11).

The species included in each community are provided in Suppl. material 10. The open deciduous forest community included the highest number of species unique to that community (Fig. 12); the willow community had the fewest unique species. Four species (*Calamagrostis canadensis* (Michx.) P.Beauv. (Poales: Poaceae), *Setophaga coronata* (Passeriformes: Parulidae), *Loxia leucoptera* (Passeriformes: Fringillidae) and *Junco hyemalis* (Passeriformes: Emberizidae)) were members of all five communities.

The open deciduous forest community (3 sites) of 114 member species was characterised by the shrubs *Alnus viridis* (Chaix) DC. (Fagales: Betulaceae) and *Oplopanax horridus* Miq. (Apiales: Araliaceae) under an open hardwood overstorey of *Betula neoalaskana* Sarg. (Fagales: Betulaceae) or *Populus × hastata* Dode (Malpighiales: Salicaceae). Other species included *Calamagrostis canadensis*, *Catharus ustulatus* (Nuttall, 1840) (Passeriformes: Turdidae), *Dryopteris expansa* (C.Presl) Fraser-Jenk. & Jermy (Polypodiales: Dryopteridaceae), *Empidonax alnorum* Brewster, 1895 (Passeriformes: Tyrannidae), *Equisetum arvense* L. (Equisetales: Equisetaceae), *Fannia brooksi* Chillcott, 1961 (Diptera: Fanniidae), *Gymnocarpium dryopteris* Newm. (Polypodiales: Cystopteridaceae), *Parmelia sulcata* Taylor (Lecanorales: Parmeliaceae), *Poecile*

atricapillus (Linnaeus, 1766) (Passeriformes: Paridae), *Setophaga coronata* and *Trientalis europaea* L. (Ericales: Primulaceae).

The upland mixed forest community (22 sites) of 94 species included an overstorey of *Betula neoalaskana*, *Picea glauca* (Moench) Voss (Pinales: Pinaceae), and *Populus tremuloides* Michx. (Malpighiales: Salicaceae) with a diverse understorey of *Rosa acicularis* Lindl. (Rosales: Rosaceae), *Chamerion angustifolium* (L.) J.Holub (Myrtales: Onagraceae), *Calamagrostis canadensis*, *Vaccinium vitis-idaea* L. (Ericales: Ericaceae), *Lycopodium annotinum* L. (Lycopodiales: Lycopodiaceae) and *Linnaea borealis* L. (Dipsacales: Caprifoliaceae). Additional species included *Catharus ustulatus*; *Cornus canadensis* L. (Cornales: Cornaceae); *Equisetum pratense* Ehrh. (Equisetales: Equisetaceae); *Geocaulon lividum* Fernald (Santalales: Santalaceae); *Gymnocarpium dryopteris*; Hybotidae sp. BOLD:ACX4896 (Diptera: Hybotidae), *Hylocomium splendens* W.P.Schimper, 1852 (Hypnales: Hylocomiaceae); *Hypogymnia physodes* (L.) Nyl. (Lecanorales: Parmeliaceae); *Junco hyemalis*; *Lobaria pulmonaria* (L.) Hoffm. (Peltigerales: Lobariaceae); *Ochlerotatus communis* (De Geer, 1776) (Diptera: Culicidae); *Orthilia secunda* (L.) House (Ericales: Ericaceae); *Parmelia sulcata*; *Pleurozium schreberi* Mitten, 1869 (Hypnales: Hylocomiaceae); *Sanionia uncinata* Loeske, 1907 (Hypnales: Amblystegiaceae); *Setophaga coronata*; and *Trientalis europaea*.

The shrub-sedge bog community (3 sites) was characterised by the presence of *Andromeda polifolia* L. (Ericales: Ericaceae), *Betula glandulosa* Michx. (Fagales: Betulaceae), *Carex rotundata* Wahlenb. (Poales: Cyperaceae), *Dictyna arundinacea* (Linnaeus, 1758) (Araneae: Dictynidae), *Eudorylas* sp. BOLD:ACZ4721 (Diptera, Pipunculidae), *Ledum palustre* L. (Ericales: Ericaceae), *Myrica gale* L. (Fagales: Myricaceae), *Passerculus sandwichensis* (J.F.Gmelin, 1789) (Passeriformes: Emberizidae), and *Vaccinium oxycoccos* L. (Ericales: Ericaceae).

The single shrub site straddled a small stream where a *Salix commutata* Bebb (Malpighiales: Salicaceae), *Salix pulchra* Cham., *Myrica gale*, *Calamagrostis canadensis* and other wetland plants grew. The full list of species at this site can be obtained by searching through Suppl. material 6 for site label "SK24".

The black spruce forest community (11 sites) was characterized by *Picea mariana* Britton, Sterns & Poggenb. (Pinales: Pinaceae), *Ledum palustre*, *Vaccinium oxycoccos*, *Betula glandulosa*, *Empetrum nigrum* L. (Ericales: Ericaceae), *Rubus chamaemorus* L. (Rosales: Rosaceae) and *Vaccinium vitis-idaea*. Other frequent species in black spruce forest sites were *Catharus ustulatus*; *Hypogymnia occidentalis* L.H.Pike (Lecanorales: Parmeliaceae); *Hypogymnia physodes*; *Junco hyemalis*; *Pleurozium schreberi*; *Regulus calendula* (Linnaeus, 1766) (Passeriformes: Regulidae); *Sphagnum angustifolium* C.E.O.Jensen, 1896 (Sphagnales: Sphagnaceae); *Sphagnum fuscum* Klinggräff, 1872; and *Vaccinium uliginosum* L. (Ericales: Ericaceae).

Discussion

New distribution records

Remarkably, 26% of the described arthropod species which we documented appeared to be new records for Alaska. We believe that this is partially due to the use of HTS methods, which identified many species—especially small Diptera—that have been under-surveyed in Alaska. An extensive DNA barcode library of insects from Canada (Hebert et al. 2016) enabled identification of many Alaskan species for which DNA barcodes had not been obtained by Sikes et al. (2017a).

We expect that many more arthropod species could be found even in our small study area, based on species accumulation curves, especially if rare community types were intentionally sought. The Diptera, especially smaller species, are diverse in temperate and higher latitudes, with many more species expected to be described (Hebert et al. 2016, Morinière et al. 2019). Despite a similarly high expected diversity of Diptera in Alaska, the state lacks a dedicated Diptera taxonomist and there have been few efforts to inventory the Alaskan species of Diptera.

Most of the species, newly reported for Alaska, are widespread in northern North America, so finding them in our study area was not particularly surprising. We did not consider the five species of arthropods that were apparently new to North America to be non-native because they may be trans-Beringian species, a well-documented distribution pattern in the flora and fauna of Alaska (Hultén 1968, Scudder 1979, Danks et al. 1997).

Our detection of *Lathrapanteles heleios*, a species previously known only from southern Ontario and considered for inclusion in Species Candidate Lists of the Committee on the Status of Endangered Wildlife in Canada (Fernandez-Triana 2014), indicates that it may be much more widespread than previously thought. Alternatively, our detection may represent a closely-related congener.

We do not consider the new distribution records presented here, based on metabarcoded DNA samples, to be as verifiable as specimen-based records. We regard the new distribution records documented here as tentative until verified by specimen-based collections. However, we applied appropriate means to filter out potential false positive occurrences and we carefully scrutinised all identifications resulting in potential new distribution records. We provided the sequence data via Zenodo (Bowser 2019), GenBank, Arctos and as supplementary material here (Suppl. materials 4, 5) so that all of our identifications can be checked.

We propose that an integrative combination of specimen-based morphological identifications (Sikes et al. 2017b), DNA barcode library building through specimen-based Sanger sequencing (Hebert et al. 2016, Morinière et al. 2019) and inventories by HTS metagenomic methods for bulk samples (Morinière et al. 2019) should be pursued on Alaskan National Wildlife Refuges to enable efficient future biomonitoring through HTS methods.

Non-native species and changing assemblages

We did not observe any non-native plants in the sites included in this study. This suggests that non-native plants are rare in the study area, particularly beyond the immediate footprint of human disturbance.

In contrast, we found three non-native animal species within the study area. *Dendrobaena octaedra* had already been widely documented on the Kenai National Wildlife Refuge (Saltmarsh et al. 2016). This surface-dwelling earthworm is parthenogenic (Omodeo 1955), can be spread by vehicles (Cameron et al. 2008) and is typically amongst the first species of European earthworms to invade forests in northern North America (Hale et al. 2005, Cameron et al. 2007).

Heterarthrus nemoratus, a non-native sawfly that mines leaves of birches, was first collected in Alaska in 2004 (Snyder et al. 2007). By 2015, this winged species had become widespread in southern Alaska (FS-R10-FHP 2016). It is likely that nearly all birch forest on the Kenai Peninsula is inhabited by *Heterarthrus nemoratus*.

Deroceras agreste, a Eurasian slug and agricultural pest, has only recently been documented from Alaska. As of this writing, only two other records of this species from Alaska have been published (GBIF 2019a), both from the Kenai Peninsula and both after 2016. *Deroceras agreste* has also been intercepted at "Juneau Old Docks", Juneau, Alaska on 13 June 2016 (unpublished data from U.S. Department of Agriculture, Animal and Plant Health Inspection Service, provided by Christopher H. Secary, Alaska Department of Natural Resources). At present, it appears that *D. agreste* on the Kenai Peninsula is restricted to areas close to human development.

Sitta candensis has become common on the Kenai Peninsula only recently. As recently as 1959, this species was not known to occur in Southcentral Alaska (Gabrielson and Lincoln 1959). In 1968, it was considered accidental on the Kenai Moose Range (the former name of KNR) with only 3 previously-documented sightings (U.S. Fish and Wildlife Service 1968). In 1973, it was considered to be rare in the North Gulf Coast region of Alaska (Isleib and Kessel 1973). The oldest records of *Sitta candensis* from the Kenai Peninsula available on GBIF are two observations that date from 1979; from 2015 to 2018, 595 to 855 observations per year were recorded (GBIF 2019b).

Regulus satrapa has become a relatively common species during the breeding season over the past 20 years on the northern portion of the Kenai Peninsula, based on Breeding Bird Surveys (Pardieck et al. 2019). Gabrielson and Lincoln (1959) described this species' distribution as occurring on the Kenai Peninsula, but highlighted the southern Kenai Peninsula habitat types similar to Kodiak Island and the coastal areas of Prince William Sound where it was common. It was listed as uncommon during the breeding season on the Kenai Moose Range (U.S. Fish and Wildlife Service 1968).

Contopus cooperi was identified as a priority species of concern by the Boreal Partners in Flight Working Group (Boreal Partners in Flight 2005). The global population decline of the species combined with potential threats to preferred habitat types prompted this designation. While anecdotal information indicates a continued decline of this species across the Kenai Peninsula, we detected it with regularity.

The detection of *Dendrobaena octaedra* and *Heterarthrus nemoratus* in parts of the Slikok watershed that are distant from obvious human disturbance means that the forest assemblage in which they were found is now a hybrid assemblage *sensu* Hobbs et al. (2009). These are two of several species that have recently become part of the hybrid assemblages of the KNWR, successfully occupying even areas that are far from human disturbance.

Other non-native species that have become widespread on KNWR within the last 100 years, but were not detected in the current study, include *Canis latrans* Say, 1823 (Thurber and Peterson 1991); *Eriocampa ovata* (Linnaeus, 1760) (U.S. Fish and Wildlife Service, Region 7 2010); *Lupinus polyphyllus* Lindl. (U.S. Fish and Wildlife Service, Region 7 2010); *Monsoma pulveratum* (Retzius, 1783) (Kruse et al. 2010, U.S. Fish and Wildlife Service, Region 7 2010); *Profenusa thomsoni* (Konow, 1886) (Snyder et al. 2007, FS-R10-FHP 2016); *Taraxacum officinale* F.H.Wigg. (U.S. Fish and Wildlife Service, Region 7 2010); *Thymallus arcticus* (Pallas, 1776) (State of Alaska, Department of Fish and Game 1978); and *Trichodectes canis* (de Geer, 1778) (Schwartz et al. 1983, Wolstad 2010). Some of these species, (e.g. *Eriocampa ovata*, *Monsoma pulveratum*, and *Profenusa thomsoni*), were probably present in our study area, but we may have failed to detect them due to rarity of these species, patchy distribution of these species, mismatch of our temporal sampling windows with the activity of these species or low probability of detection using our methods. *Taraxacum officinale* was common in our study area, but it was mostly restricted to areas of human disturbance. *Canis latrans* was present in our study area, but we did not use methods designed to detect mammals.

Communities

The communities which we observed fit within previous vegetation classifications in this region. Our open deciduous forest community corresponded to the I.B.2 open broadleaf forest and I.B.2.c open balsam poplar (black cottonwood) forest classes of Viereck et al. (1992) and the *Alnus crispa* ssp. *sinuata*-*Echinopanax horridum*, *Betula papyrifera*/*Echinopanax horridum* and *Populus balsamifera* ssp. *trichocarpa*/*Echinopanax horridum* classes of DeVelice et al. (1999). Our upland mixed forest community lined up with the I.B broadleaf forest and I.C mixed forest classes of Viereck et al. (1992) and the Lutz spruce-paper birch cover types, paper birch cover types and *Picea X lutzii*-*Populus tremuloides*/*Vaccinium vitis-idaea* class of DeVelice et al. (1999). Our shrub-sedge bog community fitted the II.C.2.j sweetgale-graminoid bog class of Viereck et al. (1992) and the *Eriophorum angustifolium*-*Trichophorum caespitosum* and *Myrica gale*/*Eriophorum angustifolium* classes of DeVelice et al. (1999). Our willow community was like the II.C.2.g willow open low scrub class of Viereck et al. (1992) and the *Myrica gale*/

Calamagrostis canadensis class of DeVelice et al. (1999). Our black spruce forest community fitted the I.A.2.f black spruce open needleleaf forest and I.A.3.d black spruce needleleaf woodland classes of Viereck et al. (1992) and the *Picea mariana/Vaccinium vitis-idaea* community of Viereck et al. (1992).

Methodological comments

While we selected a sample frame, plot sizes and other parameters carefully, generally using methods consistent with Morton et al. (2009), we recognise that these could be optimised. In particular, it would be useful to determine an optimal sweep net sample area and an optimal sequencing depth. Using a sweep net with a smaller mesh size would improve collection rates of more minute arthropods in the future (Derek Sikes, University of Alaska Museum, personal communication).

It was apparent from the species accumulation curve of arthropods that many more species remain to be collected in our study area. We believe that one reason for this pattern is low probability of detection for many species using our sweep net and metabarcoding methods.

We intentionally limited our efforts to testing methods that could efficiently be deployed over large, remote areas and deliver information on a statistically useful sample size of sampling locations. This limited the available sampling methods to active sampling methods and extraction methods. Of these, we chose sweep-net sampling because of its simplicity, because the required equipment could be compact and light, because it samples a wide diversity of insect groups (Collet 2004) and because samples could be stored and processed later. Many studies have shown sweep-net sampling to be an effective sampling method (e.g. Spafford and Lortie 2013, Shweta and Rajmohana 2018), but the weaknesses and limitations of this method are also well known. Drawbacks of the sweep net method include variation amongst collectors, differing results depending on the time of day and weather.

We recognise that other sampling methods (e.g. malaise traps) would be superior for maximising the number of species observed (Marshall et al. 1994). Morinière et al. (2019) demonstrated that malaise traps, processed by HTS methods, can be an effective method for biomonitoring Diptera. We agree with other authors that no single method samples terrestrial invertebrates exhaustively and that employing a variety of methods would deliver the most comprehensive biological inventories (Marshall et al. 1994, Yi et al. 2012).

Amongst primer pairs, differences in binding to DNA templates lead to amplification biases, affecting both read abundances and detections of species, so that any single primer set will lead to detections of a subset of species (Elbrecht and Leese 2017, Hajibabaei et al. 2019). The *mICOLLintFIHCO2198* primer pair which we used amplifies well across a broad range of arthropod taxa (Leray et al. 2013, Brandon-Mong et al. 2015), but, in future efforts, we would consider selecting the *mICOLLintFjgHCO2198* primer pair, which has a more degenerate reverse primer and amplifies well across a broader range of arthropod taxa than the *mICOLLintFIHCO2198* pair (Elbrecht and Leese 2017).

In future efforts, we would consider using the mBRAVE platform (<http://www.mbrave.net>, Ratnasingham 2019) for metabarcoding analyses. An analysis on this cloud-based platform with standardised analytical steps should be more easily repeatable than the implementation of the SCVUC COI metabarcode pipeline that we used, especially for non-specialists. Ideally, we would like to demonstrate methods that would be more accessible to non-specialists, so that metabarcoding can become more of a standard practice for biomonitoring. For beginners, Liu et al. (2019) recommended graphical-based platforms, including mBRAVE.

Other comments

At least some of the patterns which we documented are the result of interannual variation related to cycles of the boreal forest. For example, the high frequency of occurrence of *Loxia leucoptera* that we documented was likely related to the irruptiveness of this species. *Loxia leucoptera* is common on the Kenai Peninsula, but it is known to be erratically migratory (Gabrielson and Lincoln 1959).

Conclusion

In the past, monitoring all but a small subset of biodiversity has been logistically and economically intractable (Dallmeier et al. 2013). However, we demonstrated practical and efficient methods that could be repeated for monitoring of a large portion of biodiversity. The combination of observation-based, specimen-based and HTS methods which we used were effective for documenting species distributions and species assemblages within the study area, although there is room for improvement, particularly in the detection of arthropod species. Biomonitoring, using such methods, could provide the kinds of data necessary for meeting the broad conservation mandates of KNWR and other Alaska National Wildlife Refuges.

Future Directions

In future efforts, we intend to survey additional hyperdiverse portions of the terrestrial biota, including soil arthropods, soil fungi and soil bacteria, thus yielding even more complete community assemblages.

Acknowledgements

We are grateful to the USGS Advanced Research Computing team for use of the Yeti Supercomputer. We thank Daniel Bogan and Don Buckle for reviewing the list of new Alaskan distribution records. Derek Sikes of the University of Alaska Museum and Kathryn Baer of the USDA Forest Service reviewed drafts of this manuscript and provided helpful comments that substantially improved it. We thank Christopher H. Secary from the Alaska Department of Natural Resources for sharing an unpublished occurrence record of *Deroceras agreste* in Alaska.

Ethics and security

All collecting and shipping of biological material was performed in accordance with applicable laws and permitting requirements. Specifically, all collecting and processing of arthropod sampling conformed with guidelines established by the US Fish and Wildlife Service's Institutional Animal Care and Use Committee and were also permitted under the State of Alaska Department of Fish and Game Fish Resource Permit# SF2016-082.

Author contributions

John Morton, Matt Bowser, Todd Eskelin and Dawn Robin Magness designed the project. Matt Bowser, Todd Eskelin, Jennifer Hester and Dawn Magness led field work. Mariah McInnis, Joel Stone and Rebekah Brassfield assisted with field work. Annie Dziergowski and Tracy Melvin helped with field work as volunteers. Rebekah Brassfield, Matt Bowser and Joel Stone processed sweep-net samples. Matt Bowser conducted the analysis and led writing of the manuscript.

Conflicts of interest

The authors declare that no competing interests exist.

References

- 105th Congress (1997) 111 Stat. 1252 - National Wildlife Refuge System Improvement Act of 1997. In: 105th Congress (Ed.) United States Statutes at Large. 111. U.S. Government Printing Office URL: <https://www.govinfo.gov/app/details/STATUTE-111/STATUTE-111-Pg1252>
- 96th Congress (1980) 94 Stat. 2371 - Alaska National Interest Lands Conservation Act of 1979. In: 96th Congress (Ed.) United States Statutes at Large. 94. U.S. Government Printing Office URL: <https://www.govinfo.gov/app/details/STATUTE-94/STATUTE-94-Pg2371>
- AKEPIC (2016) Alaska Exotic Plant Information Clearinghouse database. Alaska Natural Heritage Program, University of Alaska, Anchorage. Release date: 2016-6-07. URL: <http://aknhp.uaa.alaska.edu/maps/akepic/>
- Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. Journal of Molecular Biology 215 (3): 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Aylagas E, Borja Á, Irigoien X, Rodríguez-Ezpeleta N (2016) Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Frontiers in Marine Science 3: 96-96. <https://doi.org/10.3389/fmars.2016.00096>

- Baird D, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology* 21 (8): 2039-2044. <https://doi.org/10.1111/j.1365-294x.2012.05519.x>
- Bivand R, Lewin-Koh N (2018) maptools: Tools for handling spatial objects. R package version 0.9-4. URL: <https://CRAN.R-project.org/package=maptools>
- Bivand R, Rundel C (2018) rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.4-2. URL: <https://CRAN.R-project.org/package=rgeos>
- Bivand R, Keitt T, Rowlingson B (2018) rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.3-6. URL: <https://CRAN.R-project.org/package=rgdal>
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolk T, Kreps J, Langille MG, Lee J, Ley R, Liu Y, Lofffield E, Lozupone C, Maher M, Marotz C, Martin B, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson IMS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CH, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Knight R, Caporaso JG (2018) QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints* 6: e27295v2-e27295v2. <https://doi.org/10.7287/peerj.preprints.27295v1>
- Boreal Partners in Flight (2005) Landbirds of concern included in the ADF&G's Comprehensive Wildlife Conservation Plan. U.S. Geological Survey, 35 pp. URL: https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/CWCS_landbirds.pdf
- Bowser M, 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-e10792. <https://doi.org/10.3897/bdj.5.e10792>
- Bowser M (2019) Raw metagenomic data from sweep net samples collected in 2016 as part of the Slikok Creek Watershed Biotic Inventory. Zenodo. Release date: 2019-11-20. URL: <https://doi.org/10.5281/zenodo.3245460>
- Bowser M, Burr S, Davis I, Dubois G, Graham E, Moan J, Swenson S (2019) A test of metabarcoding for Early Detection and Rapid Response monitoring for non-native forest pest beetles (Coleoptera). *Research Ideas and Outcomes* 5: e48536-e48536. <https://doi.org/10.3897/rio.5.e48536>
- Brandon-Mong G-J, Gan H-M, Sing K-W, Lee P-S, Lim P-E, Wilson J-J (2015) DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research* 105 (6): 717-727. <https://doi.org/10.1017/s0007485315000681>
- Brunsdon C, Chen H (2014) GISTools: Some further GIS capabilities for R. R package version 0.7-4. URL: <https://CRAN.R-project.org/package=GISTools>

- Bush A, Compson Z, Monk W, Porter T, Steeves R, Emilson E, Gagne N, Hajibabaei M, Roy M, Baird D (2019) Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution* 7: 434-434. <https://doi.org/10.3389/fevo.2019.00434>
- Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11 (12): 2639-2643. <https://doi.org/10.1038/ismej.2017.119>
- Cameron E, Bayne E, Clapperton MJ (2007) Human-facilitated invasion of exotic earthworms into northern boreal forests. *Ecoscience* 14 (4): 482-490. [https://doi.org/10.2980/1195-6860\(2007\)14\[482:HIOEEI\]2.0.CO;2](https://doi.org/10.2980/1195-6860(2007)14[482:HIOEEI]2.0.CO;2)
- Cameron E, Bayne E, Coltman D (2008) Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. *Molecular Ecology* 17: 1189-1197. <https://doi.org/10.1111/j.1365-294X.2007.03603.x>
- Chamberlain S (2018) bold: Interface to Bold Systems API. 0.8.6. URL: <https://CRAN.R-project.org/package=bold>
- Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43 (4): 783-791. <https://doi.org/10.2307/2531532>
- Chen H (2018) VennDiagram: Generate high-resolution Venn and Euler plots. R package version 1.6.20. URL: <https://CRAN.R-project.org/package=VennDiagram>
- Chiu C, Wang Y, Walther B, Chao A (2014) An improved nonparametric lower bound of species richness via a modified good-turing frequency formula. *Biometrics* 70 (3): 671-682. <https://doi.org/10.1111/biom.12200>
- Collet D (2004) Pilot study for sampling insect diversity. U.S. Fish & Wildlife Service, Kenai National Wildlife Refuge, Soldotna, Alaska, 10 pp. <https://doi.org/10.7299/X7W37WN9>
- Dallmeier F, Szaro RC, Alonso A, Comiskey J, Henderson A (2013) Framework for assessment and monitoring of biodiversity. In: Levin SA (Ed.) *Encyclopedia of Biodiversity*. 3. Academic Press, Waltham, Massachusetts. [ISBN 9780123847201]. <https://doi.org/10.1016/B978-0-12-384719-5.00316-6>
- Danks HV, Downes JA, Larson DJ, Scudder GGE (1997) *Insects of the Yukon: Characteristics and history*. Biological Survey of Canada Monograph series No. 2. In: Danks HV, Downes JA (Eds) *Insects of the Yukon*. Biological Survey of Canada (Terrestrial Arthropods), Ottawa, Ontario. URL: <https://biologicalsurvey.ca/assets/file/105> [ISBN 0-9692727-8-2].
- Dapporto L, Ramazzotti M, Fattorini S, Vila R, Talavera G, Dennis RL (2015) recluster: ordination methods for the analysis of beta-diversity indices. R package version 2.8. URL: <https://CRAN.R-project.org/package=recluster>
- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology* 26: 5872-5895. <https://doi.org/10.1111/mec.14350>
- DeVelice RL, Hubbard CJ, Boggs K, Boudreau S, Potkin M, Boucher T, Wertheim C (1999) *Plant community types of the Chugach National Forest: Southcentral Alaska*. Technical Publication R10-TP-76. USDA Forest Service, Chugach National Forest, Alaska Region, Anchorage, Alaska, 376 pp. URL: https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprd3814845.pdf

- Edgar R (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv 081257 <https://doi.org/10.1101/081257>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26 (19): 2460-2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science* 5: 11. <https://doi.org/10.3389/fenvs.2017.00011>
- Fernandez-Triana J (2014) Towards the conservation of parasitoid wasp species in Canada: Preliminary assessment of Microgastrinae (Hymenoptera: Braconidae). *Biodiversity Data Journal* 2: e1067. <https://doi.org/10.3897/bdj.2.e1067>
- FS-R10-FHP (2016) Forest health conditions in Alaska - 2015. USDA Forest Service, Alaska Region, Anchorage, Alaska. URL: http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fseprd491888.pdf
- Gabrielson I, Lincoln F (1959) *Birds of Alaska*. The Stackpole Company and the Wildlife Management Institute, Harrisburg, Pennsylvania and Washington, D.C, xiv + 922 pp.
- GBIF (2016) GBIF Occurrence Download. GBIF.org. Release date: 2016-6-08. URL: <https://doi.org/10.15468/dl.wcg2yo>
- GBIF (2019a) GBIF Occurrence Download. GBIF.org. Release date: 2019-11-26. URL: <https://doi.org/10.15468/dl.jhmbjr>
- GBIF (2019b) GBIF Occurrence Download. GBIF.org. Release date: 2019-12-20. URL: <https://doi.org/10.15468/dl.wvlboj>
- Gibson J, Shokralla S, Curry C, Baird D, Monk W, King I, Hajibabaei M (2015) Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PLOS One* 10 (10): e0138432. <https://doi.org/10.1371/journal.pone.0138432>
- Goldstein P, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *BioEssays* 33 (2): 135-147. <https://doi.org/10.1002/bies.201000036>
- Hajibabaei M, Shokralla S, Zhou X, Singer GC, Baird D (2011) Environmental barcoding: a Next-Generation Sequencing approach for biomonitoring applications using river benthos. *PLOS One* 6 (4): e17497. <https://doi.org/10.1371/journal.pone.0017497>
- Hajibabaei M, Baird D, Fahner N, 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 B: Biological Sciences* 371 (1702): 20150330. <https://doi.org/10.1098/rstb.2015.0330>
- Hajibabaei M, Porter T, Wright M, Rudar J (2019) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLOS One* 14 (9): e0220953. <https://doi.org/10.1371/journal.pone.0220953>
- Hale C, Frelich L, Reich P, Pastor J (2005) Effects of European earthworm invasion on soil characteristics in northern hardwood forests of Minnesota, USA. *Ecosystems* 8: 911-927. <https://doi.org/10.1007/s10021-005-0066-x>
- Handel C, Cady M (2004) Alaska landbird monitoring survey: Protocol for setting up and conducting point count surveys. USGS Alaska Science Center, Anchorage, Alaska, 40 pp. URL: http://alaska.usgs.gov/science/biology/bpif/monitor/alms/ALMSprotocol_2004.pdf

- Hebert PN, Ratnasingham S, Zakharov E, Telfer A, Levesque-Beaudin V, Milton M, Pedersen S, Jannetta P, deWaard J (2016) Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702): 20150333. <https://doi.org/10.1098/rstb.2015.0333>
- Hijmans R (2018) raster: Geographic Data Analysis and Modeling. R package version 2.8-4. URL: <https://CRAN.R-project.org/package=raster>
- Hobbs R, Higgs E, Harris J (2009) Novel ecosystems: implications for conservation and restoration. *Trends in Ecology & Evolution* 24 (11): 599-605. <https://doi.org/10.1016/j.tree.2009.05.012>
- 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 (5): 345-354. URL: https://www.asprs.org/a/publications/pers/2015journals/PERS_May_2015/HTML/index.html#345
- Hultén E (1968) *Flora of Alaska and neighboring territories*. Stanford University Press, Stanford, California, 1032 pp. [ISBN 0804706433]
- Iannone R (2020) DiagrammeR: Graph/Network Visualization. R package version 1.0.5. URL: <https://CRAN.R-project.org/package=DiagrammeR>
- Isleib ME, Kessel B (1973) *Birds of the North Gulf Coast - Prince William Sound Region, Alaska*. Biological papers of the University of Alaska, No. 14. University of Alaska Press, Fairbanks, Alaska. [ISBN 0-912006-39-0]
- Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37 (142): 547-579. <https://doi.org/10.5169/SEALS-266450>
- Jenkerson CB, Maersperger T, Schmidt G (2010) eMODIS: A user-friendly data source. U.S. Geological Survey OpenFile Report 2010–1055. U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia, 10 pp. URL: <http://pubs.usgs.gov/of/2010/1055/>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30 (4): 772-780. <https://doi.org/10.1093/molbev/mst010>
- Kruse J, Zogas K, Hard J, Lisuzzo N (2010) New pest in Alaska and Washington. The Green Alder Sawfly - *Monsoma pulveratum* (Retzius). Pest Alert R10-PR-022. United States Department of Agriculture, Forest Service, State and Private Forestry, Alaska Region URL: https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5303926.pdf
- Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (Eds) *Nucleic acid techniques in bacterial systematics*. John Wiley and Sons, New York, 329 pp. [ISBN 0471929069].
- Lawrence AP, Bowers MA (2002) A test of the 'hot' mustard extraction method of sampling earthworms. *Soil Biology and Biochemistry* 34 (4): 549-552. [https://doi.org/10.1016/S0038-0717\(01\)00211-5](https://doi.org/10.1016/S0038-0717(01)00211-5)
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10 (1): 34. <https://doi.org/10.1186/1742-9994-10-34>

- Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47 (W1): W256-W259. <https://doi.org/10.1093/nar/gkz239>
- Liu M, Clarke L, Baker S, Jordan G, Burridge C (2019) A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology* <https://doi.org/10.1111/een.12831>
- Lobo J, Shokralla S, Costa MH, Hajibabaei M, Costa FO (2017) DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports* 7 (1): 15618. <https://doi.org/10.1038/s41598-017-15823-6>
- MacKenzie D, Nichols J, Lachman G, Droege S, Royle JA, Langtimm C (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83 (8): 2248-2255. [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:esorwd\]2.0.co;2](https://doi.org/10.1890/0012-9658(2002)083[2248:esorwd]2.0.co;2)
- MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2005) *Occupancy estimation and modeling*. Academic Press, 344 pp. URL: <https://www.elsevier.com/books/occupancy-estimation-and-modeling/mackenzie/978-0-12-088766-8> [ISBN 9780120887668]
- Marshall SA, Anderson RS, Roughley RE, Behan-Pelletier VM, Danks HV (1994) Terrestrial arthropod biodiversity: Planning a study and recommended sampling techniques. Supplement to the Bulletin of the Entomological Society of Canada 26 (1): 1-33. URL: <https://biologicalsurvey.ca/public/Bsc/Controller/Page/briefs/planningastudy.pdf>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17 (1): 10-12. <https://doi.org/10.14806/ej.17.1.200>
- McCune B, Grace J (2002) *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, Oregon. [ISBN 0-9721290-0-6]
- Morinière J, Balke M, Doczkal D, Geiger M, Hardulak L, Haszprunar G, Hausmann A, Hendrich L, Regalado L, Rulik B, Schmidt S, Wägele J, Hebert PN (2019) A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular Ecology Resources* 19 (4): 900-928. <https://doi.org/10.1111/1755-0998.13022>
- 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. In: McWilliams W, Moisen G, Czaplewski R (Eds) *Forest Inventory and Analysis (FIA) Symposium 2008; October 2123, 2008; Park City, UT. Proc. RMRSP56CD*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, 17 pp. URL: <http://www.treeseearch.fs.fed.us/pubs/33332>
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) *vegan: Community Ecology Package*. R package version 2.5-3. URL: <https://CRAN.R-project.org/package=vegan>
- Omodeo P (1955) *Cariologia dei Lumbricidae*. Il contributo. *Caryologia* 8 (1): 135-178. <https://doi.org/10.1080/00087114.1955.10797555>
- Pardieck KL, Ziolkowski DJ, Lutmerding M, Aponte V, Hudson MR (2019) North American breeding bird survey dataset. Version 2018.0. U.S. Geological Survey, Patuxent Wildlife Research Center. URL: <https://doi.org/10.5066/P9HE8XYJ>

- Penev L, Mitchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Tuama ÉÓ, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. Research Ideas and Outcomes 3: e12431. <https://doi.org/10.3897/rio.3.e12431>
- Porter T, Hajibabaei M (2018a) Automated high throughput animal CO1 metabarcoding classification. Scientific Reports 8 (1). <https://doi.org/10.1038/s41598-018-22505-4>
- Porter T, Hajibabaei M (2018b) Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology 27 (2): 313-338. <https://doi.org/10.1111/mec.14478>
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. PLOS ONE 5 (3): e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). Molecular Ecology Notes 7 (3): 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Ratnasingham S, Hebert PN (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. PLOS One 8 (7): e66213. <https://doi.org/10.1371/journal.pone.0066213>
- Ratnasingham S (2019) mBRAVE: The Multiplex Barcode Research And Visualization Environment. Biodiversity Information Science and Standards 3 <https://doi.org/10.3897/biss.3.37986>
- R Core Team (2018) R: A language and environment for statistical computing. 3.5.1. URL: <https://www.R-project.org/>
- Revell LJ (2012) phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3: 217-223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Reynolds J (1977) The earthworms (Lumbricidae and Sparganophilidae) of Ontario. Royal Ontario Museum, Toronto, Ontario, ix + 141 pp. <https://doi.org/10.5962/bhl.title.60740>
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584. <https://doi.org/10.7717/peerj.2584>
- Saltmarsh DM, Bowser M, Morton J, Lang S, Shain D, Dial R (2016) Distribution and abundance of exotic earthworms within a boreal forest system in Southcentral Alaska. NeoBiota 28: 67-86. <https://doi.org/10.3897/neobiota.28.5503>
- Schwartz CC, Stephenson R, Wilson N (1983) *Trichodectes canis* on the gray wolf and coyote on Kenai Peninsula, Alaska. Journal of Wildlife Diseases 19: 372-373. <https://doi.org/10.7589/0090-3558-19.4.372>
- Scudder GGE (1979) Present patterns in the fauna and flora of Canada. Memoirs of the Entomological Society of Canada 111, Supplement S108: 87-179. <https://doi.org/10.4039/entm111108087-1>
- Shweta M, Rajmohana K (2018) A comparison of sweep net, yellow pan trap and malaise trap for sampling parasitic Hymenoptera in a backyard habitat in Kerala. Entomon 43 (1): 43-44.
- Sigovini M, Keppel E, Tagliapietra D (2016) Open Nomenclature in the biodiversity era. Methods in Ecology and Evolution 7 (10): 1217-1225. <https://doi.org/10.1111/2041-210x.12594>

- Sikes D, Bowser M, Morton J, Bickford C, Meierotto S, Hildebrandt K (2017a) Building a DNA barcode library of Alaska's non-marine arthropods. *Genome* 60 (3): 248-259. <https://doi.org/10.1139/gen-2015-0203>
- Sikes D, Bowser M, Daly K, Høye T, Meierotto S, Mullen L, Slowik J, Stockbridge J (2017b) The value of museums in the production, sharing, and use of entomological data to document hyperdiversity of the changing North. *Arctic Science* 3 (3): 498-514. <https://doi.org/10.1139/as-2016-0038>
- Simpson A, Eyler MC, Sikes D, Bowser M, Sellers E, Guala GF, Cannister M, Libby R, Kozlowski N (2019) A comprehensive list of non-native species established in three major regions of the United States. Version 2.0, 2019. U.S. Geological Survey. URL: <https://doi.org/10.5066/P9E5K160>
- Snyder C, MacQuarrie CK, Zogas K, Kruse J, Hard J (2007) Invasive species in the last frontier: distribution and phenology of birch leaf mining sawflies in Alaska. *Journal of Forestry* 105 (3): 113-119.
- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 28: 1409-1438.
- Spafford R, Lortie C (2013) Sweeping beauty: is grassland arthropod community composition effectively estimated by sweep netting? *Ecology and Evolution* 3 (10): 3347-3358. <https://doi.org/10.1002/ece3.688>
- State of Alaska, Department of Fish and Game (1978) Alaska's fisheries atlas. 2. State of Alaska, Department of Fish and Game
- Tande G, Lipkin R (2003) Wetland sedges of Alaska. Alaska Natural Heritage Program, Environmental and Natural Resources Institute, Anchorage, Alaska, 138 pp. URL: <http://aknhp.uaa.alaska.edu/botany/alaskan-wetland-sedge-species/>
- Thurber J, Peterson R (1991) Changes in body size associated with range expansion in the coyote (*Canis latrans*). *Journal of Mammalogy* 72 (4): 750-755. <https://doi.org/10.2307/1381838>
- U.S. Fish and Wildlife Service (1968) Birds of the Kenai National Moose Range. Refuge Leaflet 197-R2. United States Department of the Interior, Fish and Wildlife Service, Bureau of Sport Fisheries and Wildlife URL: <https://catalog.data.gov/dataset/birds-of-the-kenai-national-moose-range-1963>
- U.S. Fish and Wildlife Service, Region 7 (2010) Comprehensive conservation plan, Kenai National Wildlife Refuge. U.S. Fish and Wildlife Service, Region 7, Anchorage, Alaska, 532 pp. URL: https://www.fws.gov/uploadedFiles/Region_7/NWRS/Zone_2/Kenai/PDF/USFWS_2010_Kenai_CCP.pdf
- U.S. Geological Survey (2016) USDA Geospatial Data Gateway version of the Watershed Boundary Dataset (WBD). U.S. Geological Survey. URL: <http://nhd.usgs.gov/data.html>
- USGS Advanced Research Computing (2019) USGS Yeti Supercomputer. U.S. Geological Survey. URL: <https://doi.org/10.5066/F7D798MJ>
- Viereck LA, Dyrness CT, Batten AR, Wenzlick KJ (1992) The Alaska vegetation classification. General technical report PNW-GTR-286. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, 280 pp. <https://doi.org/10.2737/PNW-GTR-286>
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73 (16): 5261-5267. <https://doi.org/10.1128/aem.00062-07>

- Watts C, Dopheide A, Holdaway R, Davis C, Wood J, Thornburrow D, Dickie IA (2019) DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques. *Austral Entomology* 58 (3): 675-686. <https://doi.org/10.1111/aen.12384>
- Welsh SL (1974) *Anderson's Flora of Alaska*. Brigham Young University Press, Provo, Utah, 724 pp. [ISBN 0842507051]
- Wickham H (2007) Reshaping data with the reshape package. *Journal of Statistical Software* 21 (12): 1-20.
- Wolstad T (2010) *Trichodectes canis*, an invasive ectoparasite of Alaskan wolves: detection methods, current distribution, and ecological correlates of spread. University of Alaska Fairbanks, Fairbanks, Alaska. URL: <http://hdl.handle.net/11122/7279>
- Woodward A, Beever E (2010) Framework for ecological monitoring on lands of Alaska National Wildlife Refuges and their partners. U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia, 94 pp. <https://doi.org/10.3133/ofr20101300>
- Yi Z, Jinchao F, Dayuan X, Weiguo S, Axmacher J (2012) A comparison of terrestrial arthropod sampling methods. *Journal of Resources and Ecology* 3 (2): 174-182. <https://doi.org/10.5814/j.issn.1674-764x.2012.02.010>
- Yu D, Ji Y, Emerson B, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3 (4): 613-623. <https://doi.org/10.1111/j.2041-210x.2012.00198.x>

Endnotes

*1 Kenai National Wildlife Refuge Entomology Collection:
https://www.fws.gov/refuge/Kenai/what_we_do/science/specimens.html
<http://arctos.database.museum/KNWR ento>
<https://www.gbif.org/dataset/cb0eb5bf-4759-4aa1-92b5-4143b0a64e15>

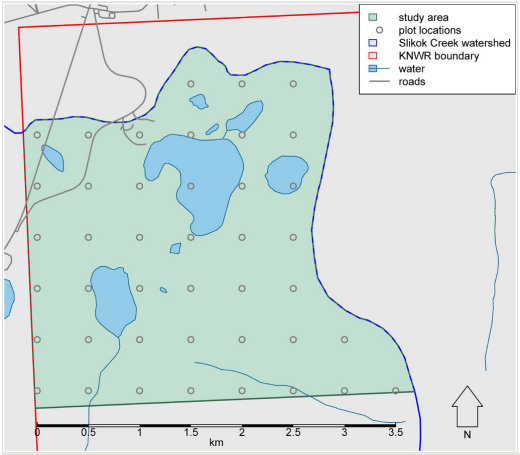


Figure 1.
Map of the Slikok Creek watershed study area and sampling design.

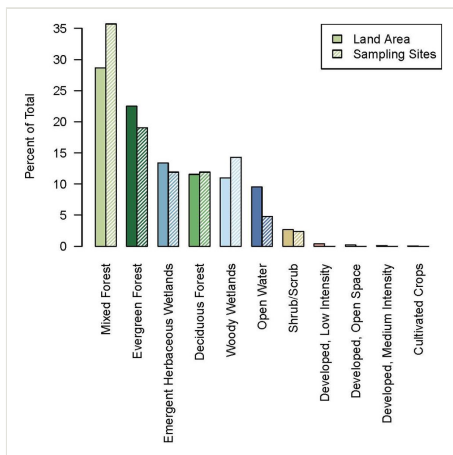


Figure 2.

Breakdown of the study area and sample frame by cover classes. The cover classes and colours of the bars are those defined in the National Land Cover Database (NLCD).



Figure 3.

A sampling site marked with a fibreglass rod and temporary flagging (image details: <https://doi.org/10.7299/X7F1901H>).

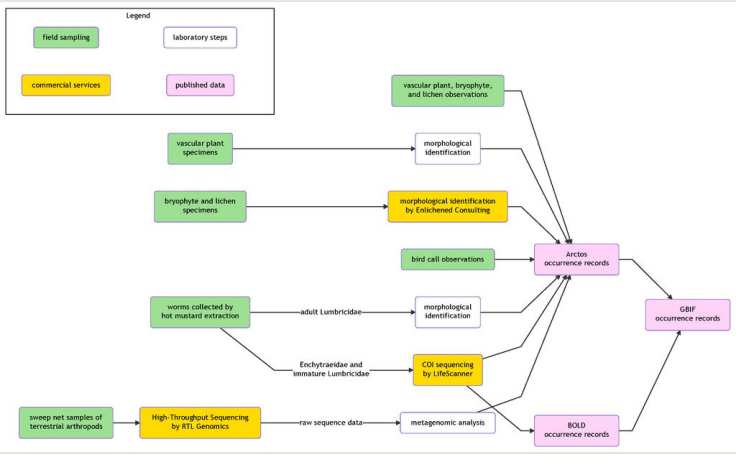


Figure 4.

Flowchart illustrating workflows from field sampling to publication of occurrence data. The flowchart was generated using the DiagrammeR package (Iannone 2020).

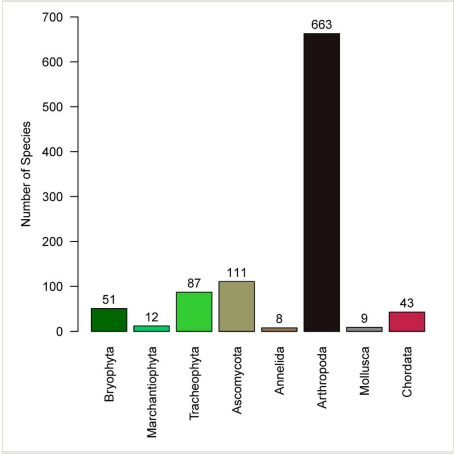


Figure 5.
Total numbers of species observed in each phylum.

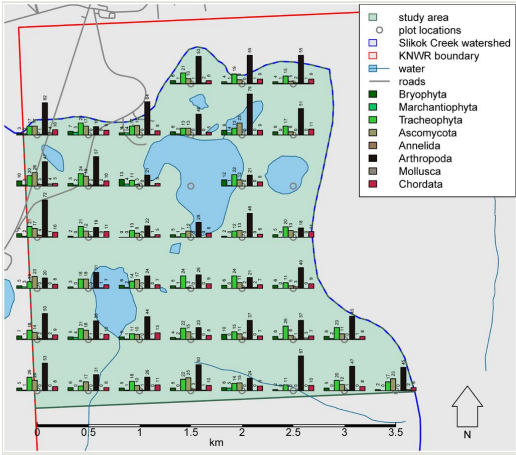


Figure 6.
Map showing numbers of species by phyla documented at each site.

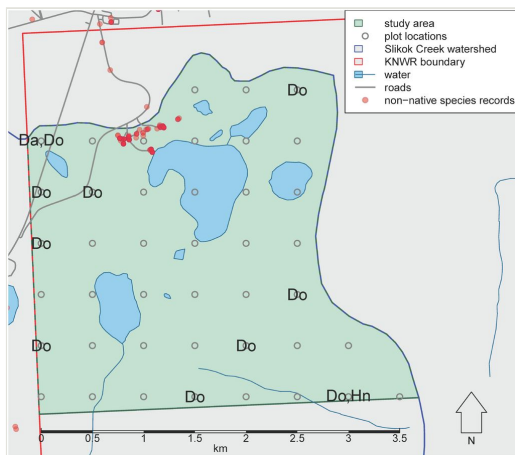


Figure 7.

Locations where non-native species were detected. Da: *Deroceras agreste*. Do: *Dendrobaena octaedra*. Hn: *Heterarthrus nemoratus*. Red dots signify non-native species records from AKEPIC (2016) and GBIF (2016).

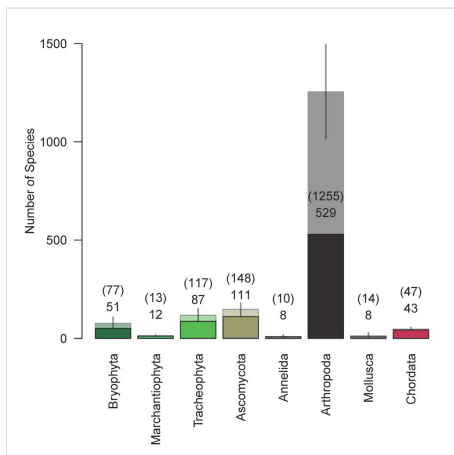


Figure 8.

Observed and estimated numbers of species from phyla in the analysis dataset. Darker boxes and lower numbers are observed numbers of species; paler boxes and upper numbers in parentheses are Chao estimates of the total species pool. Error bars are $2\times$ the standard errors of the Chao estimates except that the lower bounds of error bars were truncated at the observed numbers of species.

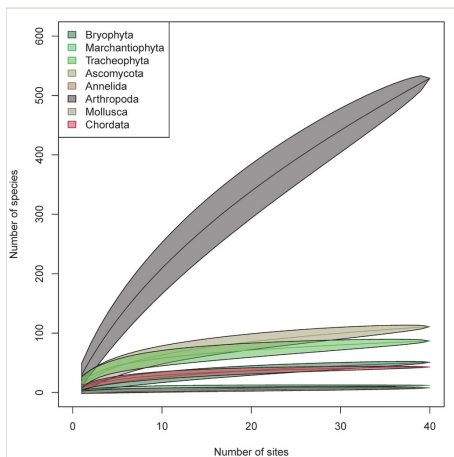


Figure 9.

Species accumulation curves based on the analysis dataset. Centre lines represent the estimates and the polygons indicate $\pm 2 \times$ standard error of the estimates.

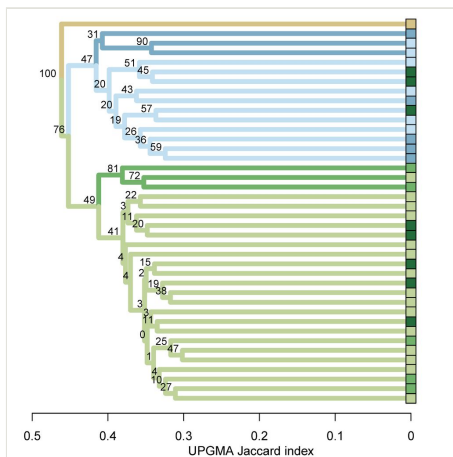


Figure 10.

UPGMA tree of communities at sampling sites, based on Jaccard coefficients. Colour-filled boxes represent sampling sites. Colours in filled boxes correspond to colours of land cover classes from Homer et al. (2015) and Fig. 2. Five groups have been highlighted by coloured tree branches.

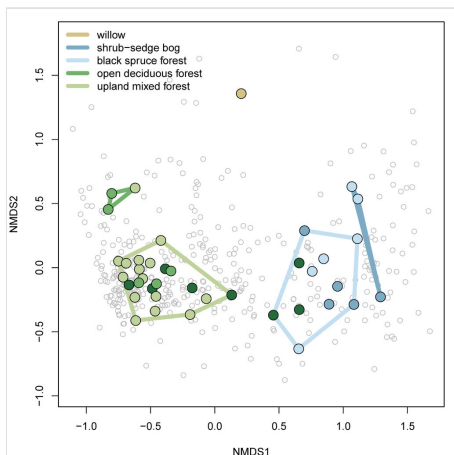


Figure 11.

Nonmetric Multidimensional Scaling biplot, NMDS axes 1 and 2. Colours of sites (filled circles) correspond to colours of land cover classes from Homer et al. (2015) and Fig. 2. Polygons are convex hulls of the groups identified by the cluster analysis (Fig. 10). Species are represented by open, grey circles.

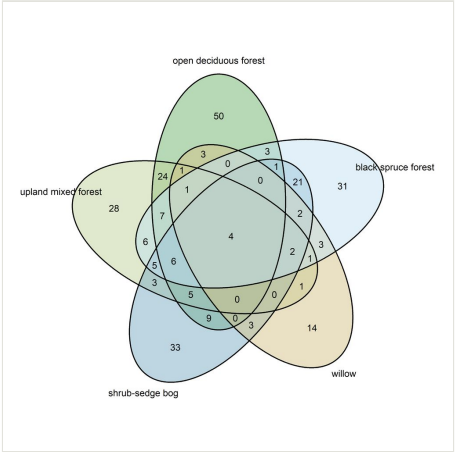


Figure 12.
Venn diagram of species included in the community groupings.

Table 1.

Observed and estimated numbers of species by phyla. Chao: Chao estimator. SE: estimate of the standard error of the chao estimate. Percent observed: percentage of species observed based on the Chao estimate. Slope: the number of species added per plot at the 39th plot.

Phylum	Observed	Chao	SE	Percent observed	Slope
Annelida	8	10	4	80%	0.05
Arthropoda	529	1255	119	42%	8.35
Ascomycota	111	148	16	75%	0.99
Bryophyta	51	77	15	66%	0.56
Chordata	43	47	4	91%	0.20
Marchantiophyta	12	13	2	92%	0.05
Mollusca	8	14	7	58%	0.10
Tracheophyta	87	117	16	74%	0.63

Supplementary materials

Suppl. material 1: Slikok Creek watershed study area map (KMZ)

Authors: Matthew L. Bowser

Data type: geographic features

Brief description: Slikok Creek watershed study area map in Keyhole Markup Language Zipped format.

Filename: Slikok_Creek_Watershed_Project_Map.kmz - [Download file](#) (2.54 kb)

Suppl. material 2: DNA extraction methods

Authors: Kelli Brooks

Data type: laboratory methods

Brief description: These are the DNA extraction methods that were provided by RTL Genomics used for sweep-net samples of terrestrial invertebrates.

[Download file](#) (2.51 kb)

Suppl. material 3: Lichen and bryophyte identifications

Authors: Trevor Goward and Curtis Björk

Data type: identifications

Brief description: This spreadsheet contains the original identifications of specimens shipped to Trevor Goward for identification. In the file, the "Arctos_GUID" field contains the original Globally Unique Identifier of the Kenai National Wildlife Refuge's herbarium samples from which the specimens came. Related data are available on-line via Arctos. For example, a sample with a GUID of KNWR:Herb:10449 is available at the URL <http://arctos.database.museum/guid/KNWR:Herb:10449>.

[Download file](#) (39.47 kb)

Suppl. material 4: ASV table

Authors: Matthew L. Bowser

Data type: ASV table

Brief description: In this spreadsheet, rows represent ASVs and columns correspond to Arctos Globally Unique Identifiers of bulk sweep-net samples.

[Download file](#) (610.29 kb)

Suppl. material 5: ASV sequences

Authors: Matthew L. Bowser

Data type: ASV sequences in FASTA format

Brief description: This file contains ASV sequences from sweep-net samples in FASTA format.

[Download file](#) (697.83 kb)

Suppl. material 6: Raw occurrence data from the Slikok watershed biotic inventory

Authors: Matthew L. Bowser, Rebekah Brassfield, Annie Dziergowski, Todd Eskelin, Jennifer Hester, Dawn Robin Magness, Mariah McInnis, Tracy Melvin, John M. Morton and Joel Stone

Data type: occurrences

Brief description: This dataset was downloaded from Arctos on 19 November 2019. Coordinate uncertainties vary greatly by method, ranging from 3 m for 0.25 m² earthworm quadrats to hundreds of metres for birds observed on variable circular plots.

[Download file](#) (2.78 MB)

Suppl. material 7: Arthropod species newly reported from Alaska

Authors: Matthew L. Bowser

Data type: species checklist

[Download file](#) (22.36 kb)

Suppl. material 8: BOLD TaxonID Tree for SlikokOtu1170

Authors: Matthew L. Bowser

Data type: phylogenetic tree

Brief description: BOLD TaxonID Tree for SlikokOtu1170 generated by BOLD's Identification Engine, 26.December.2019

[Download file](#) (8.45 kb)

Suppl. material 9: Analysis dataset

Authors: Matthew L. Bowser, Rebekah Brassfield, Annie Dziergowski, Todd Eskelin, Jennifer Hester, Dawn Robin Magness, Mariah McInnis, Tracy Melvin, John M. Morton and Joel Stone

Data type: occurrences

Brief description: This is the derived dataset created for the purpose of analysis. It includes only species-resolution identifications and only the 80 sweep-net samples taken from the east half of each plot.

[Download file](#) (2.24 MB)

Suppl. material 10: Community composition

Authors: Matthew L. Bowser

Data type: species lists

Brief description: This file lists the species assigned to each of five community types.

[Download file](#) (18.39 kb)