NRC Research Press

ARTICLE

DNA barcode data reveal biogeographic trends in Arctic non-biting midges

Torbjørn Ekrem, Elisabeth Stur, Matthew G. Orton, and Sarah J. Adamowicz

Abstract: Chironomid flies (non-biting midges) are among the most abundant and diverse animals in Arctic regions, but detailed analyses of species distributions and biogeographical patterns are hampered by challenging taxonomy and reliance on morphology for species-level identification. Here we take advantage of available DNA barcode data of Arctic Chironomidae in BOLD to analyse similarities in species distributions across a northern Nearctic – West Palearctic gradient. Using more than 260 000 barcodes representing 4666 BINs (Barcode Index Numbers) and 826 named species (some with interim names) from a combination of public and novel data, we show that the Greenland chironomid fauna shows affinities to both the Nearctic and the West Palearctic regions. While raw taxon counts indicate a strong Greenland – North American affinity, comparisons using Chao's dissimilarity metric support a slightly higher similarity between Greenland and West Palearctic chironomid communities. Results were relatively consistent across different definitions of species taxonomic units, including morphologically determined species, BINs, and superBINs based on a ~4.5% threshold. While most taxa found in Greenland are shared with at least one other region, reflecting circum-Arctic dispersal, our results also reveal that Greenland harbours a small endemic biodiversity. Our exploratory study showcases how DNA barcoding efforts using standardized gene regions contribute to an understanding of broad-scale patterns in biogeography by enabling joint analysis of public DNA sequence data derived from diverse prior studies.

Key words: Diptera, zoogeography, biogeography, Canada, Norway, Svalbard.

Résumé : Les chironomidés (moucherons non-piqueurs) sont parmi les animaux les plus abondants et diversifiés dans les régions arctiques, mais des analyses détaillées des distributions de ces espèces et de leur répartition biogéographique sont entravées par une taxonomie difficile et une dépendance à la morphologie pour l'identification des espèces. Ici, les auteurs tirent profit des données disponibles en matière de codes à barres de l'ADN pour les chironomidés arctiques au sein de la base de données BOLD pour analyser les similarités pour ce qui est des distributions des espèces au long d'un gradient nordique Néarctique-Paléarctique occidental. Au moyen de plus de 260 000 codes à barres représentant 4666 BIN (« Barcode Index Numbers ») et 826 espèces nommées (dont certaines de manière intérimaire) à partir d'une combinaison de données publiques et inédites, les auteurs montrent que la faune des chironomidés du Groenland montre des ressemblances à la fois avec celles des régions Néarctique et Paléarctique occidentale. Bien que des décomptes bruts indiquent une grande affinité Groenland – Amérique du Nord, des comparaisons à l'aide de l'indice de dissimilarité de Chao suggèrent une similarité légèrement plus élevée entre les communautés du Groenland et du Paléarctique occidental. Les résultats étaient relativement cohérents en faisant appel à diverses définitions des unités taxonomiques spécifiques, incluant les espèces définies sur une base morphologique, des BIN et des superBIN à un seuil de 4,5 %. Bien que la plupart des taxons du Groenland soient partagés avec au moins une autre région, ce qui reflète une distribution circum-arctique, les résultats révèlent aussi que le Groenland recèle une petite biodiversité endémique. Cette étude exploratoire démontre comment les efforts en matière de codage à barres au moyen de régions géniques standardisées contribuent à la compréhension de la biogéographie à grande échelle en rendant possible l'analyse conjointe de données moléculaires publiques dérivées de diverses études antérieures. [Traduit par la Rédaction]

Mots-clés : Diptera, zoogéographie, biogéographie, Canada, Norvège, Svalbard.

Introduction

The Arctic is home to about 21 000 described species, many of which are specifically adapted to a cold climate (CAFF 2013). As this diversity is under pressure from several sources, perhaps most notably a warmer climate, it is increasingly important to have a sufficient overview of species distributions and taxonomic boundaries. Arthropods form a major part of terrestrial and freshwater biodiversity in the Arctic (Coulson et al. 2014; Danks 1981; Høye and Sikes 2013), and many species are sensitive to changes in their surroundings. Hence, increased knowledge of their ecology, phylogeny, and biogeography can help us understand the present diversity patterns as well as predict future change.

Flies of the family Chironomidae, or non-biting midges, are extremely common in the Arctic. In some high Arctic regions they can be the dominating invertebrates (Wirta et al. 2016), especially of freshwater habitats (Chertoprud et al. 2017) where their imma-

Received 15 May 2018. Accepted 24 September 2018.

T. Ekrem and E. Stur. Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway; Centre for Biodiversity Genomics, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON N1G 2W1, Canada.
M.G. Orton and S.J. Adamowicz.* Centre for Biodiversity Genomics, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON N1G 2W1, Canada; Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada.
Corresponding author: Torbjørn Ekrem (email: torbjorn.ekrem@ntnu.no).
*Sarah J. Adamowicz currently serves as an Associate Editor; peer review and editorial decisions regarding this manuscript were handled by Dario Lijtmaer.

ture stages are an important food source for fish (Berg et al. 2010). They also play a central role in the decomposition of organic matter and other parts of the nutrient cycle (Baranov et al. 2016; Berg 1995; Shang et al. 2013). Non-biting midges are key components of Arctic terrestrial food chains (Wirta et al. 2015), and their abundance, particularly in the high Arctic, make them essential food for nesting birds (Hodkinson et al. 1996). Many chironomids have specific habitat requirements and are often nutrient and temperature sensitive (Brodersen and Anderson 2002). This makes them good environmental indicators. In concert with general high diversity and wide distributions, ecological preferences make chironomids interesting study objects in biogeography and palaeoecology (Armitage et al. 1995; Brooks et al. 2007; Ferrington 2008). However, the taxonomy of many chironomids frequently is challenging, and they can be difficult to identify using morphology. Thus, the use of genetic metabarcoding to analyse and monitor community change can clearly be beneficial and is already being applied in freshwater ecology (Andujar et al. 2018; Bista et al. 2017).

Apart from a few isolated coastal areas, most of the Arctic was completely covered in ice during the last glacial maximum (Jakobsson et al. 2014). Moreover, the few ice-free land masses that did exist must have been fairly inhospitable to the majority of terrestrial invertebrates. Thus, most of the terrestrial arthropod species in the Arctic region today likely dispersed into their current ranges over the last 10 000 – 12 000 years. The exact colonization routes differ between organisms and regions, and details are still unknown for many taxa. However, for the insects of Greenland, it is shown that while some groups, such as Lepidoptera, Trichoptera, and Araneae, have taxonomic affinities with the Nearctic fauna, others, such as Coleoptera and Collembola, are clearly dominated by Palearctic elements (Böcher 2015). It is remarkable that Greenland taxa with poor dispersal capabilities show greatest similarities with the Palearctic and not the geographically closer Nearctic. Thus, hypotheses for the origin of these taxa in Greenland includes dispersal by birds or humans, or interglacial survival, for instance along the east coast (Böcher 2015). It has been argued that most non-biting midges of Greenland have likely immigrated from North America as a large percentage of the known fauna consist of species with Nearctic or Holarctic distributions (Lindegaard 2015). However, there are Palearctic elements present in eastern Greenland, and several regions are likely not investigated well enough to have a clear picture of other major immigration routes. It is obvious, however, that the diversity in southern Greenland is poorer than regions of similar latitude in North America and Eurasia, indicating that the continent was mostly populated from the north (Lindegaard 2015).

Although the arthropod diversity of some Arctic regions is comparatively well known (Böcher et al. 2015), there is a general lack in knowledge of true distributions, ecology, and even species boundaries for many taxa (Hodkinson et al. 2013). This is at least partly due to the reliance of morphology for identification and different taxonomic traditions. For instance, the use of morphological keys to Central European Chironomidae when identifying stream macroinvertebrates in Svalbard can lead to false positives and extended northward distributions of boreal taxa (own observations). The use of short, standardized gene fragments to identify specimens to species level (i.e., DNA barcoding (Hebert et al. 2003)) has changed the game and brought more objectivity into the documentation of biodiversity (Hubert and Hanner 2015). In Chironomidae, the use of DNA barcodes has aided discovery of cryptic species, associated life stages, and documented diversity that otherwise would remain undetected (Anderson et al. 2013; Ekrem et al. 2010; Lin et al. 2018; Silva et al. 2013; Stur and Ekrem 2011, 2015).

Observed biogeographical patterns might depend on the concept used to define species as the analytical entity. For instance, the use of the phylogenetic species concept often results in more species than the biological species concept within the same taxonomic group, and finer taxonomic resolution can infer different biogeographical patterns (Peterson 2006). However, previous studies of endemism and macroecological patterns in Afrotropical birds (Dillon and Fjeldså 2005), southeast Asian primates (Nijman and Meijaard 2008), Siberian freshwater snails (Vinarski and Kramarenko 2015), and world Ichneumonidae (Jones et al. 2012) all observed the generally same patterns when analysing datasets of different taxonomic resolution at the species level. Thus, it can be informative to use different definitions (or levels) of species when analysing biogeographical patterns to assess the consistency of patterns at different phylogenetic depths, which may approximately correlate with evolutionary time.

DNA barcode data in large collaborative databases such as the Barcode of Life Data System (BOLD; Ratnasingham and Hebert 2007) are perfectly suited to explore species distributions and biogeographical patterns since they contain many data points across a broad geographical scale. Moreover, the marker standardization inherent in DNA barcoding efforts enables large analyses of sequence data originating from multiple published sources, and allows analysis at different species levels. A recent study investigated patterns in North American Noctuidae (Lepidoptera) (Zahiri et al. 2017), but we are unaware of papers that have compared extensive georeferenced DNA barcode data of terrestrial or aquatic invertebrates across the Arctic or northern Atlantic. Our aim is to explore trans-Atlantic diversity patterns of Arctic Chironomidae using DNA barcodes to see if current public DNA sequence information reveals new biogeographical information for this biologically important family in northern regions. Specifically, we aim to resolve whether the chironomid fauna of Greenland exhibits a stronger affinity to the northern Nearctic or western Palearctic, which would indicate overarching dispersal trends across the Arctic Atlantic region. Moreover, our exploratory study considers multiple definitions of species units, including morphological species, Barcode Index Numbers (BINs) (Ratnasingham and Hebert 2013), and "superBINs" (BINs clustered at the ~4.5% divergence level), and asks whether biogeographic patterns are similar at different taxonomic levels.

Materials and methods

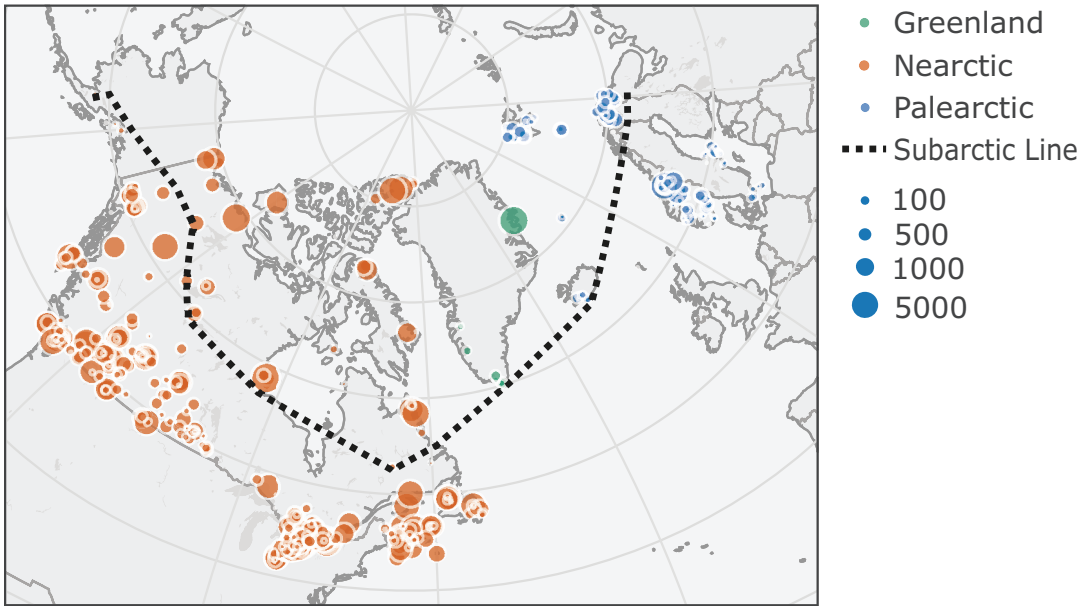
Field sampling, sorting, and identification

Chironomids were sampled in connection with a variety of projects and excursions in Churchill (Manitoba, Canada), Finnmark (Norway), Svalbard (Norway), Iceland, and Greenland. Adult males and females were typically collected using a sweep net or Malaise trap, while aquatic immatures were sampled with kick-nets, drift nets, or van Veen grabs, and terrestrial immatures with Berlese funnels. Specimens were preserved in ethanol (85% for adults, 96% for immatures) and kept cool at 5 °C until further processing. Samples were sorted under a stereomicroscope, and a selection of specimens of each morphotype was slide mounted in Euparal for identification in a compound microscope. A tissue sample (usually one or two legs) was dissected off prior to slide mounting and used for DNA extraction.

Molecular analysis of collected specimens

DNA was extracted using standard insect tissue extraction protocols at the Canadian Centre for DNA Barcoding (CCDB), University of Guelph. PCR and Sanger sequencing used either the Folmer primer set (LCO1490 + HCO2198) (Folmer et al. 1994), the Lep primers (LepF1 + LepR1) (Hebert et al. 2004), or a cocktail of these (C_LepFolF + C_LepFolR) (Hernández-Triana et al. 2014). Novel sequences, trace-files, and metadata are publicly available through

Fig. 1. Map of study region, including Canada, Greenland, and Scandinavia, from which Chironomidae DNA barcode records were analyzed. Circles represent individual sites where barcoding occurred. Sites are colour coded according to region and sized according to the ln-transformed number of barcodes per site. Dotted black line represents the Conservation of Arctic Flora and Fauna (CAFF) subarctic zone boundary.



the Barcode of Life Data Systems (BOLD) (Ratnasingham and Hebert 2007) under the dataset DS-ARCTCHIR, doi:10.5883/DS-ARCTCHIR.

Database mining and data filtering

In addition to the novel data provided here, we mined for northern trans-Atlantic public data using the BOLD API. Data were retrieved based on taxonomy (Chironomidae) and geography, including all available records from Alaska (USA), Canada, Finland, Greenland, Iceland, Norway, and Sweden. These regions were chosen as they contain a high number of records in BOLD and are bordering the northern Atlantic. A total of 258 666 records matched these criteria at the time of download (20 October 2017).

For both public and novel data, records were filtered to retain those having a BIN identifier, latitude and longitude coordinates, and a sequence present for the cytochrome *c* oxidase subunit 1 (COI) 5′ marker, using R, version 3.4.2 (R Core Team 2018), with the package readr, version 1.1.1 (Wickham et al. 2017). The records were further quality checked by one-by-one manual examination of those BINs that deviated greatly from the typical patterns of genetic variability within family Chironomidae, and removed if these could be attributed to misidentifications or contaminations. The dataset size was first reduced by selecting the first available full-length barcode sequence (658 bp) from each BIN whenever possible, or alternatively the longest sequence if no 658 bp sequences were present in the BIN. Next, an alignment of the selected sequences was created using the muscle algorithm (Edgar 2004) (settings: diags = TRUE, gapopen = −3000 and maxiters = 2) from the R package muscle, version 3.18.0 (available from http://bioconductor.org/packages/muscle/), and a pairwise distance matrix of sequences was generated using the Tamura–Nei (Tamura and Nei 1993) distance model (TN93) from the package ape, version 4.1 (Paradis et al. 2004). For each included sequence, we calculated its average pairwise distance to all other sequences in the dataset and displayed these in a boxplot. Those sequences that were more than 1.5 times the inter-quartile distance above the upper quantile were exported for manual examination. By examining the specimen photographs and using the taxon identification tool in BOLD we investigated the outliers and removed the records of one BIN (BOLD:ACZ1013) from the dataset.

After combining the novel and public data, the quality-filtered dataset (named FULL) contained in total 261 523 chironomid barcode records distributed across the northern trans-Atlantic region (Fig. 1). The FULL dataset was further filtered by only including records originating north of the subarctic boundary as defined by the Programme for the Conservation of Arctic Flora and Fauna (CAFF) shapefile data using the packages raster (Hijmans 2017), sp (Pebesma and Bivand 2005), and rgeos (Bivand and Rundel 2017) for filtering of the CAFF shapefile retrieved 5 December 2017 from CAFF/ABDS - GeoNetwork Catalogue - The Arctic Biodiversity Data Service (ABDS) (http://geo.abds.is/geonetwork/srv/eng/catalog.search#/home). This data subset (named ARCTIC) consisted of 66 765 records distributed in sub-Arctic and Arctic trans-Atlantic regions (Fig. 1, above dotted line). To visualize the variability in sampling/sequencing effort among sites (Fig. 1), the sizes of the circles were scaled to the ln-transformed number of barcodes plus one. Sites were defined by first rounding GPS coordinates, which are in decimal degree format in BOLD, to the nearest 0.1. Following rounding, each unique combination of latitude and longitude was treated as a site.

Taxonomic units for analysis

Biogeographic analyses were performed at three taxonomic levels. First, we used BINs, which are assigned by BOLD for uploaded records that are 500 bp in length or greater within the COI barcode marker, and that contain fewer than 1% undetermined nucleotides. The BIN algorithm performs preliminary clustering of sequences based upon a threshold of 2.2% (p-distances) and then refines the clusters on the basis of continuity versus discontinuity in the pattern of genetic divergences among neighbouring sequences (Ratnasingham and Hebert 2013). Second, we generated "superBINs" clustered at the ~4.5% sequence divergence threshold to yield clusters that are closer to biological or Linnaean species of chironomids (Ekrem et al. 2007; Lin et al. 2015). Being based on DNA sequences alone, this method has the advantage that it can be applied to the full chironomid public dataset, including records that lack species-level taxonomic assignments. Third, we

analyzed distributional patterns using morphological species-level identifications. Manual curation of the taxonomic dataset was required to remove obvious errors in name assignments for some taxa (see methods on filtering described above). The curated dataset bearing species-level identifications consisted of 98 168 records.

For the superBIN clustering analysis, we considered BINs to be units that would either remain separate or could be grouped together by using a higher sequence divergence threshold for clustering. Thus, BINs are entirely nested within superBINs. After performing an alignment of a single maximal-length sequence per BIN using the muscle algorithm, a pairwise distance matrix was estimated using the TN93 distance model. We suggest that it is preferable to use a model that takes unequal base frequencies into account, given that insect mitochondrial genomes are highly AT-biased. Our choice of a distance metric, as well as clustering threshold, for the superBINs was also informed by the results of Lin et al (2015), who used the HKY model to perform GMYC analysis, which seeks to group sequences into evolutionary species. The matrix was then used as the basis for clustering the sequences using a 4.5% distance threshold in the R package DECIPHER, version 2.4.0 (Wright et al. 2012). For the FULL dataset, superBINs generated with a 4.0% and 5.0% threshold were also generated for comparison to see if slightly different levels of threshold would make a large impact on the number of clusters, and to investigate roughly which threshold level compared best with the number of assigned species in our taxonomically best-curated data. Single-linkage clustering, a type of hierarchical clustering, was selected to mimic as closely as possible methods from Lin et al. (2015). All other sequences within BINs were then assigned to their superBIN for biogeographic analysis. Since there also is genetic divergence within each BIN, the complete superBINs in most cases show slightly larger genetic divergence than 4.5%.

Biogeographic analysis

Each record was defined as belonging to the Nearctic (Canada and Alaska, USA), Greenland, or Palearctic (Iceland, Norway, Finland, Sweden) regions. For each region, completeness of sampling was assessed through examining a BIN accumulation curve by number of specimens sampled (Fig. 3A), generated using base R packages, version 3.4.2 (R Core Team 2018), or by site (Fig. 3B), generated using the function specaccum in the package vegan, version 2.5-1 (Oksanen et al. 2018). Resampling was performed 100 times for each iteration per region. Sites were defined by coordinates down to 0.1 decimal degree.

For each taxonomic level (BINs, superBINs, species), we ascertained which geographic region(s) were occupied by each taxon. These data were visualized through Venn diagrams showing the unique biodiversity, as well as overlap in composition, among the Nearctic, Greenland, and Palearctic regions.

To quantify compositional similarity we used Chao et al.'s (2005) pairwise similarity indices that takes unseen species in unequal sample sizes into account. The index was calculated for each pair of geographic zones on species, BINs, and superBINs using the R package vegan, version 2.5-1 (Oksanen et al. 2018). The Chao et al. (2005) index is returned as a dissimilarity value varying from 0 (highest similarity) to 1 (highest dissimilarity, no overlap in community composition).

Data accessibility and reproducibility of analyses

The full dataset used in biogeographic analyses and Novel R code is available through GitHuB: https://github.com/m-orton/Arctic-Biogeographical-Analysis-Pipeline.

Results

Taxonomic richness across different definitions of species units

The FULL dataset contained a total of 261 523 barcode records, representing 4666 BINS, 3072 superBINs (~4.5% divergence threshold), and 826 morphologically defined species. By contrast, the ARCTIC-filtered dataset included 66 765 barcode records, which encompassed 1520 BINs, 1096 superBINs, and 548 morphological species. In both datasets, there was a similar reduction in the MOTU count when moving from BINs (2.2% clustering seed threshold) to superBINs, with a 34% decline in the MOTU count in the FULL dataset and a 28% decline in the ARCTIC region when going from BINs to superBINs. These results therefore indicate that a substantial fraction of chironomid BINs have close relatives that are within 4.5% sequence divergence (Fig. 2).

The number of MOTUs at the 4.0% divergence threshold was slightly higher than the richness at the 4.5% threshold, while the number of MOTUs at 5.0% threshold was comparatively lower. For example, for the FULL dataset, the 4.0% clustering threshold yielded 3223 MOTUs, while the 5.0% divergence threshold resulted in 2622 clusters. The number of MOTUs generated with the 4.5% threshold corresponded best with the number of identified species for our taxonomically best curated data from Greenland and the Palearctic. Thus, downstream biogeographic analyses on superBINs were performed for the 4.5% threshold.

Sampling completeness among geographic regions

Accumulation curves of BINs for the ARCTIC dataset reveal differences in sampling completeness among regions (Fig. 3). In the specimen-based analysis (Fig. 3A), the curve for Greenland indicates the lowest taxon richness and the highest level of sampling completeness. By contrast, the Palearctic region has the steepest accumulation curve, reflecting the lowest sampling completeness, while the efforts in North America are more reflective of broad-scale barcoding of samples, in which more duplicates of common species are sequenced. When looking at the BIN richness per site barcoded, though (Fig. 3B), the difference between the Palearctic and Nearctic regions is not as apparent; the accumulation curve by site is steeper for North America, reflecting the greater geographic spread of sampling sites. The low number of sites for Greenland also becomes clearly visible in this graph.

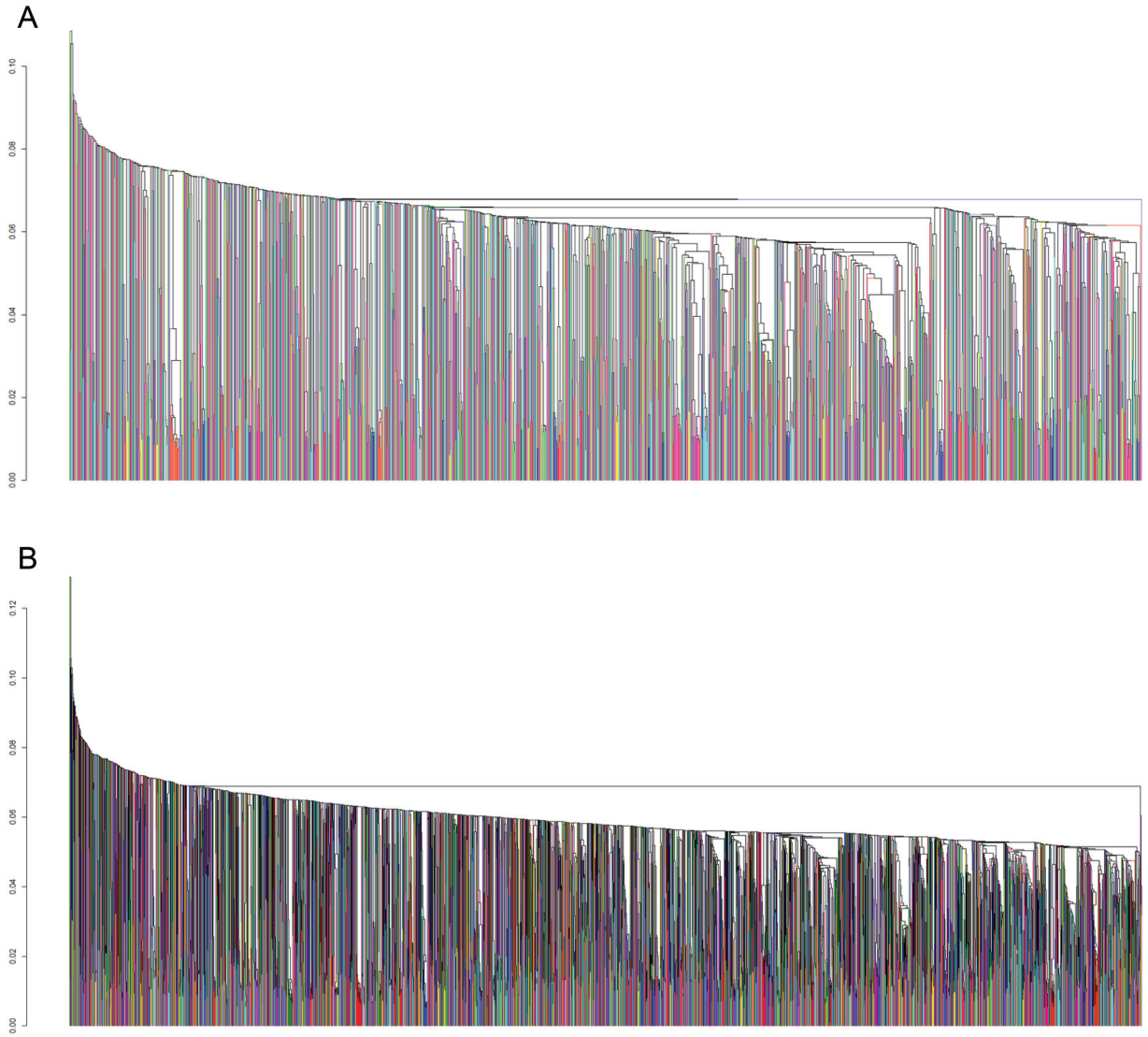
Trans-Arctic biogeography

Based on raw taxon counts, the western Palearctic and northern Nearctic regions shared more taxa with one another than either did with Greenland, while Greenland shared more taxa with the Nearctic region than with the Palearctic region for BINs and superBINs (Fig. 4). For example, for the FULL dataset, 48 BINs were share among all three regions; 10 BINs were shared between Greenland and West Palearctic, while 73 were in common between Greenland and the Nearctic. For the species category, Greenland shared roughly the same number of taxa with both the northern Nearctic and the western Palearctic regions (Fig. 4).

By contrast, based upon the Chao dissimilarity metric, which considers sampling intensity and unsampled species, Greenland displays a higher affinity for the Palearctic region, as indicated by a lower dissimilarity index (Table 1). This trend was also consistent across different taxon units (BINs, superBINs, and species) as well as between the FULL and ARCTIC datasets (Table 1).

In an additional analysis of BINs from the FULL dataset, the eastern side of Greenland displayed a slightly higher Palearctic affinity (Chao et al. dissimilarity index of 0.895) than Nearctic (0.921), mirroring the pattern in the FULL dataset when considering Greenland in its entirety. Similarly, western Greenland (which is poorly sampled in comparison to the east) exhibited a near-even but slightly higher Palearctic (0.944) affinity than Nearctic (0.951).

Fig. 2. Phenogram of the (A) ARCTIC and (B) FULL DNA barcode datasets, whereby each tip contains one sequence selected from each BIN (Barcode Index Number). The branches are coloured by superBINs (with colour palette repeating along the figure), demonstrating the high prevalence of BINs with close relatives that merge into superBINs at the ~4.5% clustering level.



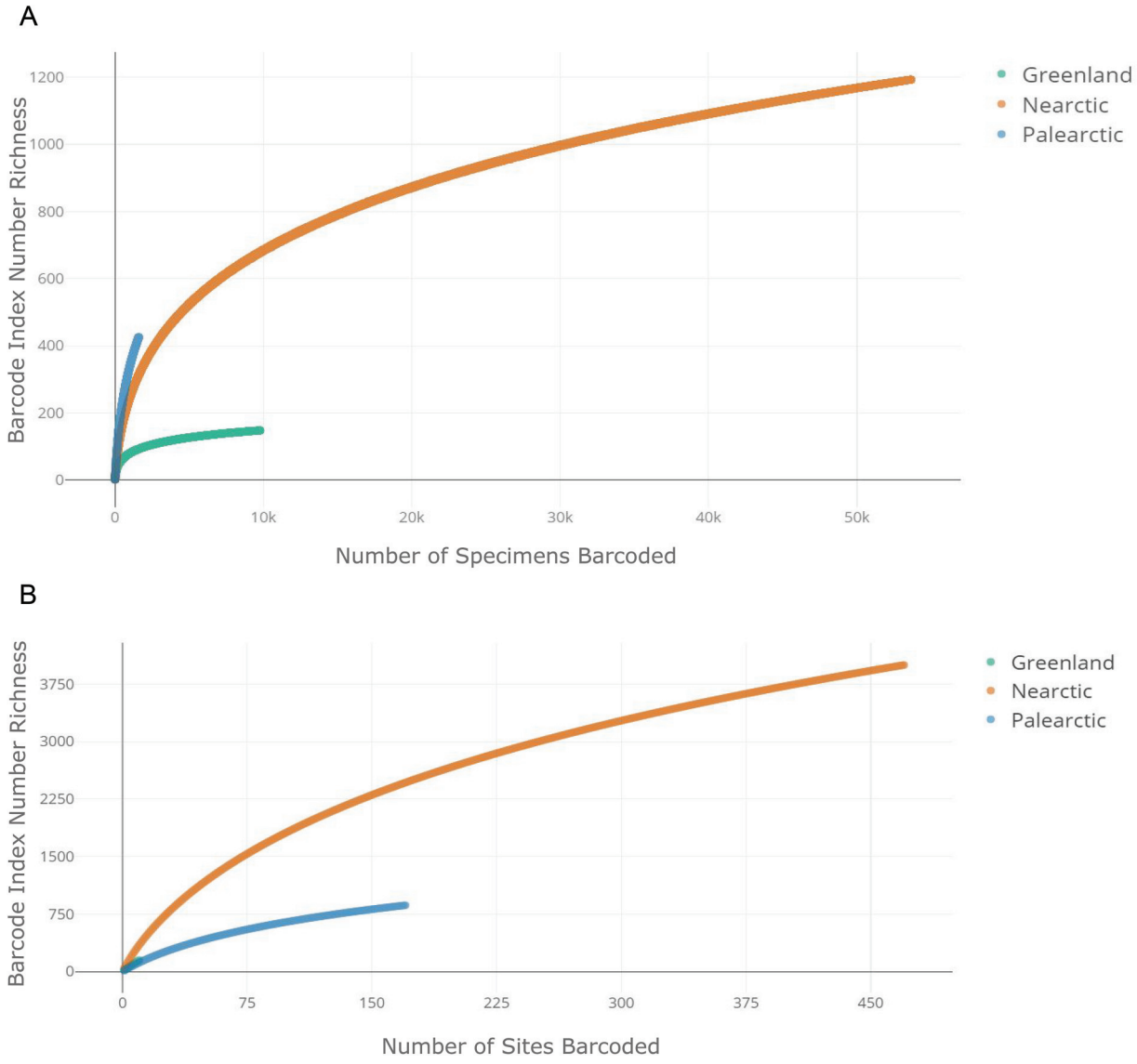
Therefore, all datasets supported that the Greenland chironomid fauna is slight more similar to the western Palearctic fauna than the northern Nearctic.

Each region exhibited more endemic biodiversity when considering BINs than superBINs (Fig. 4). For example, in the ARCTIC dataset, Greenland exhibits 21 endemic BINs but just 7 endemic superBINS, and only a single endemic morphological species (an unnamed species of the genus *Smittia*). The ARCTIC dataset exhibited a greater proportion of taxa shared among regions than the FULL dataset. For example, in the FULL dataset, 6.4% of BINs are shared between two or more regions, while in the ARCTIC dataset, 13.7% of BINs are shared between two or more regions.

Discussion

The Greenland chironomid fauna clearly has affinities to both the northern Nearctic and western Palearctic regions. In data that are uncorrected for differences in the sampling effort, the Greenland fauna shares more BINs and superBINs with the Nearctic region, while shared morphological species are more similar between the regions (Fig. 4). The major reason for this difference is that a higher percentage of the barcode records are identified to species in Greenland and the Palearctic compared to the Nearctic. When using a metric that accounts for unseen species in unequal sampling regimes, the Greenland fauna shows a stronger connection to the Palearctic (Table 1). This is rather surprising given the closer geographic proximity that Greenland has to the Nearctic, and current knowledge on the Chironomidae fauna of Greenland has indicated stronger Nearctic relations (Lindegaard 2015). However, Palearctic connections of the Greenland entomofauna are found in aphids, beetles, and springtails (Böcher 2015), and transport of chironomids to Greenland by the Polar easterlies is not unlikely as midges can be transported over considerable distances as aerial plankton (Glich 1960; Holzapfel and Perkins 1969). Moreover, long-distance dispersal routes from the east have been documented in vascular plants (Alsos et al. 2015), even with sea ice and driftwood as potential dispersal agents (Alsos et al. 2016). Of importance to our results is that the Greenland dataset included in our analyses is dominated by a major project in eastern Greenland (Wirta et al. 2016), and that the western part of the continent is comparatively poorly sampled for DNA barcode data, but even the sampled chironomid community in western Greenland is slightly more similar to the sampled Palearctic community. While

Fig. 3. BIN accumulation curves for the western Palearctic (blue line), northern Nearctic (orange line), and Greenland (green line) regions. (A) BIN richness by number of specimens barcoded. (B) BIN richness by number of sites barcoded. A total of 100 replicates were performed per region.



a comparison of eastern and western Greenland would be interesting once further data become available, our results suggest a stronger Palearctic affinity for the chironomid fauna of Greenland.

It is challenging to account for differences in sampling method and intensity when calculating comparative diversity metrics, and even indices that account for missing shared species are impacted negatively by sampling bias (Chao et al. 2005). The data used in our study originate from projects with different scopes and of different magnitudes. They are therefore not easy to use in direct comparisons, even if all projects have used similar techniques for acquiring specimens. For instance, sampling intensity has been considerably higher in the Canadian Arctic than in the other regions (Fig. 3), and Greenland has been most intensively sampled at one locality (Zackenberg), perhaps leading to an accumulation curve that levels off a bit early (Fig. 3A). The chironomid data from the Palearctic region was generated with the goal of registering as many species as possible in the barcode reference library, and comparatively few specimens of each sampled morphotypes were selected for analysis. The Palearctic data are therefore not based on random sampling, and as evidenced by the BIN richness by site curve (Fig. 3B), the steep accumulation curve in Fig. 3A is not representative for our actual progress in achieving species coverage in the European Arctic. The strong differences in sampling intensity and variability in the geographic coverage of sampling among circum-arctic regions likely contribute to the conflicting results between the raw data and the dissimilarity metric. Our study highlights the potential contributions that large-scale DNA barcoding campaigns can make to macroecology and biogeography. With further DNA barcoding campaigns being undertaken globally, species coverage is expected to approach completion over time, particularly for taxonomic groups that are readily captured using passive collecting techniques (such as Malaise traps). Future research could involve repeating our study as well as comparing biogeographic patterns among taxonomic groups, combined with statistical hypothesis testing for trends.

As is true for most taxonomic challenging groups, the identification of chironomids to species is influenced by the taxonomic knowledge (published and unpublished), taxonomic tradition, access to literature, and reference collections. Thus, the comparison

Fig. 4. Venn diagrams displaying the overlap and unique components of biodiversity in the Palearctic, Nearctic, and Greenland regions under differing definitions of species units. The three left panels show the FULL DNA barcode dataset, right panels show the ARCTIC dataset, with species units defined according to BINs (top), superBINs at the ~4.5% clustering threshold (middle row), and morphological species (bottom).

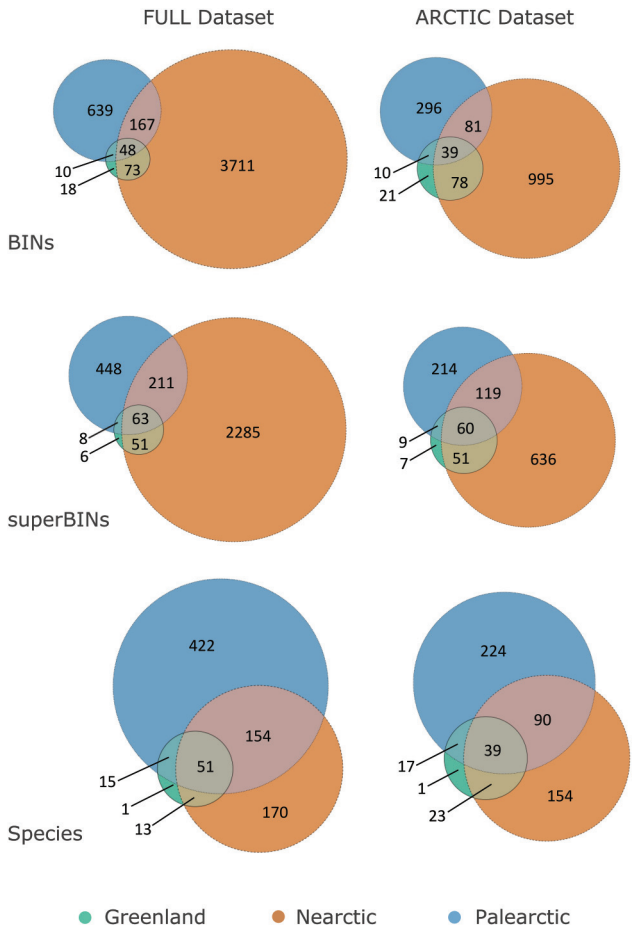


Table 1. Community dissimilarity among regions using Chao et al.'s (2005) index as implemented in the R package vegan.

	FULL dataset		ARCTIC dataset	
	Greenland	Nearctic	Greenland	Nearctic
BINs	*N* = 4666		*N* = 1520	
Nearctic	0.928	—	0.775	—
Palearctic	**0.895**	0.846	**0.671**	0.780
SuperBINs (~4.5%)	*N* = 3072		*N* = 1096	
Nearctic	0.901	—	0.676	—
Palearctic	**0.868**	0.788	**0.593**	0.638
Species	*N* = 826		*N* = 548	
Nearctic	0.925	—	0.833	—
Palearctic	**0.814**	0.573	**0.687**	0.630

Note: 0 = most similar; 1 = most dissimilar. Values in bold indicate pairwise regions with highest community similarity for Greenland.

of morphologically determined records originating from different identifiers can lead to both false negative and false positive connections between regions. In our dataset for instance, specimens from Greenland identified as *Limnophyes anderseni* Sæther group with specimens from northern Norway that are identified as *L. vrangelensis* Makarchenko and Makarchenko (BIN BOLD:AAM6308). The two species are morphologically very similar, but are separated from each other by different setation on the thorax (Makarchenko and Makarchenko 2001; Sæther 1990). DNA barcode data suggest that a taxonomic revision of these species is needed. A similar situation exists for the species pairs *Diamesa bohemani* Goetghebuer and *D. zernyi* Edwards (BIN: BOLD:AAB5113), and *Cricotopus sylvestris* (Fabricius) and *C. glacialis* Edwards (BIN BOLD:AAA5299), as well as several other species in our dataset. These species pairs show minor morphological differences, but share the same BIN and probably should be regarded as synonyms.

The use of genetic clustering data, such as DNA barcode BINs or superBINs, eliminates discrepancies in taxonomic interpretation and provides a more objective basis for comparison, but will fail to detect cases where clearly different morphological species share the same BIN. One such case in our dataset concerns the species *Procladius dentus* Roback and a species with the interim name *Procladius* sp. ES04. The male genitalia of these two taxa are so different that they would be regarded as separate species by morphology (Fig. 5). Nevertheless, widely distributed records from northern Canada and northern Norway share the same BIN (BOLD:AAD7251). Examination of our dataset indicates, however, that BIN sharing by different true species is rarer than BIN discrepancy caused by differences in taxonomic interpretation.

There are also examples of taxonomic mismatch as a result of differences in opinion in our dataset. For instance, *Psectrocladius barbimanus* (Edwards) and *P. sokolovae* Zelentsov & Makarchenko recorded from Greenland share a BIN (BOLD:AAD4703) with *P. limbatellus* (Holmgren) from Spitsbergen. Since Spitsbergen is locus typicus for the older *P. limbatellus*, and a different BIN (BOLD:AAD0483) holds records of *P. barbimanus* from many other localities in the Arctic region, we regard the BIN discordance to be due to misidentification of the Greenland records. However, it does not need to be species-level misidentifications to create discordance: *Tanytarsus heliomesonyctios* Langton is likely widely distributed in the Arctic. It was originally described from Ellesmere Island (Langton 1999) and later recorded from Svalbard (Stur and Ekrem 2011). Since the species is parthenogenetic, at least in the most northern regions, and *Tanytarsus* females are impossible to identify to species level by morphology, identifications based on adults from Greenland are listed as *Tanytarsus* sp., resulting in a taxonomic mismatch with records in the same BIN (BOLD:AAC2863) from other regions.

In our study, the use of BIN and morphological species approaches produced similar results when accounting for differences in sampling effort (Table 1), and the overall biogeographic patterns were similar among the three different approaches for defining taxonomic units for analysis. Thus, any differences in taxonomic interpretation did not influence the biogeographic trends observed using Chao et al.'s (2005) index.

As expected, the observed endemism was higher at a lower clustering threshold, and the number of superBINs was closer to the number of morphological species in Greenland and the Palearctic, regions with the higher species-level identification coverage. There was a 28%–34% reduction in MOTU count when going from BINs (2.2% seed clustering threshold) to superBINs (~4.5% threshold), meaning that nearly one-third of BINs were collapsed into another (and therefore ~60% of BINs had a close nearest neighbour BIN). Previous studies have indicated that a 4%–5% threshold corresponds best with morphological species boundaries in Chironomidae (Ekrem et al. 2007; Lin et al. 2015). This is in contrast with findings in other insect groups such as butterflies and moths, bees, tachinid flies, and fungus gnats in which the BINs largely reflects what is considered morphological species (Ortiz et al. 2017; Pohjoismäki et al. 2016; Ratnasingham and Hebert 2013; J. Kjærandsen, personal communication, November 2017). However, for several insect groups, such as mayflies, caddisflies, stoneflies, and some beetles and heteropterans, multiple nearest neighbour BINs are found to belong to the same species (Hendrich et al. 2014; Morinière et al. 2017; Raupach et al. 2014).

Fig. 5. Male genitalia of *Procladius dentus* (left) and *Procladius* sp. 04ES (right). Photographed with differential interference contrast at 400× magnification using a Leica DM6000 compound microscope. Scale bars = 200 μm.

We can only speculate why there is an observed difference in mean MOTU threshold between species belonging to different insect groups, but one reason could be that large effective population sizes are less affected by selective sweeps, and that old species with large populations preserve intraspecific genetic diversity. We are not aware of studies that on a general level compare the effective population sizes between these insect groups, and current ecological, demographic, or biogeographic data for chironomids are insufficient to analyze this hypothesized mechanism in detail. However, a simple comparison of Lepidoptera and Chironomidae collected with a Malaise trap through a whole season in a suburban garden in Norway (see Aagaard et al. 2017 for methods) shows that population sizes in the collected chironomids generally were 2.6 times higher than in moths. The trap, located 300 m from the nearest water body, collected in total 3851 specimens of Chironomidae that could be assigned to 104 species, and 1329 Lepidoptera belonging to 94 species. Although a crude measure for effective population size, we suggest these numbers indicate that there might be such differences between these insect groups. This idea would be an avenue for follow-up research at a broader taxonomic and geographic scale.

Other mechanisms may also contribute to the different patterns of molecular diversity among insect taxa. For example, wide geographical distributions with restricted gene flow between populations can lead to a similar pattern (i.e., speciation in process), as can the presence of cryptic species (Blagoev et al. 2015; Morinière et al. 2017). We think it is unlikely, however, that the latter is so common in the Chironomidae that it is found in one third of all BINs, especially since many neighboring BINs are observed from the same region and even from the same locality. Another mechanism that may structure varied patterns of genetic diversity is differential extinction rates among taxa. More generalist insect functional groups, such as filter feeders, predators, and omnivores, may be more resistant to extinction compared to groups containing more specialists, including herbivores and parasitoids. The latter groups may have both high speciation and extinction rates, on balance leading on average to more species, each containing lower intraspecific genetic diversity. Indeed, prior research on parasitoid insects, including wasp (Hymenoptera) and fly (Diptera) taxa, has revealed a very high diversity of molecular clusters, at a low clustering threshold, which largely correspond to host associations and are thought to represent species (e.g., Smith et al. 2006, 2007, 2008). We suggest this hypothesis could be tested in a comparative phylogenetic context, including taxonomic groups varying in key biological and environmental characteristics. As effective population size (N_e) data are difficult to obtain for numerous species of insects, DNA sequence data could be used as at least an indicator of N_e, given that recent comparative studies have found that patterns of mitochondrial sequence diversity largely matched predictions derived from the nearly neutral theory of molecular evolution (e.g., Mitterboeck and Adamowicz 2013; Mitterboeck et al. 2017). These findings could be used as a foundation for testing whether there are consistent differences in molecular signatures of N_e among different insect lineages having varied biological attributes, such as feeding guild.

Our study showcases the utility of public DNA barcode databases for large-scale studies of biogeographic trends. The use of DNA barcodes reduces any bias that might occur by different taxonomic interpretation of taxa, but differences in sampling regimes might still influence comparisons of organismal communities. Future work on the biogeography of Arctic Chironomidae should include more countries and geographic regions to gain a true circumpolar perspective. Moreover, the analytical approach presented here, using a large dataset downloaded through BOLD's API tool, can be expanded to include other taxonomic groups, enabling broad, multi-taxon comparisons of large biogeographic patterns and elucidating the dispersal history of life.

Acknowledgements

We thank the laboratory team at CCDB for their fantastic efforts in generating barcodes from Arctic Chironomidae, Paul Hebert for initiating chironomid barcoding of Arctic midges in Churchill and supporting Canadian Arctic sequencing through a Natural Sciences and Engineering Research Council of Canada (NSERC) Polar Year Grant, and the Norwegian Biodiversity Information Centre for funding biosurveillance projects through the Norwegian Taxonomy Initiative. We also thank Darío Lijtmaer, Tomas Roslin, and two anonymous reviewers for useful input on our manuscript. DNA barcode data in this publication were in part generated in collaboration with the Norwegian Barcode of Life Network (NorBOL) funded by the Research Council of Norway and the Norwegian Biodiversity Information Centre.

References

Aagaard, K., Berggren, K., Hebert, P.D., Sones, J., McClenaghan, B., and Ekrem, T. 2017. Investigating suburban micromoth diversity using DNA barcoding of malaise trap samples. Urban Ecosyst. **20**(2): 353–361. doi:10.1007/s11252-016-0597-2.

Alsos, I.G., Ehrich, D., Eidesen, P.B., Solstad, H., Westergaard, K.B., Schönswetter, P., et al. 2015. Long-distance plant dispersal to North Atlantic islands: colonization routes and founder effect. AoB PLANTS, **7**: plv036. doi:10.1093/aobpla/plv036. PMID:25876627.

Alsos, I.G., Ehrich, D., Seidenkrantz, M-S, Bennike, O., Kirchhefer, A.J., and Geirsdottir, A. 2016. The role of sea ice for vascular plant dispersal in the Arctic. Biol. Lett. **12**: 20160264. doi:10.1098/rsbl.2016.0264.

Anderson, A.M., Stur, E., and Ekrem, T. 2013. Molecular and morphological methods reveal cryptic diversity and three new species of Nearctic *Micropsectra* (Diptera:Chironomidae). Freshw. Sci. **32**(3): 892–921. doi:10.1899/12-026.1.

Andujar, C., Arribas, P., Gray, C., Bruce, K., Woodward, G., Yu, D.W., et al. 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. Mol. Ecol. **27**: 146–166. doi:10.1111/mec.14410. PMID:29113023.

Armitage, P.D., Cranston, P.S., and Pinder, L.C.V. 1995. The Chironomidae: Biology and ecology of non-biting midges. Chapman & Hall, London.

Baranov, V., Lewandowski, J., and Krause, S. 2016. Bioturbation enhances the aerobic respiration of lake sediments in warming lakes. Biol. Lett. **12**: 20160448. doi:10.1098/rsbl.2016.0448. PMID:27484649.

Berg, M.B. 1995. Larval food and feeding behaviour. *In* The Chironomidae. The biology and ecology of non-biting midges. *Edited by* P.D. Armitage, P.S. Cranston, and L.C.V. Pinder. Chapman & Hall, London. pp. 136–168.

Berg, O.K., Finstad, A.G., Olsen, P.H., Arnekleiv, J.V., and Nilssen, K. 2010. Dwarfs and cannibals in the Arctic: production of Arctic char (*Salvelinus alpinus* (L.)) at two trophic levels. Hydrobiologia, **652**(1): 337–347. doi:10.1007/s10750-010-0366-9.

Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., et al. 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nat. Commun. **8**: 14087. doi:10.1038/ncomms14087. PMID:28098255.

Bivand, R., and Rundel, C. 2017. rgeos: Interface to Geometry Engine - Open Source (GEOS). Available from https://cran.r-project.org/package=rgeos.

Blagoev, G.A., deWaard, J.R., Ratnasingham, S., deWaard, S.L., Lu, L., Robertson, J., et al. 2015. Untangling taxonomy: a DNA barcode reference library for Canadian spiders. Mol. Ecol. Resour. **16**(1): 325–341. doi:10.1111/1755-0998.12444. PMID:26175299.

Böcher, J. 2015. The Greenland Entomofauna: zoogeography and history. *In* The Greenland Entomofauna: an identification manual of insects, spiders and their allies. *Edited by* J. Böcher, N.P. Kristensen, T. Pape, and L. Vilhelmsen. Brill, Leiden. pp. 21–36.

Böcher, J., Kristensen, N.P., Pape, T., and Vilhelmsen, L. 2015. The Greenland Entomofauna Brill, Leiden.

Brodersen, K.P., and Anderson, N.J. 2002. Distribution of chironomids (Diptera) in low arctic West Greenland lakes: trophic conditions, temperature and environmental reconstruction. Freshw. Biol. **47**(6): 1137–1157. doi:10.1046/j.1365-2427.2002.00831.x.

Brooks, S.J., Langdon, P.G., and Heiri, O. 2007. The identification and use of Palaearctic Chironomidae larvae in palaeoecology. Technical guide. Quat. Res. Assoc. Lond. No. 10. pp. 1–275.

CAFF. 2013. Arctic biodiversity assessment. Status and trends in Arctic biodiversity. CAFF, Akureyri.

Chao, A., Chazdon, R.L., Colwell, R.K., and Shen, T.J. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett. **8**(2): 148–159. doi:10.1111/j.1461-0248.2004.00707.x.

Chertoprud, M.V., Palatov, D.M., and Dimante-Deimantovica, I. 2017. Macrobenthic communities in water bodies and streams of Svalbard, Norway. J. Nat. Hist. **51**(47–48): 2809–2825. doi:10.1080/00222933.2017.1395092.

Coulson, S.J., Convey, P., Aakra, K., Aarvik, L., Ávila-Jiménez, M.L., Babenko, A., et al. 2014. The terrestrial and freshwater invertebrate biodiversity of the archipelagoes of the Barents Sea; Svalbard, Franz Josef Land and Novaya Zemlya. Soil Biol. Biochem. **68**: 440–470. doi:10.1016/j.soilbio.2013.10.006.

Danks, H.V. 1981. Arctic Arthropods. A review of systematics and ecology with particular reference to the North American fauna. Entomological Society of Canada, Ottawa.

Dillon, S., and Fjeldså, J. 2005. The implications of different species concepts for describing biodiversity patterns and assessing conservation needs for African birds. Ecography, **28**(5): 682–692. doi:10.1111/j.2005.0906-7590.04344.x.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792–1797. doi:10.1093/nar/gkh340. PMID:15034147.

Ekrem, T., Willassen, E., and Stur, E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol. Phylogenet. Evol. **43**: 530–542. doi:10.1016/j.ympev.2006.11.021. PMID:17208018.

Ekrem, T., Stur, E., and Hebert, P.D.N. 2010. Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding. Org. Divers. Evol. **10**(5): 397–408. doi:10.1007/s13127-010-0034-y.

Ferrington, L.C. 2008. Global diversity of non-biting midges (Chironomidae; Insecta-Diptera) in freshwater. Hydrobiologia, **595**: 447–455. doi:10.1007/s10750-007-9130-1.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. **3**(5): 294–299. PMID:7881515.

Glich, P.A. 1960. Collecting insects by airplane, with special reference to dispersal of the potato leafhopper. Agricult. Res. Serv. Tech. Bull. 1222. pp. 1–16.

Hebert, P., Cywinska, A., Ball, S., and deWaard, J. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. **270**: 313–321. doi:10.1098/rspb.2002.2218.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. U.S.A. **101**(41): 14812–14817. doi:10.1073/pnas.0406166101. PMID:15465915.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P.D.N., Hausmann, A., Köhler, F., et al. 2014. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. Mol. Ecol. Resour. **15**(4): 795–818. doi:10.1111/1755-0998.12354. PMID:25469559.

Hernández-Triana, L.M., Prosser, S.W., Rodríguez-Perez, M.A., Chaverri, L.G., Hebert, P.D.N., and Gregory, R.T. 2014. Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths. Mol. Ecol. Resour. **14**(3): 508–518. doi:10.1111/1755-0998.12208. PMID:24299419.

Hijmans, R.J. 2017. raster: Geographic data analysis and modeling. Available from https://CRAN.R-project.org/package=raster.

Hodkinson, I.D., Coulson, S.J., Webb, N.R., Block, W., Strathdee, A.T., Bale, J.S., et al. 1996. Temperature and the biomass of Flying Midges (Diptera: Chironomidae) in the High Arctic. Oikos, **75**(2): 241–248. doi:10.2307/3546247.

Hodkinson, I.D., Babenko, A., Behan-Pelletier, V., Böcher, J., Boxshall, G.R., Brodo, F., et al. 2013. Terrestrial and freshwater invertebrates. *In* Arctic biodiversity assessment. Status and trends in Arctic biodiversity. *Edited by* H. Meltofte. CAFF, Akureyri. pp. 247–275.

Holzapfel, E.P., and Perkins, B.D. 1969. Trapping of air-borne insects on ships in the Pacific, Part 7. Pacific Insects, **11**(2): 455–476.

Høye, T.T., and Sikes, D.S. 2013. Arctic entomology in the 21st century. Can. Entomol. **145**(2): 125–130. doi:10.4039/tce.2013.14.

Hubert, N., and Hanner, R. 2015. DNA Barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes, **3**(1): 44–58. doi:10.1515/dna-2015-0006.

Jakobsson, M., Andreassen, K., Bjarnadóttir, L.R., Dove, D., Dowdeswell, J.A., England, J.H., et al. 2014. Arctic Ocean glacial history. Quat. Sci. Rev. **92**(Suppl. C): 40–67. doi:10.1016/j.quascirev.2013.07.033.

Jones, O.R., Purvis, A., and Quicke, D.L.J. 2012. Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. Ecography, **35**: 333–340. doi:10.1111/j.1600-0587.2011.06956.x.

Langton, P.H. 1999. *Micropsectra silvesterae* n. sp. and *Tanytarsus heliomesonyctios* n. sp., (Diptera: Chironomidae), two parthenogenetic species from Ellesmere Island, Arctic Canada. J. Kans. Entomol. Soc. **71**(3): 208–215.

Lin, X., Stur, E., and Ekrem, T. 2015. Exploring genetic divergence in a species-rich insect genus using 2790 DNA barcodes. PLoS ONE, **10**(9): e0138993. doi:10.1371/journal.pone.0138993. PMID:26406595.

Lin, X.-L., Stur, E., and Ekrem, T. 2018. DNA barcodes and morphology reveal unrecognized species in Chironomidae (Diptera). Insect Syst. Evol. **49**(4): 329–398. doi:10.1163/1876312X-00002172.

Lindegaard, C. 2015. Chironomidae (Non-biting midges). *In* The Greenland Entomofauna: an identification manual of insects, spiders and their allies. *Edited by* J. Böcher, N.P. Kristensen, T. Pape, and L. Vilhelmsen. Brill, Leiden, the Netherlands. pp. 436–545.

Makarchenko, E.A., and Makarchenko, M.A. 2001. Chironomid fauna of Orthocladiinae subfamily (Diptera, Chironomidae) of the Vrangel Island. *In* Vladimir Ya. Levanidov's Biennal Memorial Meetings. *Edited by* E.A. Makarchenko and S.K. Kohlin. Dalnauka, Vladivostok. pp. 174–186.

Mitterboeck, T.F., and Adamowicz, S.J. 2013. Flight loss linked to faster molecular evolution in insects. Proc. R. Soc. B Biol. Sci. **280**: 20131128. doi:10.1098/rspb.2013.1128.

Mitterboeck, T.F., Liu, S., Adamowicz, S.J., Fu, J., Zhang, R., Song, W., et al. 2017. Positive and relaxed selection associated with flight evolution and loss in insect transcriptomes. GigaScience, **6**(10): 1–14. doi:10.1093/gigascience/gix073. PMID:29020740.

Morinière, J., Hendrich, L., Balke, M., Beermann Arne, J., König, T., Hess, M., et al. 2017. A DNA barcode library for Germany's mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). Mol. Ecol. Resour. **17**(6): 1293–1307. doi:10.1111/1755-0998.12683. PMID:28449274.

Nijman, V., and Meijaard, E. 2008. Zoogeography of primates in insular Southeast Asia: species-area relationships and the effects of taxonomy. Contrib. Zool. **77**(2): 117–126.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. 2018. vegan: Community ecology package. Available from https://CRAN.R-project.org/package=vegan.

Ortiz, A.S., Rubio, R.M., Guerrero, J.J., Garre, M.J., Serrano, J., Hebert, P.D.N., et al. 2017. Close congruence between Barcode Index Numbers (BINs) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. Biodiv. Data J. **5**: e19840. doi:10.3897/BDJ.5.e19840.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, **20**(2): 289–290. doi:10.1093/bioinformatics/btg412. PMID:14734327.

Pebesma, E.J., and Bivand, R.S. 2005. Classes and methods for spatial data in R. R News, **5**(2): 9–13.

Peterson, A.T. 2006. Taxonomy is important in conservation: a preliminary reassessment of Philippine species-level bird taxonomy. Bird Conserv. Int. **16**(2): 155–173. doi:10.1017/S0959270906000256.

Pohjoismäki, J.L.O., Kahanpää, J., and Mutanen, M. 2016. DNA barcodes for the Northern European Tachinid Flies (Diptera: Tachinidae). PLoS ONE, **11**(11): e0164933. doi:10.1371/journal.pone.0164933. PMID:27814365.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratnasingham, S., and Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol. Ecol. Notes, **7**(3): 355–364. doi:10.1111/j.1471-8286.2007.01678.x. PMID:18784790.

Ratnasingham, S., and Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS ONE, **8**(7): e66213. doi:10.1371/journal.pone.0066213. PMID:23861743.

Raupach, M.J., Hendrich, L., Küchler, S.M., Deister, F., Morinière, J., and Gossner, M.M. 2014. Building-up of a DNA barcode library for True Bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. PLoS ONE, **9**(9): e106940. doi:10.1371/journal.pone.0106940. PMID:25203616.

Sæther, O.A. 1990. A review of the genus *Limnophyes* Eaton from the Holarctic and Afrotropical regions (Diptera: Chironomidae, Orthocladiinae). Entomol. Scand. Suppl. **35**: 1–135.

Shang, J., Zhang, L., Shi, C., and Fan, C. 2013. Influence of Chironomid larvae on oxygen and nitrogen fluxes across the sediment–water interface (Lake Taihu, China). J. Environ. Sci. **25**(5): 978–985. doi:10.1016/S1001-0742(12)60116-8.

Silva, F.L.d., Ekrem, T., and Fonseca-Gessner, A.A. 2013. DNA barcodes for species delimitation in Chironomidae (Diptera): a case study on the genus *Labrundinia*. Can. Entomol. **145**(6): 589–602. doi:10.4039/tce.2013.44.

Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., and Hebert, P.D.N. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proc. Natl. Acad. Sci. U.S.A. **103**(10): 3657–3662. doi:10.1073/pnas.0511318103. PMID:16505365.

Smith, M.A., Wood, D.M., Janzen, D.H., Hallwachs, W., and Hebert, P.D.N. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl. Acad. Sci. U.S.A. **104**(12): 4967–4972. doi:10.1073/pnas.0700050104. PMID:17360352.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., et al. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. U.S.A. **105**(34): 12359–12364. doi:10.1073/pnas.0805319105. PMID:18716001.

Stur, E., and Ekrem, T. 2011. Exploring unknown life stages of Arctic Tanytarsini (Diptera: Chironomidae) with DNA barcoding. Zootaxa, **2743**: 27–39.

Stur, E., and Ekrem, T. 2015. A review of Norwegian *Gymnometriocnemus* (Diptera, Chironomidae) including the description of two new species and a new name for *Gymnometriocnemus volitans* (Goetghebuer) sensu Brundin. ZooKeys, **508**: 127–142. doi:10.3897/zookeys.508.9874.

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**(3): 512–526. doi:10.1093/oxfordjournals.molbev.a040023. PMID:8336541.

Vinarski, M.V., and Kramarenko, S.S. 2015. How does the discrepancies among taxonomists affect macroecological patterns? A case study of freshwater snails of Western Siberia. Biodivers. Conserv. **24**: 2079. doi:10.1007/s10531-015-0934-4.

Wickham, H., Hester, J., and Francois, R. 2017. readr: Read Rectangular Text Data. Available from https://CRAN.R-project.org/package=readr.

Wirta, H.K., Vesterinen, E.J., Hambäck, P.A., Weingartner, E., Rasmussen, C., Reneerkens, J., et al. 2015. Exposing the structure of an Arctic food web. Ecol. Evol. **5**(17): 3842–3856. doi:10.1002/ece3.1647. PMID:26380710.

Wirta, H., Várkonyi, G., Rasmussen, C., Kaartinen, R., Schmidt, N.M., Hebert, P.D.N., et al. 2016. Establishing a community-wide DNA barcode library as a new tool for arctic research. Mol. Ecol. Resour. **16**(3): 809–822. doi:10.1111/1755-0998.12489. PMID:26602739.

Wright, E.S., Yilmaz, L.S., and Noguera, D.R. 2012. DECIPHER, a search-based approach to Chimera identification for 16S rRNA sequences. Appl. Environ. Microbiol. **78**(3): 717–725. doi:10.1128/AEM.06516-11. PMID:22101057.

Zahiri, R., Lafontaine, J.D., Schmidt, B.C., deWaard, J.R., Zakharov, E.V., and Hebert, P.D.N. 2017. Probing planetary biodiversity with DNA barcodes: The Noctuoidea of North America. PLoS ONE, **12**(6): e0178548. doi:10.1371/journal.pone.0178548. PMID:28570635.