Novel Pipeline for Large-Scale Comparative Population Genetics

Majoros, S.E.[1], Adamowicz, S.J.[1], Cottenie, K.[1]

[1]University of Guelph, Guelph, ON, Canada

Abstract

This study determined population genetic structure measures, compared these measures across species with different biological traits; and created efficient, reproducible, reusable programming modules that are publicly available for future research. Cytochrome C Oxidase subunit I gene sequences from Diptera (true fly) species from Greenland and Canada were used as a case study and proof of concept. We hypothesized that population genetic structure measures will be influenced by the biological traits of organisms. Data were pulled from public databases, as well as taxon-specific literature. The R pipeline includes fifteen modules that can be adapted and applied to a diverse set of animal groups, geographic regions, genes, and traits. Habitat, larval diet, geographical distance, latitude, and longitude were all significantly related to population genetic structure in Diptera. Overall, this study has created efficient, reusable bioinformatics modules, as well as provided insight into the factors affecting population genetic structure in Northern fly communities.

1 Introduction

As scientists continue to ask complex and urgent questions about our world and environments, and deal with increasingly large amounts of data, there is a demand for new methods and computational developments to perform scientific analyses (Reiter et al., 2021, Shade & Teal, 2015, Smith, 2018). Bioinformatics and computational science are becoming increasingly important to scientific research (Smith, 2018). There is also a need to ensure reproducibility, which becomes more difficult as scientific analyses increase in complexity (Reiter et al., 2021). Scientific workflows and pipelines can make this more manageable (Liew et al., 2017). Pipelines contain all the steps of an analysis, allowing for the process to be understood and repeated by others. Pipelines improve the efficiency of research, allow for repeatability of methods, and provide a foundation upon which to build future research. With more scientists needing to use some element of bioinformatics and programming, there is a need for user-friendly and accessible pipelines (Smith, 2018).

A programming pipeline includes all steps of the analysis, from data filtering and formatting to specific analyses required by the research, to export of statistical results and data visualizations. These pipelines should be reusable and made publicly available

for use by other researchers (Reiter et al., 2021). They should also be accurate, efficient (able to be run in a timely manner without unnecessary steps), well documented, and usable by other researchers.

There are some best practices to consider when creating a programming pipeline, outlined in Reiter et al. (2021) and Shade & Teal (2015) (Table 1). First, it is recommended to visualize the steps in the pipeline, for instance in a flowchart. When beginning the pipeline, thought should be put into the choice of tools and packages, and the choice of parameters for each function and analysis, to determine which best suit the analysis and research goals for the pipeline. The pipeline should include steps to optimize these parameters, and a sensitivity analysis could also be included to test the effect of different parameters on the results. When testing the pipeline, a good dataset should also be chosen as an example, and all the information regarding how this dataset was acquired should be recorded. Quality control steps should also be included in order to ensure the pipeline is working as intended. Checks should also be included to ensure the data have the correct properties. The degree of filtering and control should be decided based on the needs of the analysis. All data, tool, and parameter choices made within the analysis and code creation should be included in the commenting and documentation.

While editing the code, it is also a good idea to keep track of the different versions and edits; this can be done easily using Git, GitHub (Github, 2022), or other websites. It is important to back up any changes made to the code often, and to have a record of these changes and the reasons for them. By keeping a transparent record of the changes, other potential users can understand the process, rationale for decisions made, and potentially catch any overlooked mistakes.

The pipeline should be able to run without errors, except for helpful errors to detect data formatting issues, and checks should be included in the script to ensure it is running as intended and the results it is providing are accurate. The pipeline should be tested using a small sample of the data first, before scaling up. Another good way to test the code performance is to create figures at checkpoints, as these can act as another check for code performance and accuracy. Note should be taken of any warnings generated throughout the analysis, and users must determine how these should be addressed. It is also recommended to open the pipeline up for review and have other scientists test the code. Finally, the end results should make biological sense for the project.

The pipeline should also be reproducible and provide the same results after each run (Akalin, 2018, Shade & Teal, 2015). To ensure this, reproducibility checkpoints should be used. At each checkpoint, the programming environment should be cleared, and the script should be rerun to ensure the results are the same (Shade & Teal, 2015). By setting a random seed in the case of analyses involving randomization, developers and users can check for reproducibility of the results even if some results may naturally differ due to the nature of the analysis (Akalin, 2018). The pipeline should not only be

run on the developer's own computer, but also on other computers, to ensure they can get the same results (Akalin, 2018, Shade & Teal, 2015).

The pipeline should be accessible (Akalin, 2018, Reiter et al., 2021). As mentioned earlier, all packages and tools should be documented, as well as any versions of the packages and software (Reiter et al., 2021). To make the pipeline reusable and analysis reproducible, all relevant information needs to be made available to future users (Akalin, 2018), including the files and data used, the choices made at each step of the analysis, and the fully commented code (Reiter et al, 2021). The code itself should be organized, and labelling of variables should be consistent and descriptive (Reiter et al., 2021). Any additional information should be included in a README file, or within the commenting itself (Reiter et al., 2021). Pipelines, with associated data and documentation, need to be accessible and findable and can be shared using GitHub or similar websites (Akalin, 2018).

Programming modules have been created for a variety of topics, including in the biological sciences, and many are publicly available on GitHub, such as Dalapicolla, (2021), which investigates genetic structure and diversity. However, there are still many important questions that are unanswered, and more workflows and pipelines are needed. With increasing data availability and the continuous addition of new data to databases, there is value in making the analytical process easier and more automated. In this study we selected and combined a series of modules and functions into a modular pipeline in the R programming language that can be used to answer biological questions regarding population genetic structure and realized dispersal. Understanding dispersal patterns can help us understand how communities and ecosystems function and how these communities will react to disturbance events such as climate change or habitat fragmentation (Fobert et al., 2019, Stevens et al., 2010).

To develop programming modules that address biological questions, a thorough understanding of the biological background is needed. When developing a pipeline that focuses on realized dispersal and population genetic structure, it is important to understand the processes and the factors that influence them. Dispersal is the movement of one organism from one location to another and is important for many aspects of community ecology and evolution (Fobert et al., 2019, Stevens et al., 2010). Dispersal rates, however, are difficult to calculate. Various methods are used to study dispersal, but the differences between these methods often lead to inconsistency in the data and results between methods (Stevens et al., 2010). Dispersal can be determined by direct methods, such as mark and recapture studies and in lab experiments (Stevens et al., 2010, Steyn et al. 2016). Indirect methods include using the distribution of genetic diversity among populations to infer inter-population genetic differences and gene flow (Stevens et al., 2010). These studies can also be time consuming and technologically difficult, and factors that contribute to dispersal rates (such as physiological and behavioural tradeoffs, life history traits, environmental conditions, and other factors influencing gene flow) are often unaccounted for using one method (Downes & Lancaster, 2018, Stevens et al., 2010).

Dispersal as a general concept is the movement of organisms, but realized dispersal occurs when organisms are able disperse and then reproduce and establish a population in a new location or leave descendants in an already established population (Fig.1). Indicators of relative realized dispersal rates, while still likely influenced by many factors, are more easily determined and may be important for determining the potential of species for successful colonization. Population genetic structure measures can be used as indicators of realized dispersal. Populations experience both genetic drift, which contributes to the genetic differentiation between populations, and gene flow, which contributes to the population connectivity and allows the exchange of genetic information. By calculating population genetic structure measures, we can investigate the balance between these two factors.

One potential set of factors that may influence realized dispersal is biological traits. Relationships between dispersal and traits have been found in prior studies and diverse taxa. For example, Fobert et al. (2019) found that the larval phenotypes of marine fish species in Australia influenced their vertical placement in the water body and their dispersal ability. Steyn et al. (2016) found similar results in different phenotypes of a fruit fly species. The relationship between dispersal and biological traits is also shown in salamander species with differing traits, where sympatric sexual species were better dispersers than unisexual species, likely due to higher locomotive endurance (Denton et al., 2017). As a final example, successful invasion was found to be related to traits, such as dispersal ability, diet, reproductive rate, and environmental tolerance, in aquatic insect species by Carbonell et al. (2016).

Other studies suggest that traits are not the most important predictor of successful dispersal. Investigating primary succession and colonization in plant species in Germany, Kirmer et al. (2008) determined that the abundance of nearby species was more important than the traits of the colonizing species. Similarly, Downes & Lancaster (2018) found that the traits of Australian insects associated with dispersal ability were not necessarily good predictors of colonization and that environmental conditions were most important. Oster et al. (2009) also found that colonization was not related to functional traits of plant species in Sweden.

While many studies have investigated dispersal, most only considered differences between realized dispersal and colonization within a species or among a few closely related species or at a single location (Carbonell et al., 2016, Denton et al., 2017, Downes & Lancaster, 2018, Fobert et al., 2019, Oster et al., 2009, Steyn et al., 2016). There are still many taxa and locations that are understudied. However, the increased availability of georeferenced DNA sequence data provided an opportunity to advance knowledge across multiple taxa and geographic regions by semi-automating large-scale analysis. This study developed a novel way to study indicators of realized dispersal and population genetic structure across taxa and geographic regions. It thus provides the opportunity to discover patterns for understudied taxa and geographic regions, and lays the groundwork upon which to build further population genetic structure studies.

In this study, we created a modular pipeline by selecting and combining a series of modules and functions and using the functionality available in R. This pipeline and set of modules infer relative realized dispersal rates, using population genetic measures calculated from DNA sequences as an indicator of realized dispersal, using molecular data. The case study used to demonstrate and test the pipeline makes use of publicly available DNA sequence data and uses cytochrome c oxidase subunit 1 (CO1) sequences.

Cytochrome c oxidase subunit 1 (CO1) is a mitochondrial gene commonly used for DNA barcoding in animals and is good for identification of specimens to the species level and for species delineation and estimation of species richness using molecular methods (Hebert et al. 2003, Boyle & Adamowicz 2015). DNA barcoding involves using one or more standardized gene regions to identify specimens to species and to discover new species (Hebert et al. 2003, 2016). CO1 is expected to be more conservative than some other markers, such as microsatellites, which are commonly used for analysis below the species level in detecting differences among individuals within species (Liu et al., 2020). However, there is evidence that there is sufficient variability in CO1 to detect major population divisions and phylogeographic history (e.g., April et al., 2012, Bernatchez & Wilson, 1998, Keogh et al., 2021, Witt & Hebert, 2000). CO1 has been used to answer questions at a variety of temporal scales, including drawing conclusions about demographic history, divergence times, and events occurring millions of years ago (Attiná et al., 2021, Choi et al., 2020, Havel et al., 2000, Jossart et al., 2017, Reed et al., 2006, Stark et al., 2021, Talbot et al., 2016). Various studies have had success using CO1 to determine population genetic structure (Abuelmaali et al., 2021, Attiná et al., 2021, Choi et al., 2020, Froufe et al., 2014, Havel et al., 2000, Jossart et al., 2017, Liu et al., 2020, Meriam et al., 2015, Park et al., 2019, Pickett & David, 2018, Reed et al., 2006, Shum & Palumbi, 2021, Stark et al., 2021, Talbot et al., 2016, Troast et al., 2016, Xu et al., 2019). This success was primarily due to CO1's high rate of evolution and mutation, as well as faster coalescence rate compared to nuclear markers, which makes it useful for determining genetic differences between isolated populations (Froufe et al., 2014, Havel et al., 2000, Meriam et al., 2015, Stark et al., 2021, Xu et al., 2019). CO1 has been shown to provide accurate species identification and perform well for differentiating species, though this accuracy may be reduced as the geographic distance increases (Attiná et al., 2021).

CO1 is frequently concordant with other types of genetic data for detecting genetic variability and population genetic structure. Meriam et al. (2015) suggested that CO1 performs better than 12S and 16S ribosomal RNA (rRNA), due to its greater evolutionary rate. Choi et al. (2020) and Stark et al. (2021) also found similar results between CO1 and 16S rRNA when investigating genetic diversity, phylogeographic history, and phylogenetic relationships, and Havel et al. (2000) found similar results between CO1 and 12S rRNA. CO1 may be conservative for detecting rates of contemporary gene flow, but is expected to contain information content, due to frequent concordance with microsatellite data in the literature (Jossart et al., 2017, Talbot et al., 2016). Talbot et al. (2016) used both microsatellites and CO1 to study genetic diversity

within a bat ectoparasite species, and while CO1 exhibited less genetic variation among individuals with different host associations than microsatellites, they determined this was not due to any lack of ability in CO1 to detect differentiation, as among other populations, microsatellites showed less variation than CO1. Jossart et al. (2017) found comparable results between CO1 and the microsatellites when investigating intraspecific population genetic structure.

We wrote efficient, scalable, reproducible, and reusable modules, primarily using the R language (R Core Team, 2022). The R programming language provides a large variety of functions and packages for bioinformatics, ecological analysis, statistics, and visualization that allow analyses to be performed accurately and efficiently, with the source code available with open access. Due to R's functionality, accessibility, the number of options it provides, and large community of contributors and users, it is a good environment and language in which to create analytical modules. These modules allow for accurate, reproducible, and efficient calculation of relative realized dispersal rates using population genetic structure (i.e., degree of genetic isolation between subpopulations) and geographic range as indicators of successful dispersal. The pipeline will also determine how the biological traits of species influence the indicators of population genetic structure. Through this study, we produced a publicly available series of modules and commented code that aligns with best practices (as outlined in Akalin, 2018, Reiter et al., 2021, and Shade & Teal, 2015) and will be available for other researchers to use.

To test the efficiency and accuracy of the modules, we ran an analysis on Diptera species from two relatively well-sampled Northern regions, Greenland and Canada (Hebert et al., 2016, Wirta et al., 2015). These regions were chosen due to the Northern and Arctic environments they contain. The climate is shifting, and Arctic areas in particular are undergoing a variety of changes, including disappearing sea ice, melting permafrost, and increasing temperatures (Descamps et al., 2017, Lagger et al., 2017, Tape et al., 2018, Walseng et al., 2018, Węsławski et al., 2011). These changes affect the organisms present in the Arctic and can impact their survival. By understanding the current population genetic structure patterns present in Arctic populations, we can potentially better understand how they will be impacted by the changing climate. Greenland provides an example of the minimal data needed to run the modules and as a check of module functionality. The Canada dataset acts as a case study and shows the module performance on a larger dataset. Diptera was chosen due to the abundance of data available on databases such as the Barcode of Life Data Systems (BOLD; Ratnasingham and Hebert, 2007) and Global Biodiversity Information Facility (GBIF; GBIF.org, 2020), and large barcoding efforts by programs such as the Global Malaise Trap Program (deWaard et al., 2018, Geiger et al., 2016, Global Malaise Program, n.d.). Diptera is also a large, diverse order, possessing a variety of biological traits, including different diets, feeding modes, and habitat requirements (Marshall, 2012).

The modules developed in this study allowed me to discover population genetic structure patterns for Diptera and will allow for future research to determine realized

dispersal patterns for a wide range of taxa and environments. In this study we determined relative realized dispersal rates for Diptera, with population genetic structure measures as an indicator, and compared these rates across species with differing traits. We hypothesized that indicators of realized dispersal rates will be influenced by the traits of organisms due to the importance of traits to the dispersal and colonization process. We predicted that indicators of realized dispersal rates will be similar in species that share traits (Fig.2). While there is a large variety of traits that can be investigated, a select few were chosen for this example. We investigated habitat, adult diet, and larval diet. We expected a generalist diet and terrestrial habitat to be associated with higher successful dispersal rates. Having flexibility of food choice may allow individuals of a given species to survive in a different environment more readily, and increased connectivity between terrestrial environments compared to aquatic may result in more dispersal and geneflow. In these cases, we would expect these populations to have lower genetic differentiation, due to high dispersal and successful colonization.

2 Methods

The pipeline and modules can be found in the repository Population_Genetic_Structure_Pipeline on Github, S-Majoros at https://github.com/S-Majoros/Population_Genetic_Structure. This a public repository containing the R script and necessary data files. The modules included in the pipeline and analysis are outlined below. A flowchart showing the steps of the pipeline can be found in Figure 3 and a brief description of each module can be found in table 2.

2.1 Loading Packages and Setting up Variables (Module 1)

Module 1 acts as a precursor for all the other modules and contains all the needed packages and parameters to run each module. All packages and parameters are separated by module and can be run together or individually. Important parameters are outlined in this module and made alterable by the user. When a variable or parameter appears in italics throughout this method section, it can be adjusted by the user in this module. This is to draw attention to these important parameters and allow easy adjustment, but other parameters and aspects of the script can also be edited as needed within each module.

2.2 Extracting Data from Databases (Module 2)

The main dataset needed to run the full population genetic structure analysis has to include nucleotide sequences, geographic coordinates, and species identification. These data were all provided by the Barcode of Life Data Systems (BOLD; Ratnasingham and Hebert, 2007) database. Taxon-specific trait data are also required for Modules 11-13, and 15, and were obtained from the literature and loaded in during a later module. Additional datasets, such as the Global Biodiversity Information Facility

(GBIF; GBIF.org, 2020) dataset, can be used to provide additional geographic distribution information for the species.

Geographic regions where there has been an effort to gather DNA barcode sequence data were emphasized, as such data are often associated with GPS information and other metadata (Ratnasingham and Hebert 2007).

Data were retrieved from BOLD. Data for Diptera from Canada and Greenland were downloaded on June 24th, 2021, using geographic keywords. As of Dec. 15th, 2021, BOLD contained 3,118,612 records of Diptera representing 27,440 species and 132,855 Barcode Index Numbers (BINs; Ratnasingham and Hebert, 2013). BINs are operational taxonomic units (OTU) used by BOLD that are clusters of barcode sequences similar to species (Ratnasingham & Hebert, 2013). Data were pulled directly into the R environment using BOLD's application programming interface (API). Data were also retrieved for Diptera from Greenland from GBIF (GBIF.org (June 24th, 2021) GBIF Occurrence Download (https://doi.org/10.15468/dl.mk52hp)). These data included Diptera occurrences from Canada and Greenland, and only occurrences with geographic coordinates were retained.

2.3 Filtering the Data (Module 3)

The BOLD data include a variety of attributes and information, including the process and record IDs specific to BOLD, BIN numbers, nucleotide sequences, catalog, storing and collection information, taxonomic information, and geographic coordinates. Some records may contain further information about each specimen, including sex, lifestyle, reproduction, and habitat, but these are not always provided. For this analysis, the dataset was reduced to only the information needed for the modules, which included the process ID, BIN, family, genus, and species name, nucleotide sequence and geographic coordinates. The BOLD records were filtered to exclude those without a BIN, CO1 nucleotide sequence of at least 500 base pairs (bp), species identification, or coordinates. BINs were used to represent species. Several studies have found strong correlation between BINs and species in Diptera (Bartolini et al., 2020, Hernandez-Triana et al., 2017, Yu et al., 2020). For example, when Bartolini et al. (2020), sorted morphological *Anastrepha* (Diptera: Tephritidae) species into BINs, they created 14 BINs for the 16 species represented. Hernandez-Triana et al. (2017) created 29 BINs for 28 Simuliidae species, and Yu et al. (2020) sorted 5 Chironomidae species into seven BINs. BINs have also been shown to correspond to species in other arthropod groups such as Coleoptera, where BINs matched with species 90% of the time (Pentinsaari et al. 2014, 2016). Sequences with a high gap/N content (areas where there are deleted bases or missing information) were also removed. A threshold of 1% gap/N content was chosen because species often differ by more than 2% divergence (Ratnasingham & Hebert, 2013). Filtering out records with > 1% N and gap content is likely to yield a high-quality data set, given typical patterns of variability in CO1 in animals. The sequences were also checked for stop codons and non-biological insertions or deletions (indels) using the R package coil version 1.2.3 (Nugent et al.,

2020) and removed if these features were detected. The GBIF data were also filtered to remove occurrences without a species-level identification.

2.4 Organizing Data by Clusters (Module 4)

As mentioned above, BINs can be used to represent species in this study, but the module will also provide the option to use a different *clustering method* or *genetic distance threshold*. In our example, we used a 4% single linkage clustering threshold to group closely related sequences together into clusters, while excluding more distantly related sequences. Lin et al. (2015) determined that a threshold of 4-5% was suitable for clustering within the Diptera family Chironomidae. Similarly, Baloğlu et al. (2018) found stable results in Chironomidae across 3-5% clustering thresholds. Based on these results, 4% was deemed suitable for this study, but the threshold can be altered depending on the taxa being investigated. This option is important because, while BINs often correspond to species, the degree of concordance varies across taxa. By providing an alternative, users can more accurately assign specimens to species-like groupings across a variety of taxa or could use species names in the case of taxa with well-developed taxonomy and extensive taxonomic annotations of sequences. These results were then compared to those generated using BINs to understand how the two clustering approaches influence the results. To cluster the sequences, they were first aligned using the R package Decipher version 2.20.0 (Wright 2016), using the default parameters, and a *reference sequence*, to which the sequences were aligned and trimmed. The reference sequence was a Diptera (*Patelloa xanthura)* sequence retrieved from BOLD (Process ID: ACGAZ1590-12, BIN:AAA1222). Users can upload their own reference sequence or choose one of those provided for some commonly studied taxa. The BIN was chosen because it met the following selection criteria: from the order Diptera, contained at least 10 CO1-5P sequences, had at least one specimen photograph that matched the species-level taxonomy, and did not have taxonomic conflicts at the family level or above. The reference sequence within the selected BIN was chosen because it is 658 base pairs long, has 2 high-quality trace file chromatograms, and no missing nucleotides or stop codons.

A distance matrix was created using the model p-distance. This model was chosen as p-distance is used for the BIN algorithm and keeping the distance model consistent allows for easier comparison to the BIN-based analysis. The sequences are then clustered using any method. We chose the single linkage clustering method, since Refined single linkage is used in Ratnasingham & Hebert (2013) in the formation of BINs. To deal with computer memory constraints and reduce computing time, an option is included to randomly sample down the dataset to a maximum size for each BIN. This option is important as it allows the modules and script to be run by different researchers and makes it accessible to those who may not be able to access high-performance computing resources. If downsampling, users may consider repeating the randomization multiple times and comparing the results. This downsampling step was only used for the Canada data. A size of 100 *sequences per BIN* was chosen as the default setting to reduce strain on computer memory while still having a large sample size for analysis.

Throughout this step, the data are kept together if the dataset is small, but an option is provided to separate the data by genus until after the clustering step, as looping through the genera individually during the clustering step reduces computational demands for large datasets.

2.5 Grouping by Geographic Region (Module 5)

Populations were geographically defined by grouping the records into polygons using the package dggridR version 2.0.8 (Barnes & Sahr, 2017) and the geographic coordinates provided for each record. *Polygon length* is a variable that can be set at the beginning of the analysis. The data were then filtered further to include BINs only if they meet the minimum sample size requirement of at least 20 sequences total and to be present in at least two polygons. Each BIN within each polygon was also required to have at least 10 sequences in the polygon. These filtering steps were included in order to ensure that each BIN included a suitable amount of data for population genetic analysis. Polygons were filtered out if they contained fewer than 10 BINs that met the above sample size criterion. Accumulation curves and abundance and incidence-based Chao richness estimators (Chao, 1984) were calculated for each polygon. For our example, polygons that have at least 50% of their *expected BIN richness sampled* were included in the analysis. While having a higher richness is preferable, this may not be possible depending on the geographic regions studied. This filtering step is included in order to ensure the polygons being included in the analysis are reasonably well sampled, to yield a more general, rather than BIN-specific, analysis of population genetic structure in Diptera.

2.6 Multiple Sequence Alignment (Module 6)

Once the dataset was filtered, the sequences were aligned again within each BIN. The sequences were trimmed using the *reference sequence* and then aligned using the same parameters outlined in the clustering step. If BINs were used, this was the first multiple sequence alignment for the dataset. If clusters were used, this was the second, but this time only the clusters included in the final analysis were included.

2.7 Population Genetic Structure Measures (Modules 7 - 9)

Measures of relative realized dispersal (population connectivity), pairwise Nei's F_{ST} (Nei, 1972), Hedrick's G_{ST} (Hedrick, 2005), and Jost's D (Jost, 2008), were calculated using the package mmod version 1.3.3 (Winter, 2012) (Module 7). These measures quantify how genetic diversity is distributed among sub-populations and determine levels of genetic differentiation (Fig.4). These measures differ slightly in the way they are calculated (Verity & Nichols, 2015). Values of F_{ST} can range between 0 and a value determined by the heterozygosity in the data, while G_{ST}, and Jost's D use a standardized range of 0 – 1 (Verity & Nichols 2015), with zero indicating identical allele frequencies among populations and 1 indicating no sharing of alleles. The original F_{ST} was for two-allele data, but Nei's F_{ST} is adjusted for multi-allelic data. The measures are

calculated for each polygon pair for each BIN. For larger datasets, an option to calculate the global F_{ST}, G_{ST} and Jost's D is included. This provides one measure for each BIN. This can also be calculated for small datasets, though if fewer polygons and BINs are included, it is more feasible to examine the results between each polygon pair. An analysis of molecular variance (AMOVA) was then used (using the package poppr version 2.9.3 (Kamvar et al., 2014)) to take into account the amount of genetic difference among alleles and quantify how the genetic variability is partitioned among geographically hierarchical populations (Module 8). The default parameters of this function were used, and raw pairwise distances are calculated from the data. Similar to the population genetic structure measures, the AMOVAs give us an idea of the genetic differentiation between populations. Mismatch distributions were also measured as an additional measure of population expansion, using the package pegas version 1.1 (Paradis, 2010), with our example study targeting northern regions (Module 9). Mismatch distributions are used to investigate patterns in the distribution of pairwise genetic distances (as in Shum & Palumbi, 2021) (Rogers & Harpending, 1992). In order to determine whether the distribution is uni or multimodal, a dip test was performed. The null hypothesis is a unimodal distribution. If the distribution is unimodal, this suggests that the population is still expanding.

2.8 Phylogenetic Tree (Module 10)

The phylogenetic tree module provides the option to generate a maximum likelihood tree from the user's data, rather than uploading a tree from the literature. A well-supported, multi-gene tree may not be available from the literature for all study taxa, so the option to generate one de novo allows for the phylogenetic generalized least squares (PGLS) analysis to be performed for taxa for which the evolutionary relationships are not yet well known. The maximum likelihood tree was constructed using one representative sequence from each BIN or cluster, the centroid. A centroid sequence is defined as the sequence with the minimum average distance to all others in its BIN (or species cluster) (Orton et al., 2019). Alignments for each BIN or cluster were performed using the same process as outlined above, and distances were found using the default parameters of the function dist.dna from the package ape version 5.6.2 (Paradis et al., 2004). The centroid sequences (one sequence for each BIN or cluster) were then aligned to the *reference sequence and the outgroups*. For this case study, three sequences from Trichoptera were chosen as outgroups. Trichoptera was chosen as it is closely related to Diptera, but definitively outside the ingroup (Misof et al., 2014). Before creating the tree, the best-fit model of nucleotide evolution was estimated using the *Bayesian Information Criterion (BIC) score*. The tree was then generated using the function optim.pml from phangorn version 2.7.1 (Schliep 2011) and the following parameters: intervals of *discrete gamma distribution (the k-value)* was set to 4, and optNni, optGamma and optInv were set to true. The proportion of invariant sites was determined based on the model and a neighbour-joining (NJ) tree (Saitou & Nei, 1987) (created using NJ from phangorn version 2.7.1 (Schliep 2011)) was used as the guide tree.

2.9 Relationship between Traits and Population Genetic Structure (Modules 11 – 13)

To determine the relationship between traits and population genetic structure, three commonly used methods were used. BINs or clusters were grouped by traits, and the population measures (F_{ST}, G_{ST}, Jost's D) were compared across groups with different biological traits using an analysis of variance (ANOVA) (Module 11). A phylogenetic generalized least squares (PGLS) analysis was also performed to take into account phylogenetic relationships (Module 12). For larger datasets, such as our Canada dataset, an option to run a multiple regression was included (Module 13). If other analyses are preferred by the researcher, these modules can be replaced, and different methods can be used.

For our case study, the traits investigated were habitat, adult diet, and larval diet. Other researchers can substitute in other traits. Traits were assigned to each species/BIN using the literature, and a full list of sources is provided in appendix A. Traits were assigned at the lowest taxonomic level possible, and for BINs that contained multiple species names, the most common species name within the BIN was used. For habitat, we defined terrestrial as taxa that live primarily in land habitats, aquatic as taxa that live primarily in water bodies or associated habitats, and semi-aquatic as species that require both aquatic and terrestrial habitats. For diet, we defined predaceous taxa as those who prey on insects or other animals, phytophagous taxa as those who feed on plant material, fungivores as those who feed on fungi, saprophagous taxa as those who feed on decaying organic matter, omnivores as those who consumed material of both plant and animal origin, and parasites as those who live and feed in or on an organism of another species. If a species' diet consisted of a combination of the above, this was included as its own category. As some flies are non-feeding in the adult stage, this was also recorded as its own category.

For the PGLS, the required tree can be specified by the user or created using the phylogenetic tree step. For the case study, the topology of the phylogenetic tree was created manually in the program Mesquite (Maddison & Maddison 2019). This tree was based on the family-level Diptera tree created by Wiegmann et al. (2011) using 14 nuclear loci and complete mitochondrial genomes, as well as morphology. Haseyama et al. (2015)'s phylogenetic hypothesis based on 4 protein-coding genes, was used as an additional lower-level taxonomic tree for the relationships within the family Muscidae.

For the Canada case study, a multiple regression was performed using the lm function from the stats package version 4.1.3 (R Core Team, 2022). The global F_{ST}, G_{ST}, and Jost's D values were the response variables, and the predictor variables used for each BIN were habitat, adult and larval diets, the number of polygons each BIN was found in, the centre coordinates of each BIN range (found by creating a polygon and finding the centre coordinates using the package sp version 1.4.6 (Pebesma, 2018)),

the distance between the centre points of the two most distant polygons occupied by a given BIN or cluster, and the number of records.

2.10 Sensitivity Analysis (Module 14)

While there are too many parameters to determine the effects of each parameter on the analysis within the scope of this study, a key variable, polygon size, was chosen to illustrate the concept of a sensitivity analysis using the Diptera case study data. The size of the polygon/region influences the number of BINs and records within each region, so it is important to determine the effect it will have on estimated population genetic structure and choose a polygon size that is suitable. A range of polygon lengths was included from 50 to 500 km to see the effects of polygon size. In addition to the original 500 km analysis, the analysis was performed again using lengths of 50 and 250 km, and the global population genetic structure measure results were compared using paired t-tests.

2.11 Testing Association Between Traits and Species Ranges (Module 15)

The relationship between biological traits and geographic range size was studied, using range size as the response variable. Range size in m^2 for each species included in the analysis was determined using the coordinates provided by both the BOLD and GBIF data, and the package sf version 1.0.7. (Pebesma, 2018). Only species found in both BOLD and GBIF were used and only the coordinates of the individual records, not entire BINs were used. The range sizes were then compared across species with different biological traits using the same steps outlined in previous steps.

3 Results

3.1 Pipeline and Module Development

A novel pipeline consisting of 15 modules was developed, 14 of which are used to determine population genetic structure measures as proxies for realized dispersal, and compare these measures across species with differing biological traits. The 15th module is an example of using range size. Best practices were met during the development of these modules, as outlined in Table 1. The modules ran without error and without unnecessary computational demands, and the checks included in the script showed that the R script was working as intended. Using reproducibility checkpoints after each module also showed that the modules produced the same results after each run.

3.2 Case Study

3.2.1 Population Genetic Structure Measures

Twenty BINs, representing 18 species, across 2 geographic regions were included in the analysis for Greenland (Fig. 5a; Table 3a). All the F_{ST} values for the BINs were low, showing low levels of genetic differentiation. The results for Hedrick's G_{ST} and Jost's D were more variable, with 6 BINs having a higher value (above 0.5) and 14 having a lower value for both measures. The AMOVA analysis revealed low levels of variation between the regions for all BINs. From the figures produced by the mismatch distribution analysis, 15 BINs' populations appeared to be unimodal, and five appeared to be bimodal. However, the dip test statistics indicated that there was no significant multimodality, suggesting that all populations are still expanding. An example of a mismatch distribution is provided in figure 6. The full analysis was completed again using a 4% clustering threshold instead of BINs. Eighteen clusters were included, and the results were similar to those of the BIN analysis. The full clustering analyses results are outlined in Appendix B.

The analysis was also completed for Canada. In total, 716 BINs across 37 polygons/regions were included (Fig. 5b; Table 3b). F_{ST}, G_{ST}, and Jost's D were determined for each region pair for each BIN (Fig.7). Due to the dataset size, global F_{ST}, G_{ST} and Jost's D were also determined for each BIN. Most F_{ST} values were below 0.1, indicating low levels of population genetic differentiation (Fig. 8a). However, most Jost's D and G_{ST} values were over 0.8, indicating higher levels of genetic differentiation (Fig. 8b & 8c). The AMOVA for most BINs revealed low levels of variation between regions, with most phi values being below 0.1. The mismatch analysis revealed no significant signs of multimodality, suggesting that the populations are still expanding.

3.2.2 Comparison Across Traits

The population genetic structure measures were compared across species/BINs with different biological traits. For the Greenland BIN analysis, 11 BINs were aquatic, and nine were terrestrial. For larval diet, 11 were phytophagous, five were saprophagous, three were fungivores, and one was predaceous. For adult diet, three were saprophagous, two were phytophagous, one was predaceous, and 14 were non-feeding. Traits represented by only one BIN were dropped from the statistical analyses. For both the ANOVA and PGLS analyses, there was no significant relationship between the measures and biological traits (Fig. 9, 10 & 11). As ANOVAs assume a normal distribution, a Shapiro-Wilk test was performed to test whether this assumption was met. This assumption was violated (p-value = 0.012), so a Kruskal-Wallis test was performed. The results did not differ between the Kruskal-Wallis and ANOVA. ANOVAs also assume homogeneity of variances, so a Levene test was performed. This assumption was not violated (p-value = 0.55).

For Canada, 164 BINs were aquatic, 56 were semi-aquatic, and 496 were terrestrial. For adult diet, three BINs were fungivores, 289 were non-feeding, 28 were both phytophagous and predaceous, 112 were saprophagous, 10 were both fungivores and phytophagous, 168 were phytophagous, 103 were predaceous and three were both saprophagous and phytophagous. For larval diet, 37 were fungivores, 20 were omnivores, 221 were phytophagous, 150 were predaceous, three were both saprophagous and phytophagous, 26 were both fungivores and phytophagous, 27 were parasitic, and 232 were saprophagous. The multiple regression revealed a significant relationship between the population genetic structure measures and habitat (Fig. 12; Table 4). An aquatic lifestyle was associated with significantly higher levels of genetic differentiation for Jost's D and G_{ST}, and terrestrial lifestyles were associated with significantly lower levels of genetic differentiation. However, F_{ST} showed the opposite trend, with aquatic lifestyles showing significantly lower genetic differentiation and terrestrial lifestyles significantly higher. Larval diet was also significantly related to population genetic structure measures (Fig. 13). Having a parasitic larval stage resulted in significantly lower levels of genetic differentiation for Jost's D and G_{ST}. The F_{ST} results suggested that higher levels of genetic differentiation were observed in BINs with a phytophagous larval stage, a saprophagous larval stage, and those whose larval stage was a mix of fungivore and phytophagous. When high levels of genetic differentiation were observed, this indicated low realized dispersal rates.

A significant relationship between the indicators of population genetic structure and the distance between geographic regions was found for all measures, suggesting that the further the geographic distance, the higher the levels of genetic differentiation (Fig. 14). In F_{ST} only, a positive relationship was found between the measure and the latitude and longitude coordinates of the center of the BINs' range. Multiple regressions have a series of assumptions, so tests were performed to determine whether these assumptions were met. There were no violations of the assumptions of homoscedasticity and linearity. However, the assumptions of multicollinearity and normality were violated. The variance inflation factor was calculated and suggested that the predictor variables are correlated. Based on a chi square test, the biological traits of the species are correlated with each other (p-value = 0.00023). A Shapiro-Wilk test revealed a violation of normality (p-value = 2.2×10^{-16}), but a Q-Q plot showed that this was not a large violation.

3.2.3 Sensitivity Analysis

A sensitivity analysis investigating the impact of polygon size on the results was completed for the Canada dataset (Table 5). At a polygon size of 50 km across, there were only 662 BINs available for analysis after filtering rather than 716, 105 polygons compared to 37, and 207,151 individuals compared to 225,774. At 250 km across there were only 624 BINs, 56 polygons and 210,196 individuals.

For the 50km analysis, across all population genetic structure measures, results differed significantly from that of the 500km analysis using paired t-tests and including

only those BINs present in both datasets (Jost's D (p-value = 0.0083), G_{ST} (p-value = 0.00045), F_{ST} (p-value = $2.2x10^{-16}$)). For the trait analysis, a significant relationship with the same directionality was found with habitat and larval diet in Jost's D and G_{ST}, when comparing the 50km analysis to the original 500km analysis. In addition to the relationship in aquatic and terrestrial BINs, there was also a significant relationship between the measures and semi-aquatic BINs, which had lower levels of genetic differentiation than other habitats. The F_{ST} results no longer showed any relationship with habitat. For larval diet and Jost's D, a significant relationship was not only found in BINs with a parasitic larval stage, but also in those with a phytophagous, predaceous, and saprophagous diet. All these diets were associated with lower levels of genetic differentiation. G_{ST} echoed the results for parasitic, as well as phytophagous and predaceous. The results for F_{ST} remained the same as the original analysis. A significant relationship was still found between indicators of population genetic structure and distance, except in F_{ST}. In addition to this, with Jost's D, a significant relationship was found with the number of cells in which a BIN was found. The more cells in which a BIN was found, the lower were the levels of genetic differentiation. The relationship between latitude and longitude and F_{ST} remained the same.

For the 250km analysis, the results only significantly differed in the global G_{ST} and global F_{ST}. The results for the trait analysis remained the same as the 500km analysis for both Jost's D and G_{ST}, except for a significant relationship with semi-aquatic BINs. For F_{ST}, the trait results remained the same as the 500km analysis, except there was no longer any relationship with aquatic or phytophagous BINs. The relationships between the indicators of population genetic structure, distance, latitude, and longitude, remained the same as from the 500km analysis.

4 Discussion

4.1 A Novel Pipeline for Large-Scale Comparative Population Genetics

The focus of this study was to create a publicly available pipeline connecting a set of modules to determine indicators of population genetic structure for a variety of taxa and compare these measures across species possessing differing biological traits. Throughout development, significant effort was made to adhere to best practices and ensure the modules were easily adaptable and useful for other researchers. Fifteen modules were included in this study. These modules can be used together to recreate the analysis we performed or used independently to suit the needs of the researcher. The modules included steps to extract data from various publicly available databases, such as BOLD and GBIF, to filter and group the data into polygons, to determine measures of population genetic structure, and to compare these measures across species with differing habitats and diets using an ANOVA, PGLS, and multiple

regression. The first module includes steps to set up and adjust the important variables for all other modules. Other variables are also adjustable throughout the script. The modules provide options to deal with datasets of varying sizes, and also provide options for different analyses based on the dataset and research needs. This novel pipeline provides a way to determine and compare population genetic structure measures for a variety of taxa and geographic regions. The utility of the modules was demonstrated through the case study on Diptera species. This case study shows the power and potential of the pipeline to generate interesting patterns and draw biological conclusions at a large scale.

4.2 Case Study

To test the modules, the analysis was run with Diptera species from Greenland and Canada. Most species/BINs/clusters showed low levels of genetic differentiation in Greenland, and there was no significant relationship between the biological traits and the population genetic structure measures. Results from Canada were more mixed. There were stronger disagreements between Jost's D and G_{ST}, and F_{ST} when compared to Greenland, with G_{ST} and Jost's D suggesting high levels of genetic differentiation and F_{ST} suggesting lower. Aquatic species exhibited greater levels of genetic differentiation, while terrestrial species and those with a parasitic larval stage had lower levels of genetic differentiation. There was also a significant relationship between the indicators of population genetic structure and distance. As the geographic distance increased, so did the genetic differentiation.

The strength of the Greenland analysis lies in the fact that Greenland provides an example of the minimum amount of data required to run the modules and analysis. By using this small dataset, we were able to show the script's functionality. To run the modules, at least two geographic regions/polygons and two BINs (or clusters or species) are required. When running a large analysis, it is beneficial to run the modules first on the minimal dataset size. By beginning with a sampled-down version of a dataset, a researcher can ensure the modules work for the dataset structure, intended research, and computing environment, before scaling up to a dataset size more appropriate for drawing biological conclusions. The Greenland dataset analysis ran without error and yielded reasonable results. Since the two available populations for Greenland were located in the same area and lacked any noticeable physical barrier, there is likely a lot of gene flow between these populations and a large crossover of genetic information. It is also likely that recently deglaciated habitats in Greenland were colonized by a modest number of individuals and then the population spread to adjacent areas. This geography and history were supported by the population genetic structure measures, most of which showed very low levels of genetic differentiation between the populations. This was also supported by the AMOVA and mismatch analysis, both of which suggested the populations are not genetically isolated. Due to the module's success with analyzing this small dataset size, we were able to scale up to a larger dataset and run our case study analysis on Diptera species from Canada.

The Canada dataset was much larger than that of Greenland. Both G_{ST} and Jost's D suggested that there is genetic differentiation between regions in Canada. This could be due to the geographical distance and space, as the regions in the Canada dataset were widespread and covered various geographic regions and habitat types across Canada. Genetic differentiation has been shown in Canadian fly species in other studies, such as Maxwell et al. (2013), who found significant differentiation among regions in British Columbia and evidence of isolation by distance. They also investigated genetic differentiation using AMOVA and F_{ST} (Maxwell et al., 2013). The results of the AMOVA and of F_{ST} differed between our analyses (Maxwell et al., 2013), with lack of genetic differentiation being observed in our study. However, the other population genetic structure measures, G_{ST} and Jost's D, did detect genetic differentiation and aligned with the findings of Maxwell et al. (2013). Maxwell et al. (2013) used microsatellites instead of CO1, and the species included in their analyses, *Rhagoletis indifferens,* was not included in our analysis. Despite these differences, both studies found similar results and evidence of genetic differentiation and isolation by distance (Maxwell et al., 2013). Our modules were not only able to draw conclusions about one species, but general trends across hundreds of Diptera species, making use of public databases and DNA barcoding data. This comparison shows the potential of our modules to determine indicators of population genetic structure and produce results that are supported by the current literature.

F_{ST} suggested lower levels of genetic differentiation. Low levels of genetic differentiation are not uncommon in Diptera. Populations have shown low genetic differentiation and high population connectivity in various regions at large geographic scales, including across Southeast Asia (Olival et al., 2013) and across several European countries (Gojkovic et al., 2020).

Not only did we determine levels of genetic differentiation, but we also compared these across Diptera species possessing different biological traits. All known Diptera species that met our sampling requirements were included, and results were found across hundreds of taxa at a Canadian scale. Several interesting patterns emerged within the results. Diptera species with aquatic lifestyles showed significantly more genetic differentiation, compared to terrestrial species. This could be due to the lack of connectivity between waterways, restricting the movement of aquatic species compared to terrestrial. Krosch et al. (2011) found that dispersal was limited to within streams in a species of Chironomidae in Australia. Similarly, Delettre & Morvan (2008) found that distance from waterbodies affected Chironomidae dispersal. The reliance on waterbodies may cause aquatic species not to be able to disperse as far, potentially reducing gene flow and increasing genetic differentiation. Terrestrial species may experience more connectivity due to the lack of these barriers. The results did contain some discrepancies, with G_{ST} and Jost's D supporting the above findings, but F_{ST}, however, suggesting the opposite. Further research should be done to better understand these discrepancies, and to choose the most suitable population genetics measure for a given molecular marker, but these results show the potential of this

pipeline to uncover interesting patterns and relationships between Diptera species and populations.

The modules were also successful in uncovering interesting relationships between the indicators of population genetic structure and larval diet. Jost's D and G_{ST} found that species with a parasite life stage had significantly less genetic differentiation than Diptera with other diets. This could be due to the reliance on a host. If the host species, such as other insects or birds, is able to disperse freely, then the fly may also be able to move across these distances and maintain population connectivity. F_{ST} also displayed several different patterns. F_{ST} suggested higher levels of genetic differentiation in larvae with a phytophagous diet, saprophagous diet, and a diet that is a combination of fungivore and phytophagous. It is possible that these species rely on a certain species of plant or fungi, and therefore may not be able to disperse and be as connected and widespread as those with a more generalist diet.

The pipeline not only uncovered relationships between indicators of population genetic structure and biological traits, but also showed its ability to detect significant patterns involving other important factors through multivariate analysis. By including these in the modules, we accounted for other variables that may be affecting indicators of population genetic structure and infer whether the patterns discussed earlier were solely due to biological traits. The G_{ST}, Jost's D, and F_{ST} all found that as the geographical distance increases, so does the genetic differentiation. F_{ST} also exhibited an additional relationship between the latitude and longitude coordinates of the centre of each species' range and the indicators of population genetic structure. The geographic distance and differences in barriers and climate that occur across latitude and longitude within Canada likely influenced the geneflow and connectivity of populations, resulting in differences in genetic differentiation. Historical factors, such as population separation in different glacial refugial areas, likely also affect patterns in Northern regions.

Several analyses were performed in this study with variations in important factors such as how species or clusters were defined, and the size of the geographic regions being investigated. Even within one analysis, the results between the population genetic structure measures differed. For the Canada dataset, F_{ST} produced much different results than G_{ST} and Jost's D. This is due to the different way in which F_{ST} and G_{ST} and Jost's D are calculated. Often for F_{ST}, the value does not actually go between 0 and 1, but rather between 0 and a value determined by the heterozygosity in the data (Verity & Nichols, 2015). G_{ST} and Jost's D use a standardized range of 0 – 1 (Verity & Nichols 2015). It is recommended to use G_{ST} and Jost's D in addition to F_{ST}, as all of these measures can be informative (Verity & Nichols, 2015). The module is easily adaptable, so if some measures are better suited to a given researcher's analysis, the option is there to include some measures and exclude others. The parameters of the functions used to calculate these measures could also be adjusted and a sensitivity analysis could be run to see how these measures and the discrepancies between them are affected. There were also differences between the BIN and cluster analysis. This suggests that it is important to consider which method is best for the analysis and

dataset. Results also somewhat differed when geographic regions of different sizes were used. It is important to consider what cell size is best for the geographic region and taxa under investigation.

To summarize the main results, in Canada evidence of genetic differentiation was detected, and there were significant relationships between the indicators of population genetic structure and habitat and larval diet. There was also significant evidence of isolation by distance and a relationship between indicators of population genetic structure and latitude and longitude. These results support our hypothesis that biological traits influence indicators of realized dispersal rates and population genetic structure, as they play an important role in the overall dispersal process. This case study provides a good demonstration of the utility of the pipeline as a whole, and of the modules individually. A variety of options are provided to account for different sizes of datasets and preferences and needs of the user. The modules ran efficiently and consistently, allowing results to be generated and biological conclusions to be drawn.

4.3 Suitable Data for Usage

Though the use of CO1 has some limitations, such as the limitations in its ability to infer deep phylogenetic relationships (Boyle & Adamowicz, 2015, Wilson, 2010), this marker was suitable for our analysis. As discussed earlier, CO1 is commonly used to identify and delineate species, and studies have also found success using it when investigating population genetic structure (Abuelmaali et al., 2021, Attiná et al., 2021, Choi et al., 2020, Froufe et al., 2014, Havel et al., 2000, Jossart et al., 2017, Liu et al., 2020, Meriam et al., 2015, Park et al., 2019, Pickett & David, 2018, Reed et al., 2006, Shum & Palumbi, 2021, Stark et al., 2021, Talbot et al., 2016, Troast et al., 2016, Xu et al., 2019). CO1 sequences of species having variability in this marker, and which are separate in CO1 profile from related species, are suitable for use in this pipeline. Though only CO1 was used for this analysis, various other molecular markers are suitable for use in this pipeline, including cytB and other commonly sequenced mitochondrial markers. Small adjustments would still need to be made, as some functions in the filtering step, like those from coil version 1.2.3 (Nugent et al., 2020), are built for CO1 only, and this could be replaced by steps to check for stop codons (in case of protein-coding genes) and anomalous sequences in the marker of choice.

This script is currently suitable for CO1 sequences, any geographic region(s), and a diverse selection of animal taxa, including terrestrial, marine, and freshwater groups from taxa such as arthropods, fish, and mammals. The pipeline also has the potential to be adapted to plant species, though with further consideration into the molecular markers and biological traits used. We recognize that some cases of hybrids or recently diverged species may be lumped together when using mitochondrial DNA, but we suggest that a large-scale analysis with large sample size is expected generally to reveal robust trends, a proposition that can be tested in future analyses. The other variables in the pipeline and the traits chosen for the analysis can be adjusted in order to suit the goals of the study and the study taxa.

4.4 Possible Extensions

In this study we presented an analysis framework to combine large-scale molecular estimates of genetic differentiation based on open-access molecular data of multiple species with species trait data. These modules provide a solid basis upon which other researchers can build and further develop to suit their own needs. The modules can be used together as a whole pipeline, as done in this analysis, or can be used individually and customized for the user's needs. Modules such as those that involve specific data, such as the data loading and filtering modules, can be replaced or adapted to deal with different data sets, molecular markers, geographic regions, and taxa. The key modules of the pipeline are those that define regions/polygons and calculate and compare population genetic structure measures to biological traits and other important factors. By adapting or replacing other modules, these core modules can be applied to a variety of taxa, geographic regions, and traits.

The pipeline can be expanded upon to include a variety of different data types and answer other kinds of biological questions. The pipeline can be used for de novo phylogenetics; however, a more complex clustering model should be used if users are planning to perform phylogenetics as part of their work. These modules can also be used with phylogeographic and speciation models. The results we obtained could be investigated further by including more statistical analysis. The idea of population stability and expansion could be investigated further, and analyses other than mismatch could be explored. There have been some critiques of mismatch analyses because there are many variables and opportunities to introduce errors (Grant 2015). These models can be influenced by the mutation rates, genes being investigated, sample sizes, and population characteristics (Grant 2015). Despite these critiques, mismatch distributions have proven useful when studying population genetic structure (Shum & Palumbi, 2021), and was worth including in our study as it provided additional insight into the populations being investigated. Further research and consideration should be done into the method and how it suits the data and study goals if further research is to be done into population expansion of Diptera species.

Another possible expansion could be to examine whether indicators of population genetic structure have any relationship with the species name. As BINs were used here, and some BINs contain multiple species names, this could have resulted in some grouping together of closely related species. The way traits are assigned can also be expanded upon. Currently, parasitoids and nectivores are included under parasitic and phytophagous respectively but should be separated into their own categories due to the differences between parasitoids and parasites and the ecological importance of the pollinator role.

The flexibility of this pipeline also allows for the inclusion of other types of data, such as climatic data and other types of molecular markers, such as cytB and other mitochondrial markers. For markers such as microsatellites, more extensive adaptation would be needed. If a length-variable genetic region, e.g., ITS, is going to be used, then

the trimming of the sequence to the reference step will need to be altered. The sequence may need to be anchored at the ends in more conserved regions, or a long sequence would need to be used as the reference. The pipeline can also be expanded upon to include multi-marker data. For our study, we were limited by the trait data available in the literature for our study taxa, as well as by the geographic regions that were sampled and barcoded. Further sampling and barcoding would allow for a more complete picture of fly species population genetic structure to be drawn. For future research, the modules can be expanded to include other geographic regions, molecular markers, taxa, and traits to suit the requirements and needs of the researchers. Data from other databases can also be included, such as GBIF.

The modules provide several options to account for datasets of differing sizes, but some analyses can still be computationally demanding. Another possible expansion of the pipeline would be to adapt the script to be run-on high-performance computing platforms. If the dataset is extremely large, future researchers may consider this option rather than running the modules on their individual desktop machines. Future researchers could also make use of parallel processing.

4.5 Concluding Remarks

In conclusion, we were able to produce an efficient, reproducible, and publicly available pipeline by selecting and combining modules and functionality for determining indicators of population genetic structure and comparing these measures across species possessing different biological traits. Not only were valuable modules developed and made available for future researchers, but biological conclusions were also drawn regarding the population genetic structure of Diptera species in Canada. The hypothesis that indicators of realized dispersal rates will be influenced by the traits of organisms was supported, and traits such as habitat and larval diet were shown to be related to the levels of genetic differentiation between populations in Canada. In addition, evidence was found for isolation by distance, and a positive relationship between the population genetic structure measures and latitude and longitude was found. As our climate and world change, it is important to understand processes such as dispersal and the factors influencing them. By exploring the population genetic structure of species currently inhabiting Arctic and Northern regions, we can better understand how they will be impacted by our changing climate. By developing programming pipelines and workflows and making these accessible, we provide the opportunity for our work to be expanded on and provide tools that can help other scientists answer new questions.

5 Acknowledgements

Authors acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Genome Canada. The authors

thank Ayesha Ali, for her helpful feedback and insight on the project. They thank Tyler Elliot for testing the modules and providing important feedback on the code. They also thank Matthew Orton, Jacqueline May and Mehra Balsara, whose scripts were referenced for the pipeline. They are grateful to the researchers who share their data on public databases such as the Barcode of Life Data Systems and the Global Biodiversity Information Facility, and to all the researchers and developers who contribute to R programming packages and documentation.

6 Conflict of Interest

The Authors declare that there are no conflicts of interest.

7 Author Contributions

Samantha E. Majoros, Sarah J. Adamowicz and Karl Cottenie designed the project. Samantha E. Majoros performed the pipeline development and data analysis. Samantha E. Majoros wrote the manuscript with input and feedback from Sarah J. Adamowicz and Karl Cottenie.

8 Figures & Tables

Table 1: How each best practice, as outlined by Reiter et al. (2021), Shade & Teal (2015) and Akalin (2018), was met during the creation of our pipeline.

Best Practices	How We Did It
Visualize pipeline	-Created outline of pipeline using Biorender
Ensure all steps are included and well commented	-Code divided into modules and clearly broken down into steps -Commenting provided for each line
Ensure pipeline is general enough to be used on multiple datasets	-General variable names used -Options and alternative code provided to fit different dataset sizes
Include all information on dataset, packages used, and choices made	-Provided description of dataset and choices made within the commenting -Provided list of packages in the first module, along with installation steps
Cite all packages used and any other pipelines consulted	-Other pipelines referenced at the top of the script -Citations provided for essential packages within the manuscript
Start with a small dataset to test code before scaling up	-Started with a small dataset from Greenland

Best Practices	How We Did It
Include steps to filter dataset	-Included filtering step to remove records without a BIN, CO1 sequence, species name or coordinates -Filtered out sequences with high gap content, indels and stop codons
Include steps to optimize parameters or do sensitivity analysis	-Ran a sensitivity analysis for the size of the polygons/cells
Make pipeline as efficient as possible	-Made use of looping to reduce computational demand and processing time -Include options to sample down datasets
Create figures as checks within the code	-Included steps to generate figures throughout the code, both as checks and to show results
Include performance checks to ensure code is working properly	-Included checks to ensure code was doing what we thought it was, such as checking for NAs in the dataset after we tried to remove them
Include reproducibility checkpoints	-Ran code multiple times and compared results after each module to ensure it was the same
Open pipeline up for review	-Sent code to colleagues for early review and received positive feedback -Have publicly released code and will monitor comments on GitHub

Best Practices	How We Did It
Keep track of changes and versions	-Kept track of changes using GitHub
Make pipeline publicly available	-Shared code in public repository on GitHub

Table 2: Brief description of each module and its purpose.

Module	Purpose
Module 1: Loading Packages and Setting up Variables	Provides installation and loading steps for packages. Highlights and provides the option to alter important variables for all other modules
Module 2: Extracting Data from Databases	Provides code to load in data from databases
Module 3: Filtering the Data	Includes steps to filter the data to the needs to the user. This includes steps to reduce the data to only needed information, and to ensure high quality sequences
Module 4: Organizing Data by Clusters	Clusters the sequences by a clustering threshold specified by the user
Module 5: Grouping by Geographic Region	Groups records into polygons/cells based on geographic coordinates. Includes steps to filter the data and regions based on the needs of the user
Module 6: Multiple Sequence Alignment	Aligns the sequences to a reference sequence
Module 7: F-Statistics and Multi-Allelic Measures	Includes code to determine pairwise and global Nei F_{ST}, Hedricks G_{ST} and Jost's D
Module 8: Analysis of Molecular Variance (AMOVA)	Used to run an AMOVA

Module	Purpose
Module 9: Mismatch Distributions	Used to determine mismatch distributions. Includes code to run a dip test to determine modality
Module 10: Phylogenetic Tree	Provides the option to determine the centroid sequence for each BIN and generate a maximum likelihood tree
Module 11: Analysis of Variance (ANOVA)	Includes steps to load in trait data and run an ANOVA
Module 12: Phylogenetic Generalized Least Squares Analysis (PGLS)	Includes steps to load in trait data and a phylogenetic tree, and to run a PGLS
Module 13: Multiple Regression	Includes steps to determine factors such as number of polygons, distance between farthest polygons, center coordinates for each BINs range, and number of records. Includes steps to run a multiple regression
Module 14: Sensitivity Analysis	Includes steps to load in different datasets for comparison and steps to run a paired t-test to compare results
Module 15: Range Size Analysis	Includes steps to combine BOLD and GBIF data, determine the range size and area for each species, and compare this across species with different traits as outlined in the previous modules

Table 3: Population genetic structure measures for BINs from a) Greenland and b) Canada. The Greenland populations showed few signs of genetic differentiation across all measures. There are 716 BINs in the Canada analysis, and only the max and min values for each measure are shown. Most F_{ST} values were below 0.1, indicating low levels of genetic differentiation, but most G_{ST} and Jost's D values were above 0.8, indicating higher levels of genetic differentiation. Most AMOVA phi values were below 0.1, revealing low levels of variation between regions. The mismatch analysis revealed no significant signs of multimodality.

a) Greenland

BIN	Species Name	Region/ Cell Number	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
BOLD: AAB0080	*Corynoneura arctica*	7721, 7478	0.0078	0.31	0.3	0.012	0.001	0.058	0.98
BOLD: AAB7912	*Limnophyes brachytomus*	7721, 7478	0.0025	0.15	0.15	-0.0012	0.6	0.06	0.99
BOLD: AAC4201	*Paraphaenocladius impensus*	7721, 7478	0.0061	0.2	0.19	0.0096	0.001	0.025	1
BOLD: AAC8434	*Nephrotoma lundbecki*	7721, 7478	0.029	0.38	0.34	0.024	0.13	0.078	0.9

BIN	Species Name	Region/ Cell Number	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
BOLD: AAC9614	*Protophormia terraenovae*	7721, 7478	0.044	0.84	0.82	0.067	0.008	0.047	0.99
BOLD: AAD7061	*Diamesa arctica*	7721, 7478	0.011	0.78	0.78	0.0083	0.018	0.098	0.41
BOLD: AAE8704	*Smittia extrema*	7721, 7478	0.016	0.47	0.45	0.028	0.001	0.044	1
BOLD: AAL7378	*Cricotopus patens*	7721, 7478	0.05	0.75	0.72	0.054	0.05	0.12	0.24
BOLD: AAL7869	*Lycoriella flavipeda*	7721, 7478	0.0027	0.16	0.15	0.0047	0.001	0.016	1
BOLD: AAL9235	*Limnophyes pumilio*	7721, 7478	0.025	0.89	0.89	0.038	0.001	0.05	0.99
BOLD: AAL9801	*Drymeia groenlandica*	7721, 7478	0.042	0.43	0.38	0.063	0.001	0.025	1

BIN	Species Name	Region/ Cell Number	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
BOLD: AAM0255	*Diamesa chorea*	7721, 7478	0.014	0.23	0.2	0.021	0.004	0.04	1
BOLD: AAM9109	*Spilogona sanctipauli*	7721, 7478	0.011	0.18	0.16	-0.0017	0.431	0.025	1
BOLD: AAP8779	*Camptochaeta delicata*	7721, 7478	0.0081	0.25	0.24	0.0087	0.027	0.025	1
BOLD: AAU3704	*Limnophyes ninae*	7721, 7478	0.0062	0.36	0.35	0.0065	0.016	0.086	0.78
BOLD: AAZ6073	*Lycoriella riparia*	7721, 7478	0.0084	0.23	0.22	-0.0051	0.567	0.1	0.45
BOLD: AAZ6184	*Megaselia cirriventris*	7721, 7478	0.023	0.69	0.68	0.041	0.001	0.033	1
BOLD: ACA8845	*Phytomyza puccinelliae*	7721, 7478	0.0098	0.28	0.27	0.006	0.21	0.033	1

BIN	Species Name	Region/ Cell Number	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
BOLD: ACI8602	*Limnophyes brachytomus*	7721, 7478	0.034	0.71	0.69	0.039	0.033	0.033	1
BOLD: ACM4349	*Limnophyes brachytomus*	7721, 7478	0.0065	0.14	0.13	0.0078	0.029	0.025	1

b) Canada

Measure	Max	Min
Global F_{ST}	0.48	-0.002
Global G_{ST}	1	-0.009
Jost's D	1	-0.005
AMOVA Phi Value	0.63	-0.042
AMOVA p-value	1	0.01
Dip Test Value	0.5	0.015
Dip Test p-value	1	0.01

Table 4: Results from the multiple regression testing whether traits and geographic variables are significant predictors of population genetic structure measures for Canada for a) Jost's D, b) G_{ST}, and c) F_{ST}. Significant values are bolded.

a) Jost's D (df = 694)

Trait/Factors	t-value	p-value
Aquatic	**5.36**	**1.04e-07**
Semi-Aquatic	-1.34	0.18
Terrestrial	**-3.2**	**0.0015**
Adult Diet: Fungivore & Phytophagous	-0.55	0.58
Adult Diet: Non-Feeding	0.14	0.89
Adult Diet: Phytophagous	0.36	0.72
Adult Diet: Phytophagous & Predaceous	0.6	0.55
Adult Diet: Predaceous	0.64	0.52
Adult Diet: Saprophagous	-0.53	0.6
Adult Diet: Saprophagous & Phytophagous	-0.79	0.43
Larval Diet: Fungivore & Phytophagous	-1.72	0.085
Larval Diet: Omnivores	-0.88	0.38

Trait/Factors	t-value	p-value
Larval Diet: Parasitic	**-2.59**	**0.0097**
Larval Diet: Phytophagous	-1.06	0.29
Larval Diet: Phytophagous & Predaceous	-0.93	0.35
Larval Diet: Predaceous	-1.59	0.11
Larval Diet: Saprophagous	-0.44	0.66
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	0.027	0.98
Number of Records	-1.11	0.27
Max Distance	**2.98**	**0.003**
Longitude Coordinate	-0.15	0.88
Latitude Coordinate	0.94	0.35

b) G_{ST} (df = 694)

Trait/Factors	t-value	p-value
Aquatic	**5.6**	**3.16e-08**
Semi-Aquatic	-1.26	0.21
Terrestrial	-3.03	**0.0025**
Adult Diet: Fungivore & Phytophagous	-0.6	0.55
Adult Diet: Non-Feeding	0.15	0.88
Adult Diet: Phytophagous	0.37	0.71
Adult Diet: Phytophagous & Predaceous	0.6	0.55
Adult Diet: Predaceous	0.64	0.52
Adult Diet: Saprophagous	0.51	0.61
Adult Diet: Saprophagous & Phytophagous	-0.73	0.47
Larval Diet: Fungivore & Phytophagous	-1.66	0.097
Larval Diet: Omnivores	-0.84	0.4
Larval Diet: Parasitic	**-2.54**	**0.011**

Trait/Factors	t-value	p-value
Larval Diet: Phytophagous	-1	0.32
Larval Diet: Phytophagous & Predaceous	-0.81	0.41
Larval Diet: Predaceous	-1.51	0.13
Larval Diet: Saprophagous	-0.36	0.72
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	-0.05	0.96
Number of Records	-1.05	0.3
Max Distance	**3.13**	**0.002**
Longitude Coordinate	0.04	0.97
Latitude Coordinate	1.26	0.21

c) F_{ST} (df = 694)

Trait/Factors	t-value	p-value
Aquatic	**-2.5**	**0.013**
Semi-Aquatic	-0.2	0.84
Terrestrial	**2.21**	**0.028**
Adult Diet: Fungivore & Phytophagous	-1.18	0.24
Adult Diet: Non-Feeding	-0.11	0.91
Adult Diet: Phytophagous	0.34	0.74
Adult Diet: Phytophagous & Predaceous	-0.33	0.74
Adult Diet: Predaceous	0.72	0.48
Adult Diet: Saprophagous	-0.28	0.78
Adult Diet: Saprophagous & Phytophagous	0.21	0.83
Larval Diet: Fungivore & Phytophagous	**3.6**	**0.00034**
Larval Diet: Omnivores	1.1	0.27
Larval Diet: Parasitic	0.85	0.39

Trait/Factors	t-value	p-value
Larval Diet: Phytophagous	**2.44**	**0.015**
Larval Diet: Phytophagous & Predaceous	1.25	0.21
Larval Diet: Predaceous	1.12	0.26
Larval Diet: Saprophagous	**3.15**	**0.0017**
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	-1.34	0.18
Number of Records	0.51	0.61
Max Distance	**2.7**	**0.007**
Longitude Coordinate	**4.2**	**3.09e-05**
Latitude Coordinate	**8.33**	**4.47e-16**

Table 5: Results from the sensitivity analysis testing the influence of polygon size upon the relationship between population genetic measures and traits for a) Jost'D, b) G_{ST}, and c) F_{ST}. df for the 50km analysis = 640. df for the 250km analysis = 602. df for the 500km analysis = 694

a) Jost's D

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Aquatic	**5.36**	**1.04e-07**	**7.4**	**4.21e-13**	**3.7**	**0.00024**
Semi-Aquatic	-1.34	0.18	**-2.16**	**0.031**	**-2.24**	**0.025**
Terrestrial	**-3.2**	**0.0015**	**-4.15**	**3.76e-05**	**-3.21**	**0.0014**
Adult Diet: Fungivore & Phytophagous	-0.55	0.58	-0.099	0.92	0.96	0.34
Adult Diet: Non-Feeding	0.14	0.89	0.004	1	1.15	0.25
Adult Diet: Phytophagous	0.36	0.72	0.43	0.67	0.35	0.72
Adult Diet: Phytophagous & Predaceous	0.6	0.55	0.83	0.41	1.62	0.11
Adult Diet: Predaceous	0.64	0.52	0.64	0.52	1.2	0.23
Adult Diet: Saprophagous	-0.53	0.6	-0.2	0.84	0.57	0.57

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Adult Diet: Saprophagous & Phytophagous	-0.79	0.43	-0.32	0.75	-0.31	0.76
Larval Diet: Fungivore & Phytophagous	-1.72	0.085	-1.15	0.25	-1.54	0.12
Larval Diet: Omnivores	-0.88	0.38	-1.59	0.11	-1.84	0.066
Larval Diet: Parasitic	**-2.59**	**0.0097**	**-2.11**	**0.035**	-1.51	0.13
Larval Diet: Phytophagous	-1.06	0.29	**-2.2**	**0.029**	-0.15	0.88
Larval Diet: Phytophagous & Predaceous	-0.93	0.35	-1.81	0.071	-1.15	0.25
Larval Diet: Predaceous	-1.59	0.11	**-2.12**	**0.035**	-1.15	0.25
Larval Diet: Saprophagous	-0.44	0.66	**-2**	**0.046**	-0.72	0.47
Larval Diet: Saprophagous & Phytophagous	N/A	N/A	N/A	N/A	N/A	N/A
Number of Cells	0.027	0.98	**-2.19**	**0.029**	-1.2	0.23

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Number of Records	-1.11	0.27	0.27	0.79	-0.5	0.62
Max Distance	**2.98**	**0.003**	**6.99**	**6.78e-12**	**3.9**	**0.00011**
Longitude Coordinate	-0.15	0.88	0.22	0.82	-0.33	0.74
Latitude Coordinate	0.94	0.35	-0.69	0.49	1.63	0.1

b) G_{ST}

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Aquatic	**5.6**	**3.16e-08**	**8**	**5.94e-15**	**4.17**	**3.55e-05**
Semi-Aquatic	-1.26	0.21	**-2.12**	**0.035**	**-2.13**	**0.034**
Terrestrial	-3.03	**0.0025**	**-3.94**	**8.96e-05**	**-3.038**	**0.0025**
Adult Diet: Fungivore & Phytophagous	-0.6	0.55	-0.23	0.82	0.94	0.35
Adult Diet: Non-Feeding	0.15	0.88	-0.011	0.99	1.14	0.25
Adult Diet: Phytophagous	0.37	0.71	0.41	0.68	0.28	0.78
Adult Diet: Phytophagous & Predaceous	0.6	0.55	0.73	0.47	1.58	0.12
Adult Diet: Predaceous	0.64	0.52	0.64	0.52	1.15	0.25
Adult Diet: Saprophagous	0.51	0.61	-0.26	0.79	0.55	0.58
Adult Diet: Saprophagous & Phytophagous	-0.73	0.47	-0.3	0.76	-0.3	0.77

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Larval Diet: Fungivore & Phytophagous	-1.66	0.097	-0.95	0.34	-1.44	0.15
Larval Diet: Omnivores	-0.84	0.4	-1.44	0.15	-1.88	0.06
Larval Diet: Parasitic	**-2.54**	**0.011**	**-2.1**	**0.036**	-1.48	0.14
Larval Diet: Phytophagous	-1	0.32	**-2.045**	**0.041**	-0.1	0.92
Larval Diet: Phytophagous & Predaceous	-0.81	0.41	-1.62	0.11	-1.053	0.29
Larval Diet: Predaceous	-1.51	0.13	-2	**0.046**	-1.041	0.3
Larval Diet: Saprophagous	-0.36	0.72	-1.74	0.082	-0.64	0.52
Larval Diet: Saprophagous & Phytophagous	N/A	N/A	N/A	N/A	N/A	N/A
Number of Cells	-0.05	0.96	**-2.015**	**0.044**	-1.27	0.2
Number of Records	-1.05	0.3	0.28	0.78	-0.4	0.69

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Max Distance	**3.13**	**0.002**	**7.014**	**5.92e-12**	**4.038**	**6.08e-05**
Longitude Coordinate	0.04	0.97	0.36	0.72	-0.16	0.88
Latitude Coordinate	1.26	0.21	-0.89	0.37	1.64	0.1

c) F_{ST}

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Aquatic	**-2.5**	**0.013**	0.18	0.86	-0.98	0.33
Semi-Aquatic	-0.2	0.84	-0.45	0.65	0.83	0.41
Terrestrial	**2.21**	**0.028**	1.56	0.12	**3.37**	**0.00079**
Adult Diet: Fungivore & Phytophagous	-1.18	0.24	-0.79	0.43	-0.31	0.76
Adult Diet: Non-Feeding	-0.11	0.91	-0.39	0.7	-0.17	0.86
Adult Diet: Phytophagous	0.34	0.74	-0.42	0.68	-1.28	0.2
Adult Diet: Phytophagous & Predaceous	-0.33	0.74	-1.11	0.27	-0.66	0.51
Adult Diet: Predaceous	0.72	0.48	-0.007	0.99	-0.081	0.94
Adult Diet: Saprophagous	-0.28	0.78	-0.83	0.41	-0.61	0.54
Adult Diet: Saprophagous & Phytophagous	0.21	0.83	-0.26	0.8	-0.32	0.75

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Larval Diet: Fungivore & Phytophagous	**3.6**	**0.00034**	**3.36**	**0.00082**	1.19	0.24
Larval Diet: Omnivores	1.1	0.27	1.44	0.15	0.59	0.55
Larval Diet: Parasitic	0.85	0.39	1.066	0.29	1.74	0.082
Larval Diet: Phytophagous	**2.44**	**0.015**	**2.43**	**0.016**	**3.27**	**0.0011**
Larval Diet: Phytophagous & Predaceous	1.25	0.21	0.99	0.32	0.92	0.36
Larval Diet: Predaceous	1.12	0.26	1.87	0.062	1.12	0.26
Larval Diet: Saprophagous	**3.15**	**0.0017**	**2.7**	**0.0071**	**2.33**	**0.02**
Larval Diet: Saprophagous & Phytophagous	N/A	N/A	N/A	N/A	N/A	N/A
Number of Cells	-1.34	0.18	1.057	0.29	-0.65	0.52
Number of Records	0.51	0.61	-0.006	0.99	0.49	0.63

Trait/Factors	500km t-value	500km p-value	50km t-value	50km p-value	250km t-value	250km p-value
Max Distance	**2.7**	**0.007**	1.18	0.24	**2.021**	**0.04**
Longitude Coordinate	**4.2**	**3.09e-05**	**3.24**	**0.0012**	**4.37**	**1.49e-05**
Latitude Coordinate	**8.33**	**4.47e-16**	**3.71**	**0.00023**	**6.3**	**5.65e-10**

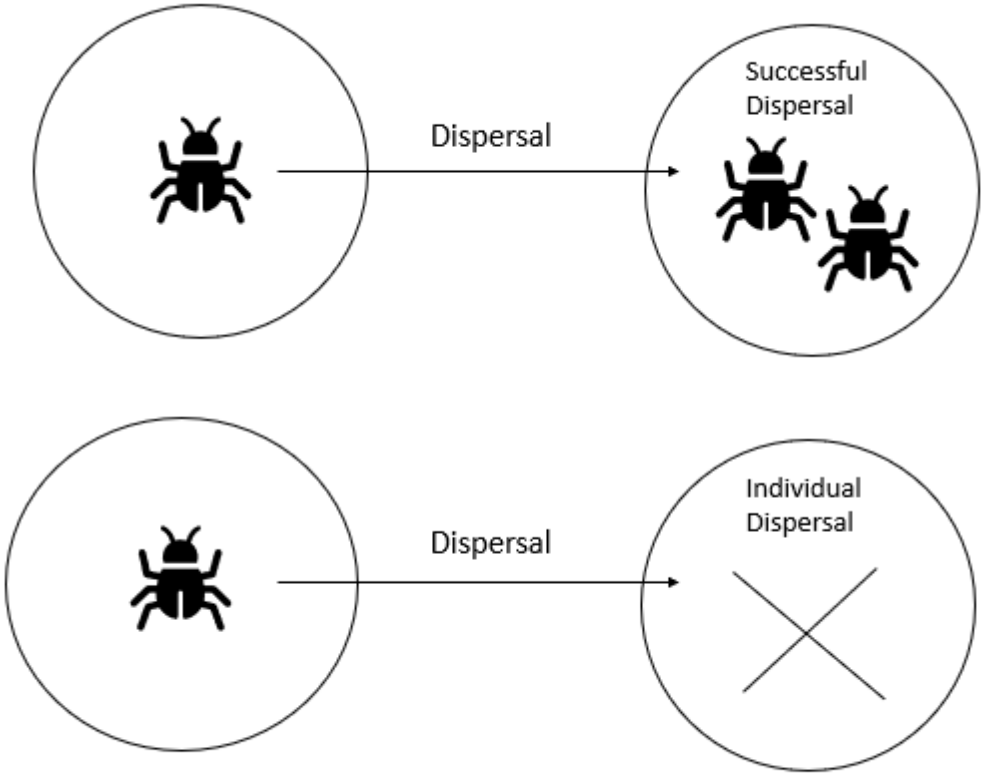
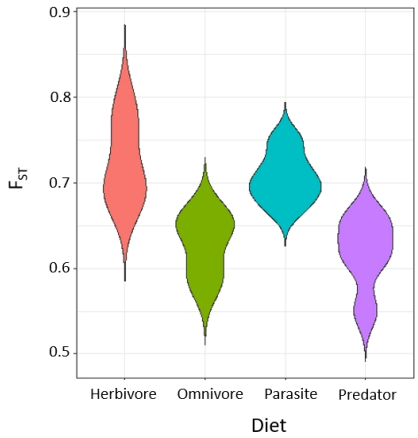


Figure 1: The concept of realized dispersal. The top image shows realized dispersal, in which the organism disperses and successfully colonizes the new environment and establishes a reproducing population or leaves genetic material in an already established population. The bottom image shows individual dispersal, but there is no successful colonization.

a)



b)

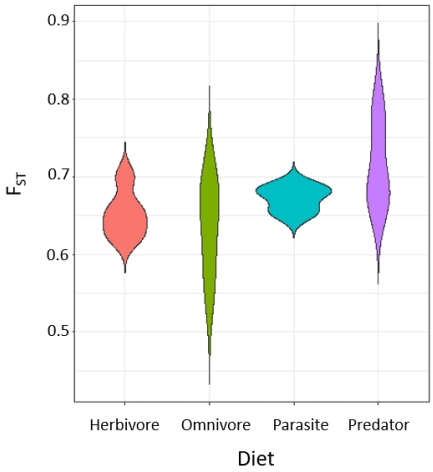


Figure 2: Violin plots based on artificial diet data showing the expected results if our hypothesis is a) supported or b) not supported. We hypothesize that species with similar traits will have similar dispersal rates due to the importance of traits to the dispersal and colonization process. We expect taxa with different traits to have different F_{ST} values. We expect that taxa with a generalist or omnivorous diet to have higher successful dispersal rates, and hence lower population genetic differentiation. If the null hypothesis is true, the F_{ST} values remain constant across feeding modes.

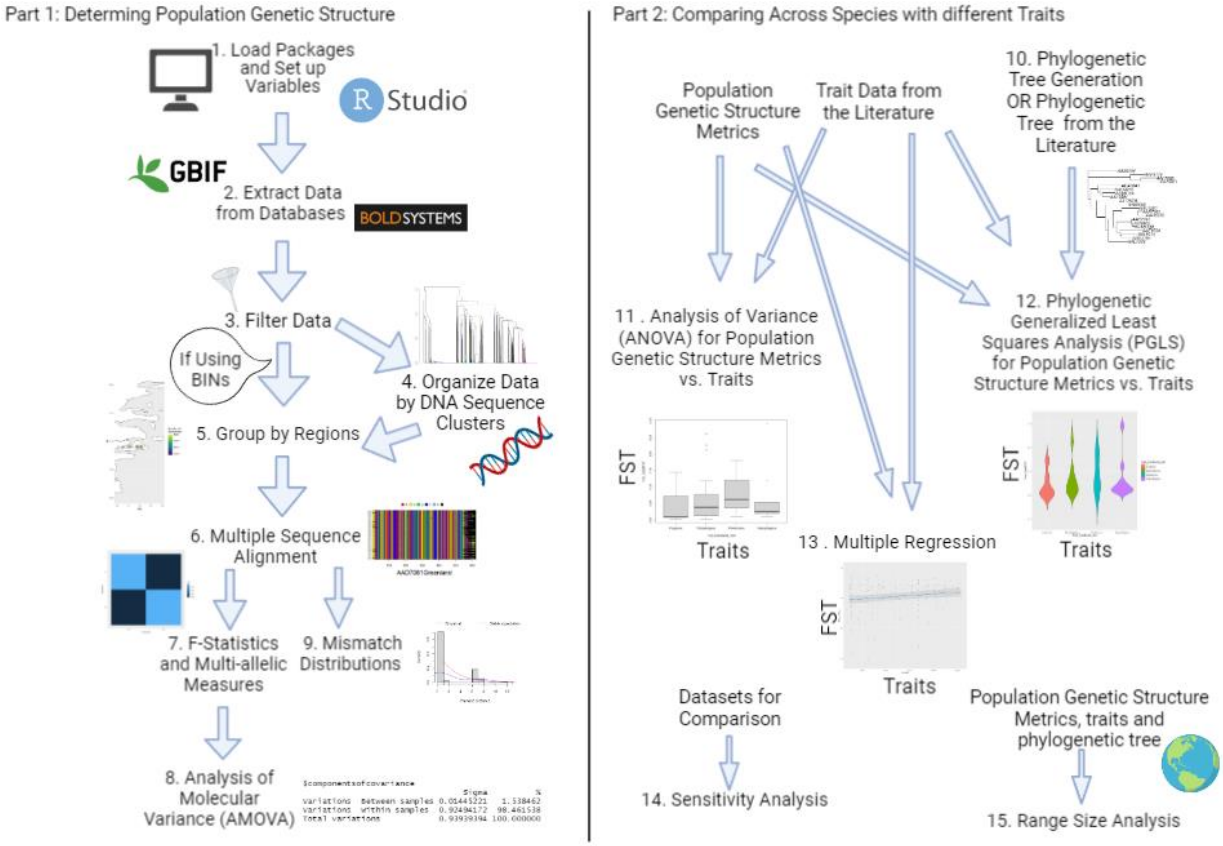


Figure 3: Steps in the population genetic strutcure pipeline. The numbered steps correspond to sections of code.

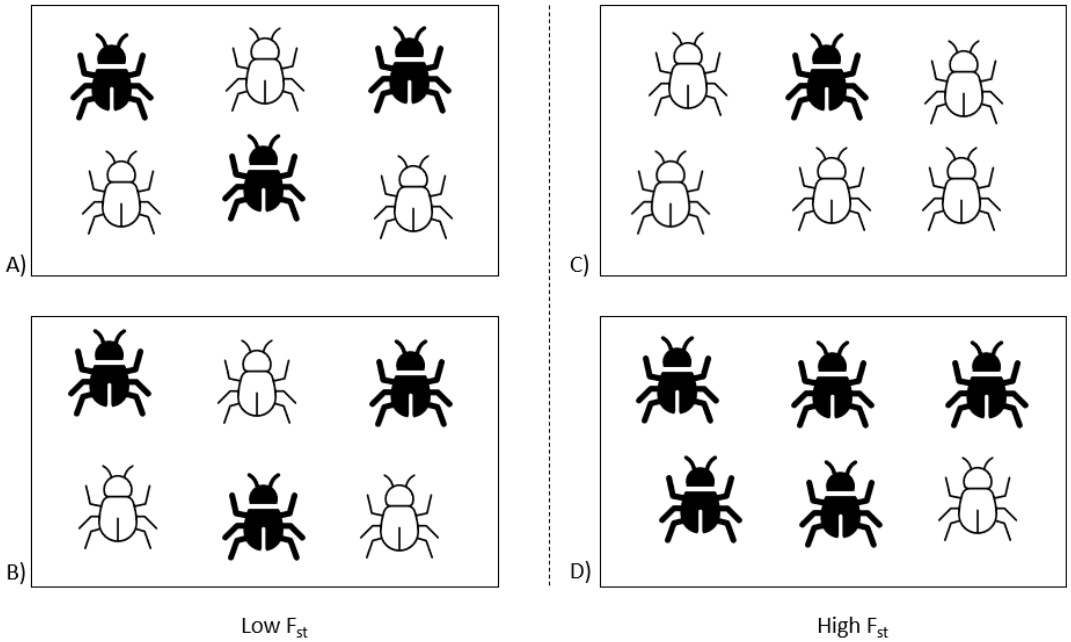
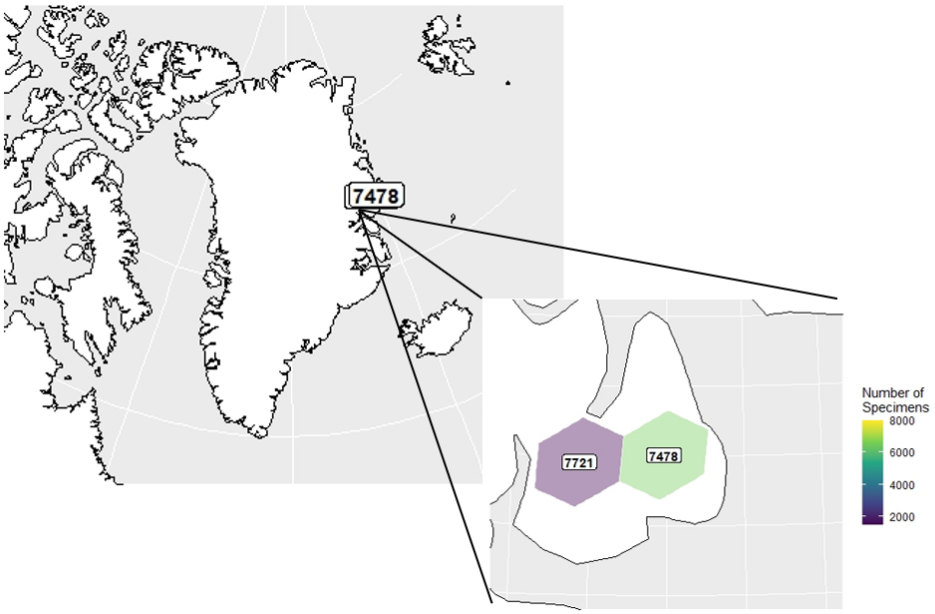


Figure 4: Conceptual illustration of low vs high F_{ST}. The colours represent different alleles (haplotypes) of CO1 within the same species. Panels A & B show low F_{ST}. The two populations share the same allele frequencies. Panels C & D show high F_{ST}. While the global allele frequencies are the same as the other panels, in this case the two sub-populations have very different allele frequencies, indicating strong population genetic differentation, due to the amount of gene flow as well as historical colonization patterns.

a) Greenland



b) Canada

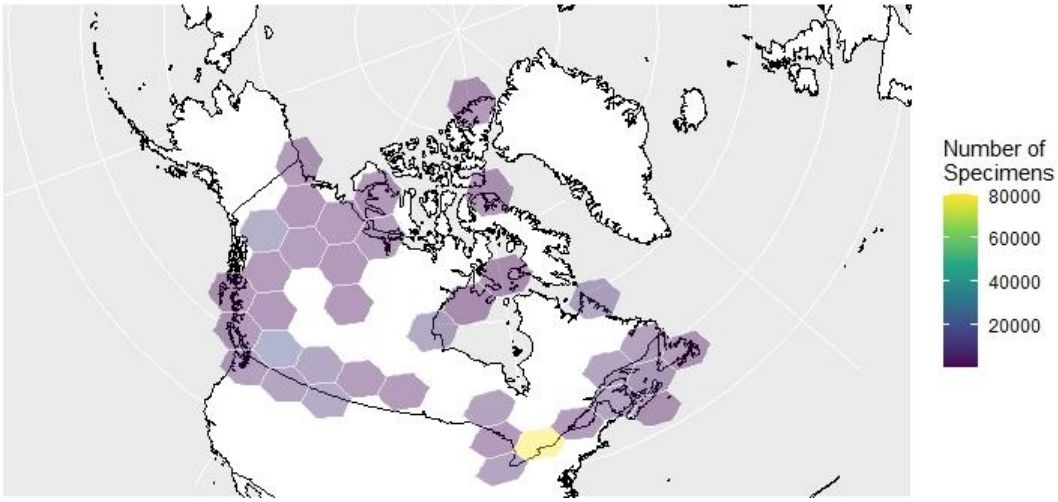


Figure 5: Regions and number of Diptera records for the case study BIN analysis in a) Greenland (polygon length = 50 km across) and b) Canada (polygon length = 500 km across). Regions labels were removed from the Canada map to avoid overcrowding.

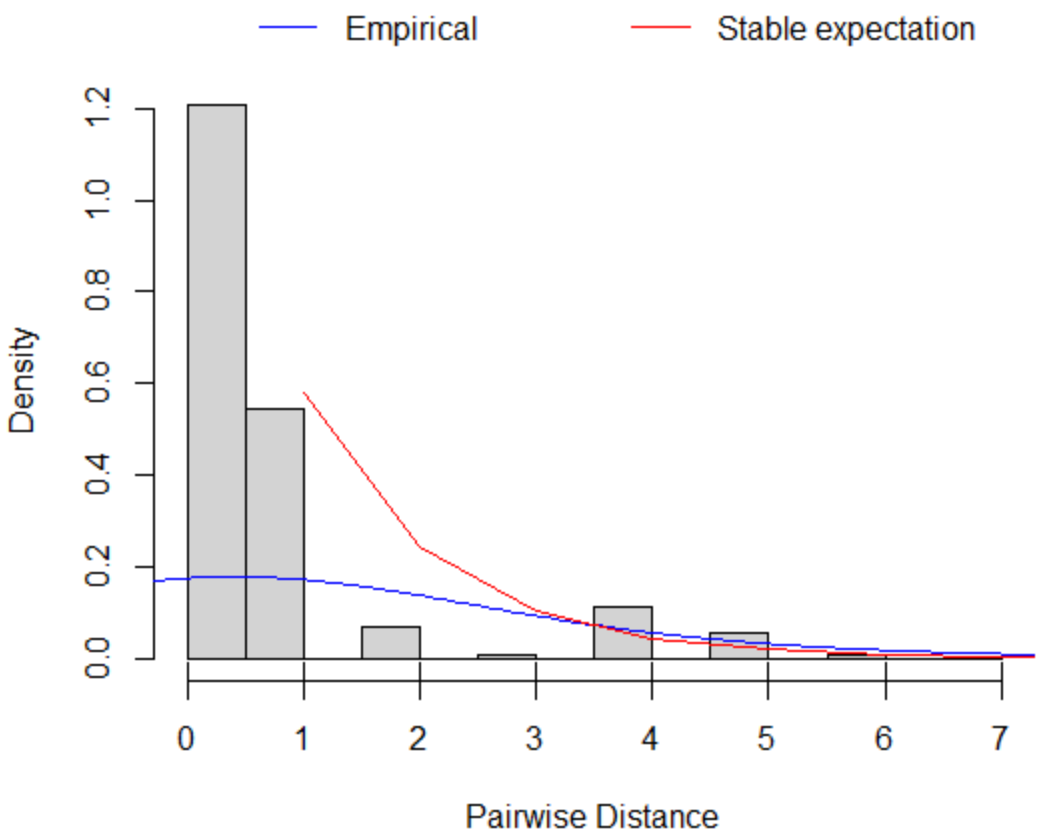
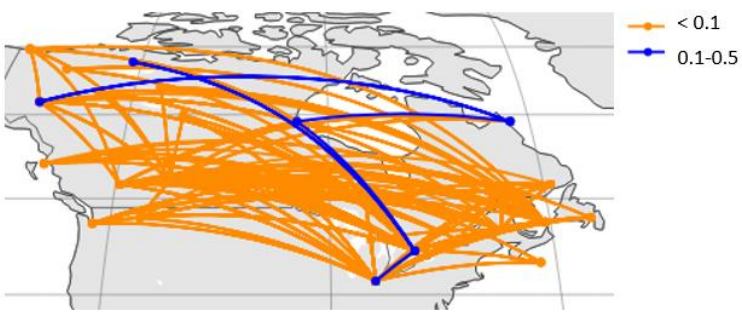
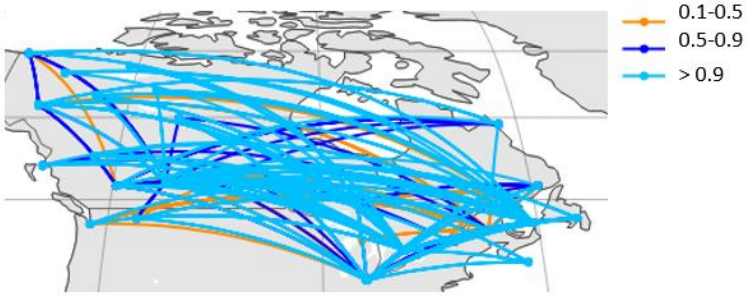


Figure 6: An example of a mismatch distribution for BIN AAB0080. There were no significant signs of mulitmodality.

a) F_{ST}



b) G_{ST}



c) Jost's D

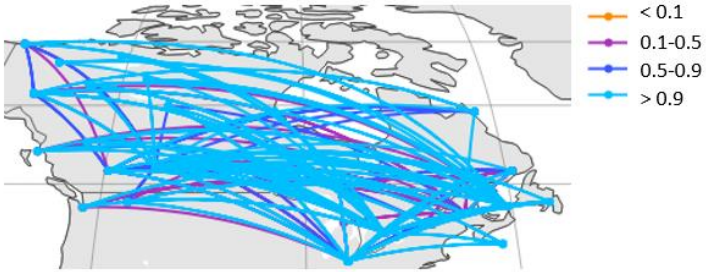


Figure 7: The pairwise a) F_{ST}, b) G_{ST}, and c) Jost's D between geographic regions in Canada for the Diptera BIN analysis. Only 10% of records from the 10 largest BINs are shown as an example. Most F_{ST} values were below 0.1, suggesting low levels of genetic differenitation, while most G_{ST} and Jost's D values were above 0.9, suggesting higher levels of genetic differentiation.

a) F_{ST} b) G_{ST} c) Jost's D

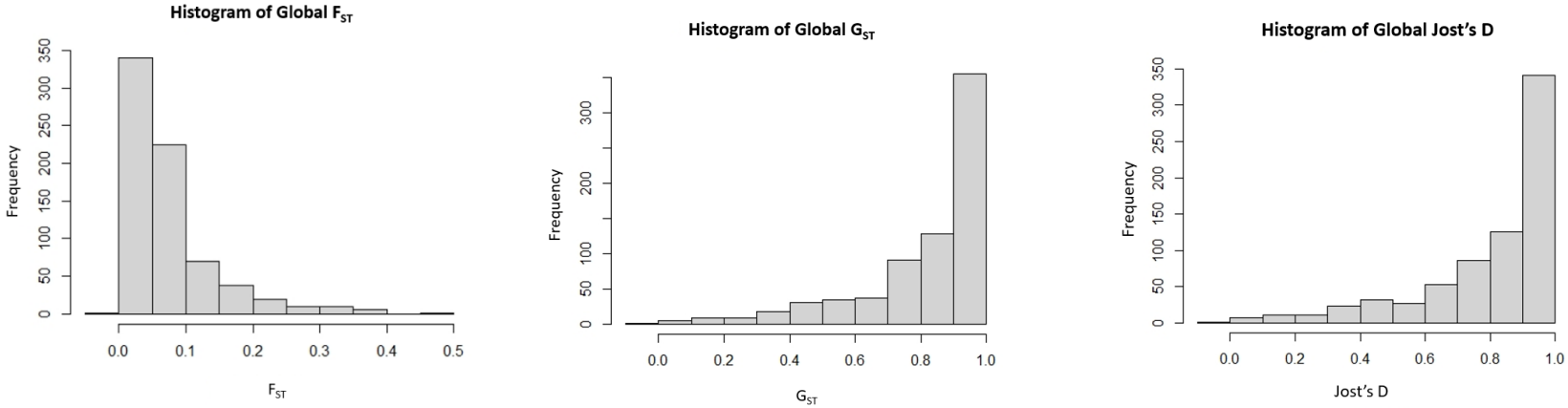
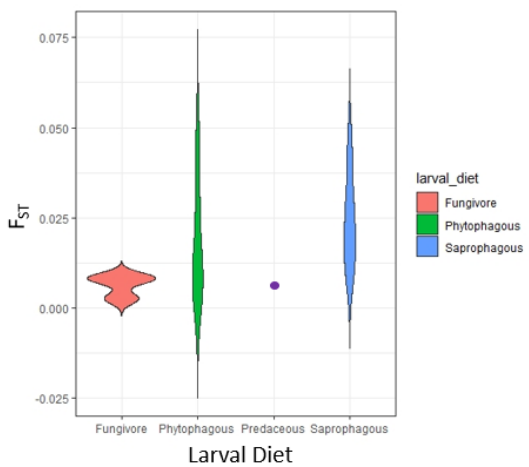
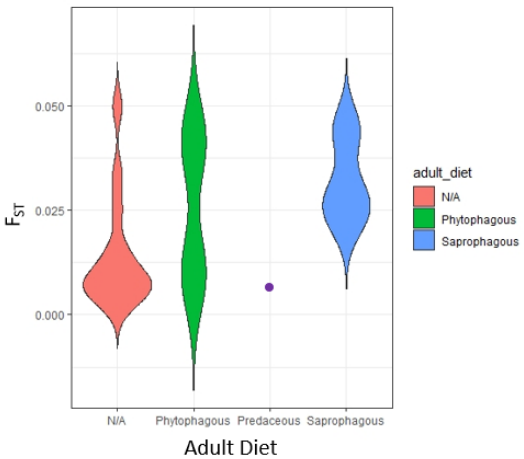


Figure 8 : The distribution of the global (one value per BIN) a)Nei F_{ST} values, b) Hedrick's G_{ST} values, and c) Jost's D values for the Canada Diptera BIN analysis (716 BINs). Most F_{ST} values were below 0.1, suggesting low levels of genetic differenitation, while most G_{ST} and Jost's D values were above 0.8, suggesting higher levels of genetic differentiation.

a) F_{ST} Larval Diet



b) F_{ST} Adult Diet



c) F_{ST} Habitat

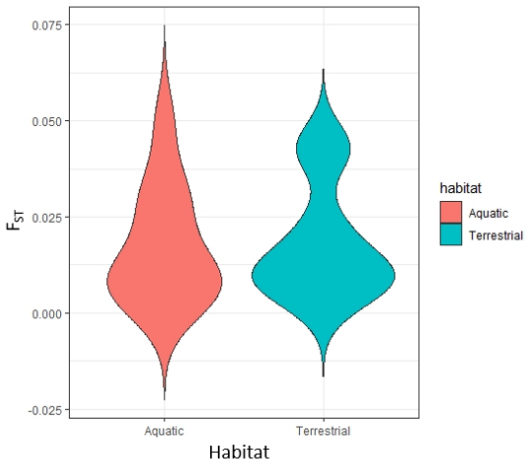
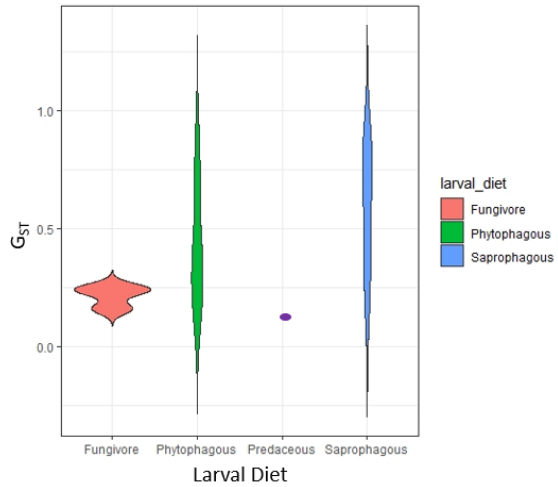
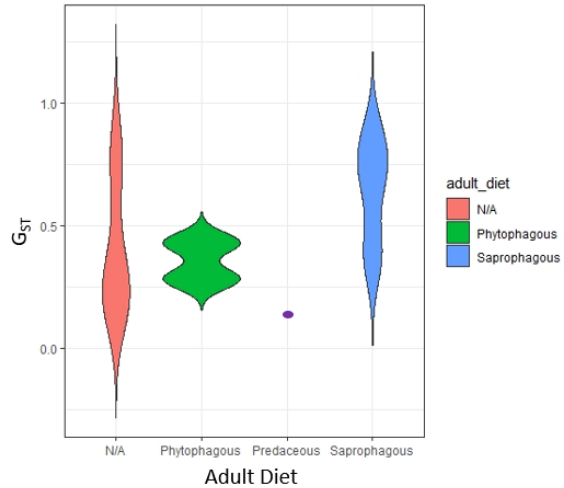


Figure 9: Violin plots showing the relationship between biological traits and population genetic structure measures for the BIN analysis for Greenland Diptera (20 BINs). Results are shown comparing Nei's F_{ST} to a) larval diet, b) adult diet, and c) habitat. There was no significant relationship between the population genetic structure measures and traits. Traits with fewer than two datapoints were dropped from the analysis, but are shown by dots on the figures.

a) G_{ST} Larval Diet



b) G_{ST} Adult Diet



c) G_{ST} Habitat

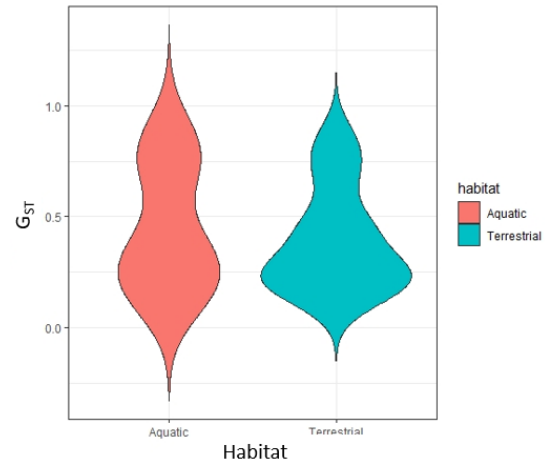


Figure 10: Violin plots showing the relationship between biological traits and population genetic structure measures for the BIN analysis for Greenland Diptera (20 BINs). Results are shown comparing Hedricks G_{ST} to a) larval diet, b) adult diet, and c) habitat. There was no significant relationship between the population genetic structure measures and traits. Traits with fewer than two datapoints were dropped from the analysis, but are shown by dots on the figures.

a) Jost's D Larval Diet

b) Jost's D Adult Diet

c) Jost's D Habitat

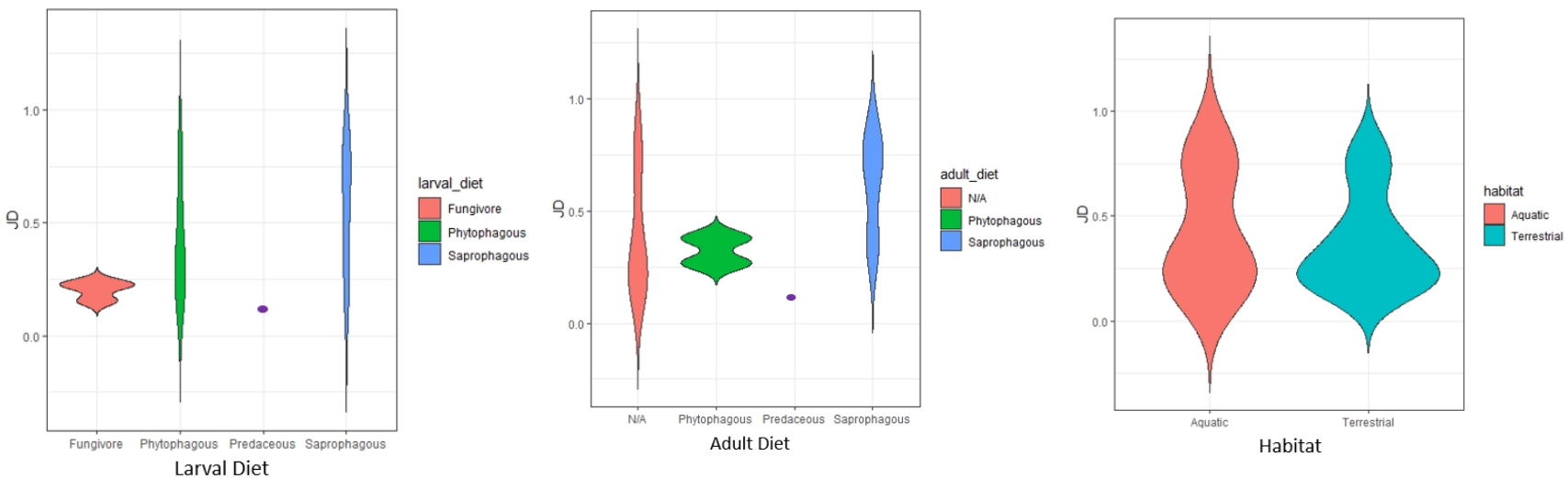


Figure 11: Violin plots showing the relationship between biological traits and population genetic structure measures for the BIN analysis for Greenland Diptera (20 BINs). Results are shown comparing Jost's D to a) larval diet, b) adult diet, and c) habitat. There was no significant relationship between the population genetic structure measures and traits. Traits with fewer than two datapoints were dropped from the analysis, but are shown by dots on the figures.

a) Jost's D Habitat

b) F_{ST} Habitat

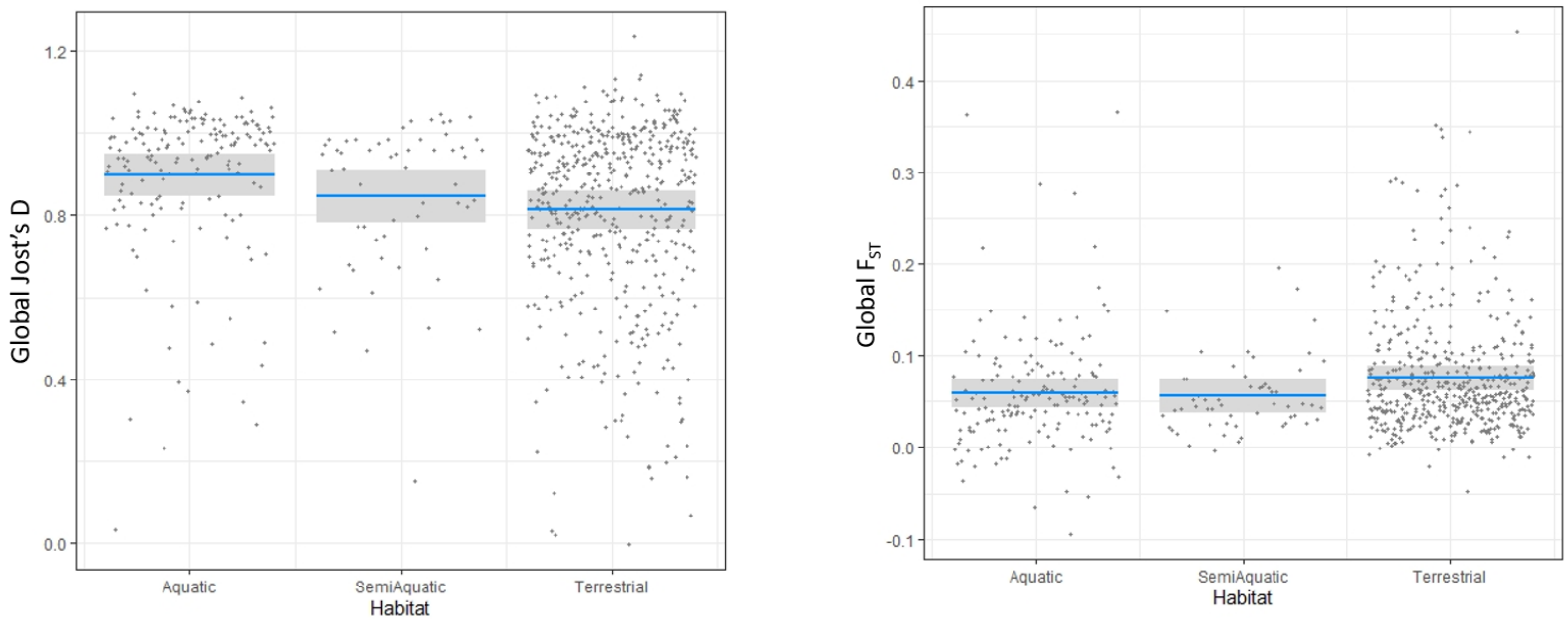
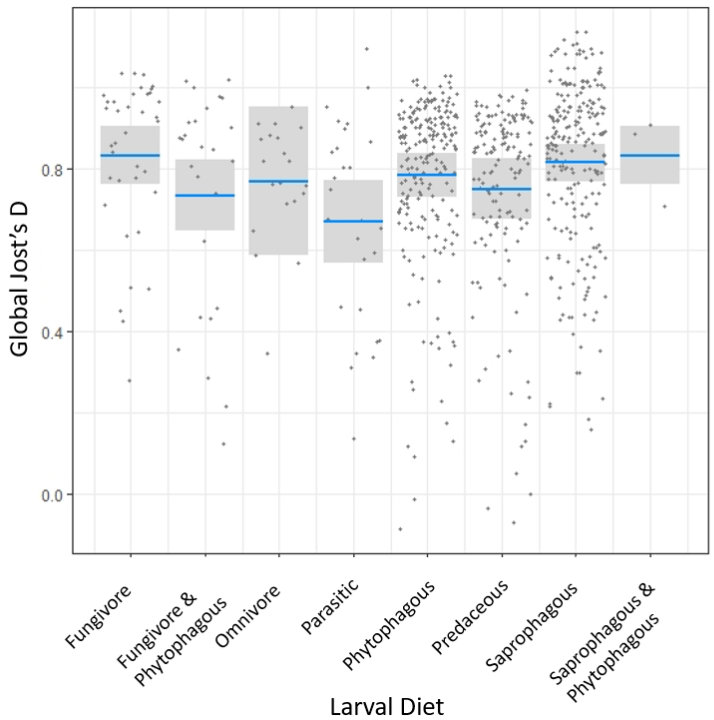


Figure 12: Plots showing the relationships between the global population genetic structure measures and habitat for the Canadian dataset of 716 Diptera BINs. Each dot represents a BIN. Only the significant relationships from the multiple regression are shown (Table 3). The relationship between habitat and i) Jost'D and ii) F_{ST} are shown. Jost's D and G_{ST} suggested that aquatic BINs show significantly higher levels of genetic differentiation, while terrestrial show significantly lower levels. F_{ST} suggests the opposite pattern.

a) Jost's D Larval Diet



b) F_{ST} Larval Diet

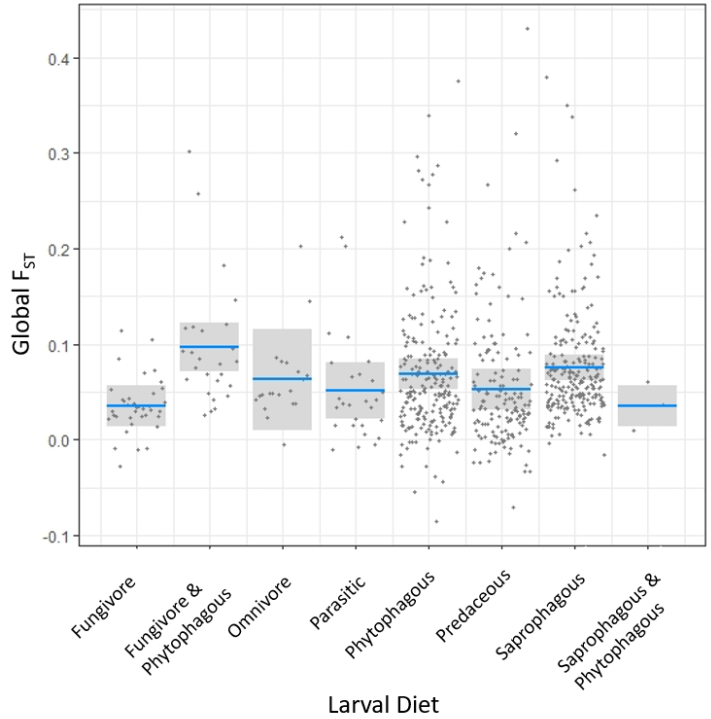
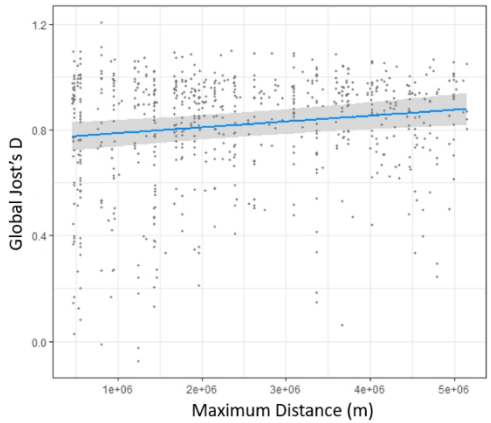
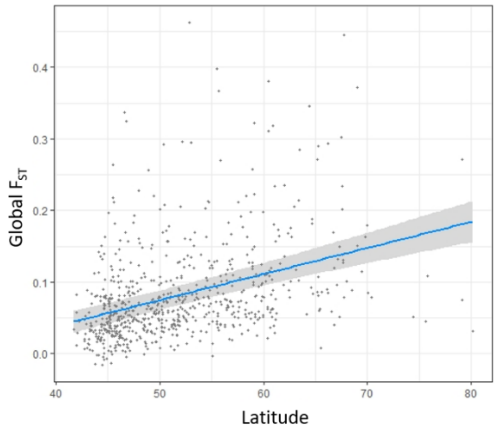


Figure 13: Plots showing the relationships between the global population genetic structure measures and larval diet for the Canadian dataset of 716 Diptera BINs. Each dot represents a BIN. Only the significant relationships from the multiple regression are shown (Table 3). The relationship between larval diet and i) Jost's D and ii) F_{ST} are shown. Jost's D and G_{ST} suggested significantly lower levels of genetic differentiation in BINs with a parasitic larval stage. F_{ST} suggested significantly higher levels of genetic differentiation in BINs with a phytophagous larval stage, a saprophagous larval stage, and those whose larval stage was a mix of fungivore and phytophagous.

a) Jost's D Max Distance



b) F_{ST} Latitude



c) F_{ST} Longitude

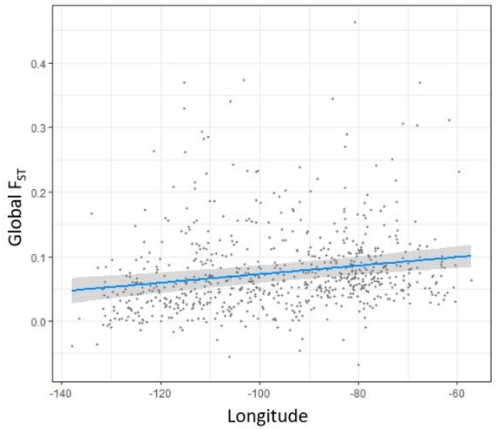


Figure 14: Plots showing the relationships between the global population genetic structure measures and geographic factors in the Canadian dataset of 716 Diptera BINs. Each dot represents a BIN. Only the significant relationships from the multiple regression are shown (Table 3). The relationship between indicators of population genetic structure and a) distance, b) latitude and c) longitude are shown. For all measures, the greater the distance, the greater the genetic differentiation. F_{ST} suggests that as latitude and longitude increase, so does genetic differentiation.

REFERENCES

Abuelmaali, S.A., Jamaluddin, J.A.F., Noaman, K., Allam, M., Abushama, H.M., Elnaiem, D.E., Ishak, I.H., Wajidi, M.F.F., Jaal, Z. & Abu Kassim, N.F. (2021). Distribution and genetic diversity of *Aedes aegypti* subspecies across the Sahelian Belt in Sudan. *Pathogens, 10*(1), 78. doi: 10.3390/pathogens10010078

Akalin, A. (2018, Oct. 1). *Scientific Data Analysis Pipelines and Reproducibility.* Toward Data Science. https://towardsdatascience.com/scientific-data-analysis-pipelines-and-reproducibility-75ff9df5b4c5

April, J., Hanner, R.H., Dion- Côté, A. & Bernatchez, L. (2012). Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology, 22*(2), 409-422. doi: 10.1111/mec.12116

Attiná, N., Núñez Bustos, E.O., Lijtmaer, D.A., Hebert, P.D.N., Tubaro, P.L. & Lavinia, P.D. (2021). Genetic variation in Neotropical butterflies is associated with sampling scale, species distributions, and historical forest dynamics. *Molecular Ecology Resources.* doi: 10.1111/1755-0998.13441

Baloğlu, B., Clews, E. & Meier, R. (2018). NGS barcoding reveals high resistance of a hyperdiverse chironomid (Diptera) swamp fauna against invasion from adjacent freshwater reservoirs. *Frontiers in Zoology, 15*(1), 31. doi: 10.1186/s12983-018-0276-7

Bartolini, I., Rivera, J., Nolazco, N. & Olortegui, A. (2020). Towards the implementation of a DNA barcode library for the identification of Peruvian species of *Anastrepha* (Diptera: Tephritidae). *PLoS ONE, 15*(1), p.e0228136. doi: 10.1371/journal.pone.0228136

Barnes, R. & Sahr, K. (2017). dggridR: Discrete global grids for R. R package version 2.0.4. https://github.com/r-barnes/dggridR/ doi: 10.5281/zenodo.1322866

Bernatchez, L. & Wilson, C.C. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology, 7*(4), 431-452. doi: 10.1046/j.1365-294x.1998.00319.x

Boyle, E.E. & Adamowicz, S.J. (2015). Community phylogenetics: assessing tree reconstruction methods and the utility of DNA barcodes. *PLoS ONE, 10*(6), p.e0126662. doi: 10.1371/journal.pone.0126662

Carbonell, J. A., Millán, A., Green, A. J., Céspedes, V., Coccia, C., & Velasco, J. (2016).

What traits underpin the successful establishment and spread of the invasive water bug *Trichocorixa verticalis verticalis*? *Hydrobiologia*, *768*(1), 273–286. doi: 10.1007/s10750-015-2556-y

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics, 11*(4), 265-270. doi:

Choi, E.H., Kim, G., Cha, S.H., Lee, J., Ryu, S.H., Suk, H.Y., Lee, Y.S., Baek, S.Y. & Hwang, U.W. (2020). Molecular phylogenetic, population genetic and demographic studies of *Nodularia douglasiae* and *Nodularia breviconcha* based on CO1 and 16S rRNA. *Scientific Reports, 10*(1), 16572. doi: 10.1038/s41598-020-72015-5

Dalapicolla, J. (2021). LanGen_pipeline_version2, Github repository, https://github.com/jdalapicolla/LanGen_pipeline_version2

Delettre, Y.R. & Morvan, N. (2000). Dispersal of adult Chironomidae (Diptera) in agricultural landscapes. *Freshwater Biology, 44*(3), 399-411. doi: 10.1046/j.1365-2427.2000.00578.x

Denton, R. D., Greenwald, K. R., & Gibbs, H. L. (2017). Locomotor endurance predicts differences in realized dispersal between sympatric sexual and unisexual salamanders. *Functional Ecology*, *31*(4), 915–926. doi:10.1111/1365-2435.12813

Descamps, S., Aars, J., Fuglei, E., Kovacs, K. M., Lydersen, C., Pavlova, O., Pedersen, Å., Ravolainen, V., & Strøm, H. (2017). Climate change impacts on wildlife in a High Arctic archipelago – Svalbard, Norway. *Global Change Biology*, *23*(2), 490–502. doi:10.1111/gcb.13381

deWaard, J.R., Levesque-Beaudin, V., deWaard, S.L., Ivanova, N.V., McKeown, J.T.A., Miskie, R., Naik, S., Perez, K.H.J., Ratnasingham, S., Sobel, C.N., Sones, J.E., Steinke, C., Telfer, A.C., Young, A.D., Young, M.R., Zakharov, E.V. & Hebert, P.D.N. (2018). Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome, 62*(3), 85-95. doi: 10.1139/gen-2018-0093

Downes, B. J., & Lancaster, J. (2018). Itinerant, nomad or invader? A field experiment sheds light on the characteristics of successful dispersers and colonists. *Freshwater Biology*, *63*(11), 1394–1406. doi:10.1111/fwb.13167

Fobert, E. K., Treml, E. A., & Swearer, S. E. (2019). Dispersal and population connectivity are phenotype dependent in a marine metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1909). doi:10.1098/rspb.2019.1104

Froufe, E., Sobral, C., Teixeira, A., Sousa, R., Varandas, S., Aldridge, D.C. & Lopes-Lima, M. (2014). Genetic diversity of the pan-European freshwater mussel *Anodonta anatina* (Bivalvia: Unionoida) based on CO1: new phylogenetic insights and implications for conservations. *Aquatic Conservation, 24*(4), 561-574. doi: 10.1002/aqc.2456

GBIF.org (2020). *GBIF Home Page.* Avaliable from https://www.gbif.org. [4 November 2020]

Geiger, M.F., Moriniere, J., Hausmann, A., Haszprunar, G., Wägele, W., Hebert, P.D.N. & Rulik, B. (2016). Testing the Global Malaise Trap Pogram-How well does the current barcode reference library identify flying insects in Germany? *Biodiversity Data Journal, 4*(4), e10671. doi: 10.3897/BDJ.4.e10671

Github. (2022). *Github.* https://github.com/

Global Malaise Trap Program (n.d.). *Global Malaise Trap Program.* Retrieved from http://globalmalaise.org/homecopy/

Gojkovic, N., Francuski, L., Ludoski, J. & Milankov, V. (2020). DNA barcode assessment and population genetic structure of aphidophagous hoverfly *Sphaerophoria scripta*: Implications for conversation biological control. *Ecology and Evolution, 10*(17), 9428-9443. doi: 10.1002/ece3.6631

Grant, W.S. (2015). Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *The Journal of Heredity, 106*(4), 333-346. doi: 10.1093/jhered/esv020

Haseyama, K.L.F., Wiegmann, B.M., Almeida, E.A.B. & de Carvalho, C.J.B. (2015). Say goodbye to tribes in the new house fly classification: A new molecular phylogenetic analysis and an updated biogeographical narrative for the Muscidae (Diptera). *Molecular Phylogenetics and Evolution, 89,* 1-12. doi: 10.1016/j.ympev.2015.04.006

Havel, J.E., Colbourne, J.K. & Hebert, P.D.N. (2000). Reconstructing the history of intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers. *Limnology and oceanography, 45*(6), 1414-1419. doi: 10.4319/lo.2000.45.6.1414

Hebert, P.D.N., Ratnasingham, S., Zakharov, E.V., Telfer, A.C., Levesque-Beaudin, V., Milton, M.A., Pedersen, S., Jannetta, P. & deWaard, J.R. (2016). Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society Biological Sciences, 371*(1702), 20150333. doi: 10.1098/rstb.2015.0333

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences, 270*(1512), 313– 21. doi: 10.1098/rspb.2002.2218

Hedrick, P.W. A standardized genetic differentiation measure. *Evolution, 59*(8), 1633-1638. doi: 10.1111/j.0014-3820.2005.tb01814.x

Hernández-Triana, L.M., Montes De Oca, F., Prosser, S.W.J., Hebert, P.D.N., Gregory, T.R. & McMurtrie, S. (2017). DNA barcoding as an aid for species identification in austral black flies (Insecta: Diptera: Simuliidae). *Genome, 60*(4), 348-357. doi: 10.1139/gen-2015-0168

Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-4026. doi:10.1111/j.1365-294X.2008.03887.x

Jossart, Q., De Ridder, C., Lessios, H.A., Bauwens, M., Motreuil, S., Rigaud, T., Wattier, R.A. & David, B. (2017). Highly contrasted population genetic structures in a host-parasite pair in the Caribbean Sea. *Ecology and Evolution, 7*(22), 9267-9280. doi: 10.1002/ece3.3413

Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, 281. doi: 10.7717/peerj.281

Keogh, S.M., Johnson, N.A., Williams, J.D., Randklev, C.R. & Simons, A.M. (2021). Gulf Coast vicariance shapes phylogeographic history of a North American freshwater mussel species complex. *Journal of Biogeography, 48*(5), 1138-1152. doi: 10.1111/jbi.14066

Kirmer, A., Tischew, S., Ozinga, W.A., Von Lampe, M., Baasch, A. & Van Groenendael, J.M. (2008). Importance of regional species pools and functional traits in colonization processes: predicting re-colonization after large-scale destruction of ecosystems. *Journal of Applied Ecology*, *45*(5), 1523-1530. doi: 10.1111/j.1365-2664.2008.01529.x

Krosch, M.N., Baker, A.M., Mather, P.B. & Cranston, P.S. (2011). Spatial population genetic structure reveals strong natal site fidelity in *Echinocladius martini* (Diptera: Chironomidae) in northeast Queensland, Australia. *Freshwater Biology, 56*(7), 1328-1341. doi: 10.1111/j.1365-2427.2010.02571.x

Lagger, C., Nime, M., Torre, L., Servetto, N., Tatián, M., & Sahade, R. (2018). Climate change, glacier retreat and a new ice-free island offer new insights on Antarctic benthic responses. *Ecography*, *41*(4), 579–591. doi:10.1111/ecog.03018

Liew, C., Atkinson, M., Galea, M., Ang, T., Martin, P. & Hemert, J. (2017). Scientific workflows: moving across paradigms. *ACM Computing Surveys, 49*(4), 1-39. doi: 10.1145/3012429

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

Liu, B., Zhang, K., Zhu, K., Shafi, M., Gong, L., Jiang, L., Liu, L., Muhammad, F. & Lu, Z. (2020). Population genetics of *Konosirus punctatus* in Chinese coastal waters inferred from two mtDNA genes (CO1 and Cytb). *Frontiers in Marine Science, 7*. doi: 10.3389/fmars.2020.00534

Maddison, W.P. & Maddison, D.R. (2019). Mesquite: a modular system for evolutionary analysis. Version 3.61. http://www.mesquiteproject.org

Marshall, S.A. (2012). *Flies. The natural history and diversity of Diptera.* Firefly Books Ltd.

Maxwell, S.A., Thistlewood, H.M.A. & Keyghobadi, N. (2013). Population genetic structure of the western cherry fruit fly *Rhagoletis indifferens* (Diptera: Tephritidae) in British Columbia, Canada. *Agricultural and Forest Entomology, 16*(1), 33-44. doi: 10.1111/afe.12029

Meriam, T., Wafa, T., Khawla, T., Tarek, H., Abdeljelil, G. & Mhamed, E. (2015). Genetic diversity and population structure of Sepia officinalis from the Tunisian cost revealed by mitochondrial CO1 sequences. *Molecular Biology Reports, 42*(1), 77-86. doi: 10.1007/s11033-014-3743-z

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Kjer, K.M. & Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science, 346*(6210), 763-767. doi: 10.1126/science.1257570

Nei, M. (1972). Genetic distance between populations. *The American Naturalist,*

106(949), 283-292. doi: 10.1086/282771

Nugent, C.M., Elliott, T.A., Ratnasingham, S., Adamowicz, S.J. (2020). Coil: an R package for cytochrome c oxidase I (CO1) DNA barcode data cleaning, translation, and error evaluation. *Genome, 63*(6), 291-305. doi: 10.1139/gen-2019-0206

Olival, K.J., Dick, C.W., Simmons, N.b., Morales, J.C., Melnick, D.J., Dittmar, K., Perkins, S.L., Daszak, P. & Desalle, R. Lack of population genetic structure and host specificity in the bat fly, Cyclopodia horsfieldi, across species of Pteropus bats in Southeast Asia. *Parasites & Vectors, 6*(1), 231. doi: 10.1186/1756-3305-6-231

Orton, M.G., May, J.A., Ly, W., Lee, D.J. & Adamowicz, S.J. (2019). Is molecular evolution faster in the tropics? *Heredity, 122*(5), 513-524. doi: 10.1038/s41437-018-0141-7

Öster, M., Ask, K., Cousins, S. A. O., & Eriksson, O. (2009). Dispersal and establishment limitation reduces the potential for successful restoration of semi-natural grassland communities on former arable fields. *Journal of Applied Ecology*, *46*(6), 1266–1274. doi:10.1111/j.1365-2664.2009.01721.x

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics, 26*(3), 419-420. doi: 10.1093/bioinformatics/btp696

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics, 2*(22), 289-290. doi:10.1093/bioinformatics/btg412

Park, S., Noh, P., Choi, Y., Joo, S., Jeong, G. & Kim, S. (2019). Population genetic structure based on mitochondrial DNA analysis of Ikonnikov's whiskered bat (*Myotis ikonnikovi* – Chiroptera: Vespertilionidae) from Korea. *Journal of Ecology and Environment, 43*(1). doi: 10.1186/s41610-019-0140-5

Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal,10*(1), 439-446. doi: 10.32614/RJ-2018-009

Pentinsaari, M., Hebert, P.D.N. & Mutanen, M. (2014). Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLoS One, 9*(9), e108651. doi:10.1371/journal.pone.0108651

Pentinsaari, M., Vos, R. & Mutanen, M. (2016). Algorithmic single-locus species delimitation: effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles*. Molecular Ecology Resources, 17*(3), 393-404. doi:10.1111/1755-0998.12557

Pickett, T. & David, A.A. (2018). Global connectivity patterns of the notoriously invasive mussel, *Mytilus galloprovincialis* Lmk using archived CO1 sequence data. *BMC Research Notes, 11*(1), 231. doi: 10.1186/s13104-018-3328-3.

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statisitcal Computing, Vienna, Austria. URL http://www.R-project.org/

Ratnasingham, S. & Hebert, P.D.N. (2007). Bold: the barcode of life data system. *Molecular Ecology Notes, 7*(3), 355-364. doi:10.1111/j.1471-8286.2007.01678.

Ratnasingham, S. & Hebert, P.D.N. (2013). A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE, 8*(7), e66213.

Reed, L.K., Nyboer, M. & Markow, T.A. (2006). Evolutionary relationships of Drosophila mojavensis geographic host races and their sister species *Drosophila arizonae*. *Molecular Ecology, 16*(5), 1007-1022. doi: 10.1111/j.1365-294X.2006.02941.x

Reiter, T., Brooks, P.T., Irber, L., Joslin, S.E.K., Reid, C.M., Scott, C., Brown, C.T. & Pierce-Ward, N.T. (2021). Streamlining data-intensive biology with workflow systems. *Gigascience, 10*(1). doi: 10.1093/gigascience/giaa140

Rogers, A.R. & Harpending, H. (1992). Population-growth makes waves in the distribution of pairwise genetic-differences. *Molecular Biology and Evolution, 9*(3), 552-569. doi: 10.1093/oxfordjournals.molbev.a040727

Saitou, N. & Nei, M. (1987). The neighbour joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406. doi: 10.1093/oxfordjournals.molbev.a040454

Schliep K.P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics, 27*(4), 592-593. doi: 10.1093/bioinformatics/btq706

Shade, A. & Teal, T.K. (2015). Computing workflows for biologists: a roadmap. *PLoS Biology, 13*(11), p.e.1002303-e1002303. doi: 10.1371/journal.pbio.1002303

Shum, P. & Palumbi, S.R. (2021). Testing small-scale ecological gradients and intraspecific differentation for hundreds of kelp forest species using haplotypes from metabarcoding. *Molecular Ecology,* 1-19. doi: 10.1111/mec.15851

Smith, D.R. (2018). Bringing bioinformatics to the scientific masses. *EMBO Reports, 19*(6), e46262. doi: 10.15252/embr.201846262

Stark, T.E., Simoes, N. & Daly, M. (2021). Phylogeography and genetic diversity of the commercially-collected Caribbean blue-legged hermit crab (Clibanarius tricolor). *Conservation Genetics, 22,* 465-482. doi: 10.1007/s10592-021-01348-z

Stevens, V. M., Turlure, C., & Baguette, M. (2010). A meta-analysis of dispersal in butterflies. *Biological Reviews*, *85*(3), 625–642. doi:10.1111/j.1469-185X.2009.00119.x

Steyn, V. M., Mitchell, K. A., & Terblanche, J. S. (2016). Dispersal propensity, but not flight performance, explains variation in dispersal ability. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1836). doi:10.1098/rspb.2016.0905

Talbot, B., Vonhof, M.J., Broders, H.G., Fenton, B. & Keyghobadi, N. (2016). Range-wide genetic structure and demographic history in the bat ectoparasite *Cimex adjunctus*. *BMC Evolutionary Biology, 16*(1), 268-268. doi: 10.1186/s12862-016-0839-1

Tape, K. D., Jones, B. M., Arp, C. D., Nitze, I., & Grosse, G. (2018). Tundra be dammed: Beaver colonization of the Arctic. *Global Change Biology*, *24*(10), 4478–4488. doi:10.1111/gcb.14332

Troast, D., Suhling, F., Jinguji, H., Sahlén, G. & Ware, J. (2016). A global population genetic study of *Pantala flavescens*. *PLoS ONE, 11*(3), p.e0148949-e0148949. doi: 10.1371/journal.pone.0148949

Verity, R. & Nichols, R.A. (2015). What is genetic differentiation, and how should we measure it – GST, D, neither or both? *Molecular Ecology, 23*(17), 4216-4225. doi: 10.1111/mec.12856

Walseng, B., Jensen, T., Dimante-Deimantovica, I.,Christoffersen, K.S., Chertoprud, M., Chertoprud, E., Novichkova, A. & Hessen, D.O. (2018). Freshwater diversity in Svalbard: providing baseline data for ecosystems in change. *Polar Biology*, *41*(10), 1995-8342005. doi: 10.1007/s00300-018-2340-3

Węsławski, J. M., Kendall, M. A., Włodarska-Kowalczuk, M., Iken, K., Kędra, M., Legezynska, J., & Sejr, M. K. (2011). Climate change effects on Arctic fjord and coastal macrobenthic diversity-observations and predictions. *Marine Biodiversity*, *41*(1), 71–85.

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Pape, T., Sinclair, B.J., Skevington, J.H., Blagoderov, V., Caravas, J., Kutty, S.N., Schmidt-Ott, U., Kampmeier, G.E., Thompson, F.C., Grimaldi, D.A., Beckenbach, A.T., Courtney, G.W., Friedrich, M., Meier, R. & Yeates, D.K. (2011). Episodic radiations in the fly tree of life. *PNAS, 108*(14), 5690-5695. doi: 10.1073/pnas.1012675108

Wilson, J. (2011). Assessing the value of DNA barcodes for molecular phylogenetics: Effect of increased taxon sampling in Lepidoptera. *PLoS One*, *6*(9), e24769. https://doi.org/10.1371/journal.pone.0024769

Winter (2012). mmod: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources, 12*(6), 1158-1160. doi: 10.1111/j.1755-0998.2012.03174.x

Witt, J.D.S. & Hebert, P.D.N. (2000). Cryptic species diversity and evolution in the amphipod genus Hyalella within central glaciated North America: a molecular phylogenetic approach. *Canadian Journal of Fisheries and Aquatic Sciences, 57*(4), 687-698. doi: 10.1139/f99-285

Wirta, H., Várkonyi,, G., Rasmussen, C., Kaartinen, R., Schmidt, N.M., Hebert, P.D.N., Barták, M., Blagoev, G., Disney, H., Ertl, S., Gjelstrup, P., Gwiazdowicz, D.J., Huldén, L., Ilmonen, J., Jakovlev, J., Jaschhof, M., Kahanpää, J., Kankaanpää, T., Krogh, P.H., Labbee, R., Lettner, C., Michelsen, V., Nielsen, S.A., Nielsen, T.R., Paasivirta, L., Pedersen, S., Pohjoismäki, J., Salmela, J., Vilkamaa, P., Väre, H., von Tschirnhaus, M., Roslin, T. (2015). Establishing a community-wide DNA barcode library as a new tool for arctic research. *Molecular Ecology Resources, 16*(3), 809-822. doi: 10.1111/1755-0998.12489

Wright, E.S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *The R Journal, 8*(1), 352-359. Doi: 10.32614/RJ-2016-025

Xu, Y., Mai, J., Yu, B., Hu, H., Yuan, L., Jashenko, R. & Ji, R. (2019). Study on the genetic differentiation of geographic populations of *Calliptamus italicus* (Orthoptera: Acrididae) in Sino-Kazakh border areas based on mitochondrial COI and COII genes. *Journal of Economic Entomology, 112*(4), 1912-1919. doi: 10.1093/jee/toz112

Yu, H., Lin, X., Wang, Q. & Wang, X. (2020). Species delimitation and life stage association of *Propsilocerus* Kieffer, 1923 (Diptera, Chironomidae) using DNA barcodes. *ZooKeys, 975*(3), 79-86. doi: 10.3897/zookeys.957.54668

APPENDICES

Appendix A: References for Diptera habitat and diet traits

Acartophthalmidae

Roháček, J. (2012). The fauna of the opomyzoid families Clusiidae, Acartophthalmidae, Anthomyzidae, Opomyzidae, Stenomicridae, Periscelididae, Asteiidae (Diptera) in the Gemer area (Central Slovakia). *Časopis Slezského Zemského Muzea, 61*(2), 97-111. doi: 10.2478/v10210-012-0011-5

Roháček, J. (2013). The fauna of the Acalyptrate families Micropezidae, Psilidae, Clusiidae, Acartophthalmidae, Anthomyzidae, Aulacigastridae, Periscelididae and Asteiidae (Diptera) in the Gemer area (Central Slovakia): supplement 1. C*asopis Slezskeho Zemskeho Muzea, 62*(2), 125-136. doi: 10.2478/cszma-2013-0014

Agromyzidae

Ayabe, Y., Ueno, T. & Harvey, J.A. (2012). Complex feeding tracks of the sessile herbivorous insect *Ophiomyia maura* as a function of the defense against insect parasitoids. *PLoS one, 7*(2), p.e32594-e32594. doi: 10.1371/journal.pone.0032594

Bañuelos, M. & Kollmann, J. (2011). Effects of host-plant population size and plant sex on a specialist leaf-miner. *Acta Oecologica, 37*(2), 58-64. doi: 10.1016/j.actao.2010.11.007

Bragard, C., Dehnen-Schmutz, K., Di Serio, F., Gonthier, P., Jacques, M., Jaques Miret, J..A., Justesen, A.F., Magnusson, C.S., Milonas, P., Navas-Cortes, J.A., Parnell, S., Potting, R., Reignault, P.L., Thulke, H., Van der Werf, W., Vicent Civera, A., Yuen, J., Zappalà, L., Czwienczek, E., Streissl, F. & MacLeod, A. (2020). Pest categorisation of *Liriomyza bryoniae. EFSA Journal, 18* (3), p.e06038-n/a. doi: 10.2903/j.efsa.2020.6038

Carapelli, A., Soltani, A., Leo, C., Vitale, M., Amri, M. & Mediouni-Ben Jemâa, J. (2018). Cryptic Diversity Hidden within the Leafminer Genus *Liriomyza* (Diptera: Agromyzidae). *Genes, 9*(11), 554. doi: 10.3390/genes9110554

Eber, S. (2004). Bottom-up density regulation in the holly leaf-miner *Phytomyza ilicis. The Journal of Animal Ecology, 73*(5), 948-958. doi: 10.1111/j.0021-8790.2004.00867.x

Gratton, C. & Welter, S.C. (1998). Oviposition preference and larval performance of *Liriomyza helianthi* (Diptera: Agromyzidae) on normal and novel host plants. *Environmental Entomology, 27*(4), 926-935. doi: 10.1093/ee/27.4.926

Hébert, J.B., Scheffer, S.J. & Hawthorne, D.J. (2016). Evidence for ecological speciation via a host shift in the holly leaf miner, *Phytomyza glabricola* (Diptera: Agromyzidae). *Ecology and Evolution, 6*(18), 6565-6577. doi: 10.1002/ece3.2358

Lambkin, C.L., Fayed, S.A., Manchester, C., La Salle, J., Scheffer, S.J. & Yeates, D.K. (2008). Plant hosts and parasitoid associations of leaf mining flies (Diptera: Agromyzidae) in the Canberra region of Australia. *Australian Journal of Entomology, 47*(1), 13-19. doi: 10.1111/j.1440-6055.2007.00622.x

Lonsdale, O. (2015). A new afrotropical species of *Phytomyza* in the subgenus *Ptochomyza* (Diptera: Agromyzidae). *African Invertebrates, 56*(3), 637-643. doi: 10.5733/afin.056.0309

Lonsdale, O. (2021). Manual of North American Agromyzidae (Diptera, Schizophora), with revision of the fauna of the “Delmarva” states. *ZooKeys, 1051*, 1-481. doi: 10.3897/zookeys.1051.64603

Marino, P.C. & Cornell, H.V. (1992). Adult movement of the native holly leafminer, *Phytomyza ilicicola* Loew (Diptera: Agromyzidae): Consequences for host choice within and between habitats. *Oecologia, 92*(1), 76-82. doi: 10.1007/BF00317265

Marino, P.C., Cornell, H.V. & Kahn, D.H. (1993). Environmental and clonal influences on host choice and larval survival in a leafmining insect. *Journal of Animal Ecology, 62*(3), 503-510. doi: 10.2307/5199

Masetti, A., Lanzoni, A. & Burgio, G. (2010). Effects of flowering plants on parasitism of lettuce leafminers (Diptera: Agromyzidae). *Biological Control, 54*(3), 263-269. doi: 10.1016/j.biocontrol.2010.05.016

Nzama, S., Olckers, T. & Zachariades, C. (2014). Seasonal activity, habitat preferences and larval mortality of the leaf-mining fly *Calycomyza eupatorivora* (Agromyzidae), a biological control agent established on *Chromolaena odorata* (Asteraceae) in South Africa. *Biocontrol Science and Technology, 24*(11), 1297-1307. Doi: 10.1080/09583157.2014.935293

Okoth, C.A., Deng, A.L., Tabu, I.M., Akutse, K.S. & Fiaboe, K.K.M. (2014). Effect of host plant on feeding, biological and morphological parameters of *Liriomyza huidobrensis* Blanchard (Diptera: Agromyzidae). *African Entomology, 22*(3), 577-588. doi: 10.4001/003.022.0315

Piwowarczyk, R., Mielczarek, L. & Guzikowski, S. (2018). First report of *Phytomyza orobanchia* (Diptera: Agromyzidae) from Poland and *Chymomyza amoena* (Diptera: Drosophilidae) on *Phelipanche ramose* (Orobanchaceae). *Florida Entomologist, 101*(3), 540-542. doi: 10.1653/024.101.0333

Roíguez-Castañeda, G., MacVean, C., Cardona, C. & Hof, A.R. (2017). What limits the distribution of *Liriomyza huidobrensis* and its congener *Liriomyza sativae* in their native niche: when temperature and competition affect species' distribution range in Guatemala. *Journal of Insect Science, 17*(4). doi: 10.1093/jisesa/iex059

Scheffer, S.J. & Hawthorne, D.J. (2007). Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): apparent discordance among marker systems. *Molecular Ecology, 16*(13), 2627-2637. doi: 10.1111/j.1365-294X.2007.03303.x

Scheffer, S.J., Lewis, M.L., Hébert, J.B. & Jacobsen, F. (2020). Diversity and host plant-use North American *Phytomyza* holly leafminers (Diptera: Agromyzidae): colonization, divergence, and specificity in a host-associated radiation. *Annals of the Entomological Society of America, 114*(1), 56-69. doi: 10.1093/aesa/saaa03

Scheffer, S.J. & Lonsdale, O. (2011). *Phytomyza omlandi* spec. nov. – The first species of Agromyzidae (Diptera: Schizophora) reared from the family Gelsemiaceae (Asteridae). *Proceedings of the Entomological Society of Washington, 113*(1), 42-49. doi: 10.4289/0013-8797.113.1.42

Scheffer, S.J. & Wiegmann, B.M. (2000). Molecular phylogenetics of the holly leafminers (Diptera: Agromyzidae: *Phytomyza*): species limits, speciation, and dietary specialization. *Molecular Phylogenetics and Evolution, 17*(2), 244-255. doi: 10.1006/mpev.2000.0830

Soltani, A., Ben Abda, M., Amri, M., Carapelli, A. & Mediouni Ben Jemâa, J. (2020). Seasonal incidence of the leaf miner *Liriomyza cicerina* Rond (Diptera: Agromyzidae) in chickpea fields and effects of climatic parameters, chickpea variety, and planting date on the leaf miner infestation rate. *Euro-Mediterranean journal for environmental integration, 5*(3). doi: 10.1007/s41207-020-00198-4

Williams, E. C. & Macdonald, O. C. (1995). Critical factors required by the nematode *Steinernema feltiae* for the control of the leafminers *Liriomyza huidobrensis*, *Liriomyza bryoniae* and *Chromatomyia syngenesiae*. *Annals of Applied Biology, 127*(2), 329-341. doi: 10.1111/j.1744-7348.1995.tb06677.x

<u>Anisopodidae</u>

Kolcsár, L., Dvořák, L. & Beuk, P.L. (2016). New records of *Sylvicola* (Diptera: Anisopodidae) from Romania. *Biodiversity Data Journal, 4*(4), p.e7861-e7861. doi: 10.3897/BDJ.4.e7861.

Ševčík, J. (2011). Anisopodidae and Dixidae (Diptera) of the Gemer region, with a new record for Slovakia. *Časopis Slezského Zemského Muzea, 60*(2), 181-184. doi: 10.2478/v10210-011-0019-2

Soli, G. & Rindal, E. (2014). The genus Sylvicola Harris, 1780 (Diptera, Anisopodidae) in Norway – with a key to the North European species. *Norwegian Journal of Entomology, 61*(2), 190-200.

Anthomyiidae

Bažok, R., Ceranić-Sertić, M., Barčić, J.I., Borošić, J., Kozina, A., Kos, T., Lemić, D. & Čačija, M. (2012). Seasonal flight, optimal timing and efficacy of selected insecticides for cabbage maggot (*Delia radicum* L., Diptera: Anthomyiidae) control. *Insects, 3*(4), 1001-1027. doi: 10.3390/insects3041001

Deasy, W., Shepherd, T., Alexander, C.J., Birch, A.N.E. & Evans, K.A. (2016). Field-based evaluation of a novel SPME-GC-MS method for investigation of below-ground interaction between brassica roots and larvae of cabbage root fly, *Delia radicum* L. *Phytochemical Analysis, 27*(6), 343-353. doi: 10.1002/pca.2634

Dreves, A.J., Dalthorp, D., Stone, A.G. & Fisher, G. (2006). Spring emergence and seasonal flight of *Delia radicum* L. (Diptera: Anthomyiidae) in western Oregon. *Environmental Entomology, 35*(2), 465-477. doi: 10.1603/0046-225X-35.2.465

Gouinguene, S.P & Stadler, E. (2006). Oviposition in *Delia platura* (Diptera, Anthomyiidae): The role of volatile and contact cues of bean. *Journal of Chemical Ecology, 32*(7), 1399-1413. doi: 10.1007/s10886-006-9058-3

Guerra, P.C., Keil, C.B., Stevenson, P.C., Mina, D., Samaniego, S., Peralta, E., Mazon, N. & Chancellor, T.C.B. (2016). Larval performance and adult attraction of *Delia platura* (Diptera: Anthomyiidae) in a native and an introduced crop. *Journal of Economic Entomology, 110*(1), 186-191. doi: 10.1093/jee/tow237

Kim, H.S., Cho, J.R., Kim, J.J., Lee, M. & Byun, M.W. (2001). Optimal radiation dose of Cobalt60 to improve the sterile insect technique for *Delia antiqua* and *Delia platura. Journal of Asia-Pacific Entomology, 4*(1), 11-16. doi: 10.1016/S1226-8615(08)60095-3

Lamb, R.J. & Boivin, G. (2018). Population variability of three Delia species (Diptera: Anthomyiidae) from the same agricultural habitat in Quebec, Canada. *Canadian Entomologist, 150*(1), 80-96. doi: 10.4039/tce.2017.52

Leuchtmann, A. (2007). Botanophila flies on Epichloë host species in Europe and North America: no evidence for co-evolution. *Entomologia Experimentalis et Applicata, 123*(1), 13-23. doi: 10.1111/j.1570-7458.2006.00518.x

Lukwinski, A.T., Hill, J.E., Khachatourians, G.G., Hemmingsen, S.M. & Hegedus, D.D. (2006). Biochemical and taxonomic characterization of bacteria associated with the crucifer root maggot (*Delia radicum*). *Canadian Journal of Microbiology, 52*(3), 197-208. doi: 10.1139/w05-123

Mesmin, X., Vincent, M., Tricault, Y., Estorgues, V., Daniel, L., Cortesero, A., Faloya, V. & Le Ralec, A. (2019). Assessing the relationship between pest density and plant damage: a case study with the belowground herbivore *Delia radicum* (Diptera: Anthomyiidae) on broccoli. *Applied Entomology and Zoology, 54*(2), 155-165. doi: 10.1007/s13355-019-00607-3

Mlynarek, J.J., MacDonald, M., Sim, K., Hiltz, K., McDonald, M.R. & Blatt, S. (2020). Oviposition, feeding preferences and distribution of *Delia* species (Diptera: Anthomyiidae) in Eastern Canadian onions. *Insects, 11*(11), 780. doi: 10.3390/insects11110780

Nieto, D.J., Letourneau, D.K., Toyama, L., Slay, C. & Kula, R.R. (2021). Surveying populations of *Delia radicum* (Diptera: Anthomyiidae) and its associated natural enemies in organic cauliflower in California. *Biocontrol Science and Technology, 31*(1), 112-118. doi: 10.1080/09583157.2020.1845609

Nilsson, U., Rännbäck, L.M., Anderson, P. & Rämert, B. (1986). Herbivore response to habitat manipulation with floral resources: a study of the cabbage root fly. *Journal of Applied* Entomology, 136(7), 481-489. doi: 10.1111/j.1439-0418.2011.01685.x

Ohta, S., Yousuke S., Koichiro T., Yukio I. & Takashi M. (2015). Comprehensive identification of odorant-binding protein genes in the seed fly, *Delia platura* (Diptera: Anthomyiidae). *Applied Entomology and Zoology, 50*(4), 457-463. doi: 10.1007/s13355-015-0353-8

Rao, S. & Baumann, D. (2004). The interaction of a Botanophila fly species with an exotic Epichloë fungus in a cultivated grass: fungivore or mutualist? *Entomologia Experimentalis et Applicata, 112*(2), 99-105. doi: 10.1111/j.0013-8703.2004.00189.x

Rao, S., Alderman, S.C., Takeyasu, J. & Matson, B. (2005). Botanophila-Epichloe association in cultivated Festuca in Oregon: evidence of simple fungivory.

Entomologia Experimentalis et Applicata, 115(3), 427-433. doi: 10.1111/j.1570-7458.2005.00309.x

Saumure, R.A., Walde, A.D. & Wheeler, T.A. (2006). Nonpredatory fly larvae (*Delia platura*: Anthomyiidae) in a nest of a Northern map turtle (*Graptemys geographica*). *Chelonian Conservation and Biology, 5*(2), 274-275. doi: 10.2744/10718443(2006)5[274:NFLDPA]2.0.CO2

Silver, N., Hillier, K. & Blatt, S. (2018). Management of Delia (Diptera: Anthomyiidae) through selectively timed planting of *Phaseolus vulgaris* (Fabaceae) in Atlantic Canada. *Canadian Entomologist, 150*(5), 663-674. doi: 10.4039/tce.2018.36

Sontowski, R., Poeschl, Y., Okamura, Y., Vogel, H., Guyomar, C., Cortesero, A. & Dam, N.M. (2022). A high-quality functional genome assembly *of Delia radicum* L. (Diptera: Anthomyiidae) annotated from egg to adult. *Molecular Ecology Resources, 22*(5), 1954-1971. doi: 10.1111/1755-0998.13594

Van der Heyden, H., Fortier, A. & Savage, J. (2020). A HRM assay for rapid identification of members of the seedcorn maggot complex (*Delia florilega* and *D. platura*) (Diptera: Anthomyiidae) and evidence of variation in temporal patterns of larval occurrence. *Journal of Economic Entomology, 113*(6), 2920-2930. doi: 10.1093/jee/toaa230

Anthomyzidae

RohaAek, J., Petrakova, L. & Tothova, A. (2019). Molecular phylogeny and timing of evolution of *Anthomyza* and related genera (Diptera: Anthomyzidae). *Zoologica Scripta, 48*(6), 745-760. doi: 10.1111/zsc.12373

Roháček, J. (2012). The fauna of the opomyzoid families Clusiidae, Acartophthalmidae, Anthomyzidae, Opomyzidae, Stenomicridae, Periscelididae, Asteiidae (Diptera) in the Gemer area (Central Slovakia). *Časopis Slezského Zemského Muzea, 61*(2), 97-111. doi: 10.2478/v10210-012-0011-5

Asilidae

Dennis, D.S., Lavigne, R.J. & Dennis, J.G. (2012). Spiders (Araneae) as prey of robber flies (Diptera: Asilidae). *Journal of Entomological Research Society, 14*(1), 65-76.

Pollock, D.A. (2021). Robber flies (Diptera: Asilidae) as predators of harvester ant workers (Hymenoptera: Formicidae: Pogonomyrmex) in Eastern New Mexico and

West Texas. *The Southwestern Naturalist, 65*(1), 19-27. doi: 10.1894/0038-4909-65.1.3

Asteiidae

Roháček, J. (2012). The fauna of the opomyzoid families Clusiidae, Acartophthalmidae, Anthomyzidae, Opomyzidae, Stenomicridae, Periscelididae, Asteiidae (Diptera) in the Gemer area (Central Slovakia). *Časopis Slezského Zemského Muzea, 61*(2), 97-111. doi: 10.2478/v10210-012-0011-5

Aulacigastridae

Roháček, J. (2013). The fauna of the Acalyptrate families Micropezidae, Psilidae, Clusiidae, Acartophthalmidae, Anthomyzidae, Aulacigastridae, Periscelididae and Asteiidae (Diptera) in the Gemer area (Central Slovakia): supplement 1. C*asopis Slezskeho Zemskeho Muzea, 62*(2), 125-136. doi: 10.2478/cszma-2013-0014

Bibionidae

Frouz, J., Lin, Q., Li, X., Abakumov, E., Brune, A. & Šustr, V. (2020). Utilization of dietary protein in the litter-dwelling larva of *Bibio marci* (Diptera: Bibionidae). *Eurasian Soil Science, 52*(12), 1583-1587. doi: 10.1134/S1064229319120032

Šustr, V., Stingl, U. & Brune, A. (2014). Microprofiles of oxygen, redox potential, and pH, and microbial fermentation products in the highly alkaline gut of the saprophagous larva of *Penthetria holosericea* (Diptera: Bibionidae). *Journal of Insect Physiology, 67*, 64-69. doi: 10.1016/j.jinsphys.2014.06.007

Calliphoridae

Andere, A.A., Platt, R.N., Ray, D.A. & Picard, C.J. (2016). Genome sequence of *Phormia regina* Meigen (Diptera: Calliphoridae): implications for medical, veterinary and forensic research. *BMC Genomics, 17*(1), 842. doi: 10.1186/s12864-016-3187-z

Bedini, S., Flamini, G., Cosci, F., Ascrizzi, R., Echeverria, M.C., Guidi, L., Landi, M., Lucchi, A. & Conti, B. (2017). Artemisia spp. essential oils against the disease-

carrying blowfly *Calliphora vomitoria. Parasites & Vectors, 10*(1), 80. doi: 10.1186/s13071-017-2006-y

Bennett, G.F. & Whitworth, T.L. (1991). Studies on the life history of some species of *Protocalliphora* (Diptera: Calliphoridae). *Canadian Journal of Zoology, 69*(8), 2048-2058. doi: 10.1139/z91-286

Berrigan, D. & Lighton, J.R.B. (1994). Energetics of pedestrian locomotion in adult male blowflies, *Protophormia terraenovae* (Diptera: Calliphoridae). *Physiological Zoology, 67*(5), 1140-1153. doi: 10.1086/physzool.67.5.30163886

Brodie, B.S., Wong, W.H.L., VanLaerhoven, S. & Gries, G. (2015). Is aggregated oviposition by the blow flies *Lucilia sericata* and *Phormia regina* (Diptera: Calliphoridae) really pheromone-mediated? *Insect Science, 22*(5), 651-660. doi: 10.1111/1744-7917.12160

Burgess, I.F. (1987). Opportunistic myiasis of nestling common blackbirds by *Lucilia silvarum. Parasitology Research, 121*(2), 763-764. doi: 10.1007/s00436-021-07402-6

Dufek, M.I., Oscherov, E.B., Damborsky, M.P. & Mulieri, P.R. (2019). Calliphoridae (Diptera) in human-transformed and wild habitats: diversity and seasonal fluctuations in the humid Chaco Ecoregion of South America. *Journal of Medical Entomology, 56*(3), 725-736. doi: 10.1093/jme/tjy234

Dykstra, C.R., Hays, J.L., Simon, M.M. & Wegman, A.R. (2012). *Protocalliphora* (Diptera: Calliphoridae) infestations of nestling red-shouldered hawks in Southern Ohio. *The Wilson Journal of Ornithology, 124*(4), 783-787. doi: 10.1676/1559-4491-124.4.783

Garrido-Bautista, J., Moreno-Rueda, G., Baz, A., Canal, D., Camacho, C., Cifrián, B., Nieves-Aldrey, J.L., Carles-Tolrá, M. & Potti, J. (1987). Variation in parasitoidism of *Protocalliphora azurea* (Diptera: Calliphoridae) by *Nasonia vitripennis* (Hymenoptera: Pteromalidae) in Spain. *Parasitology Research, 119*(2), 559-566. doi: 10.1007/s00436-019-06553-x

Gentes, M., Whitworth, T.L., Waldner, C., Fenton, H. & Smits, J.E. (2007). Tree swallows (*Tachycineta bicolor*) nesting on wetlands impacted by oil sands mining are highly parasitized by the bird blow fly *Protocalliphora* sp. *Journal of Wildlife Diseases, 43*(2), 167-178. doi: 10.7589/0090-3558-43.2.167

Gunn, A. & Bird, J. (2010). The ability of the blowflies *Calliphora vomitoria* (Linnaeus), *Calliphora vicina* (Rob-Desvoidy) and *Lucilia sericata* (Meigen) (Diptera: Calliphoridae) and the muscid flies *Muscina stabulans* (Fallén) and *Muscina*

prolapsa (Harris) (Diptera: Muscidae) to colonise buried remains. *Forensic Science International, 207*(1), 198-204. doi: 10.1016/j.forsciint.2010.10.008

Johnson, L.S., Eastman, M.D. & Kermott, L.H. (1991). Effect of ectoparasitism by larvae of the blow fly *Protocalliphora parorum* (Diptera: Calliphoridae) on nestling house wrens, *Troglodytes aedon. Canadian Journal of Zoology, 69*(6), 1441-1446. doi: 10.1139/z91-204

Kelly, M.A., Usher, M.J., Ujvari, B., Madsen, T., Wallman, J.F., Buttemer, W.A. & Hulbert, A.J. (2014). Diet fatty acid profile, membrane composition and lifespan: An experimental study using the blowfly (*Calliphora stygia*). *Mechanisms of Ageing and Development, 138,* 15-25. doi: 10.1016/j.mad.2014.02.001

Maeda, T., Tamotsu, M., Yamaoka, R. & Ozaki, M. (2015). Effects of floral scents and their dietary experiences on the feeding preference in the blowfly, *Phormia regina. Frontiers in Integrative Neuroscience, 9*, 59. doi: 10.3389/fnint.2015.00059

Magni, P.A., Pacini, T., Pazzi, M., Vincenti, M. & Dadour, I.R. (2014). Development of a GC–MS method for methamphetamine detection in *Calliphora vomitoria* L. (Diptera: Calliphoridae). *Forensic Science International, 241*, 96-101. doi: 10.1016/j.forsciint.2014.05.004

Parker, F.D., Welch, J.B. & Matlock, R.B. Jr. (1993). Influence of habitat, season and attractant on adult behavior of the screwworm (Diptera: Calliphoridae) in a tropical dry zone in Costa Rica. *Journal of Economic Entomology, 86*(5), 1359-1375. doi: 10.1093/jee/86.5.1359

Pérez, C., Segura, N.A., Patarroyo, M.A. & Bello, F.J. (2016). Evaluating the biological cycle and reproductive and population parameters of *Calliphora vicina* (Diptera: Calliphoridae) reared on three different diets. *Journal of Medical Entomology, 53*(6), 1268-1275. doi: 10.1093/jme/tjw114

Pölkki, M., Kangassalo, K. & Rantala, M.J. (2012). Transgenerational effects of heavy metal pollution on immune defense of the blow fly *Protophormia terraenovae*. *PLoS ONE, 7*(6), p.e38832. doi: 10.1371/journal.pone.0038832

Wesolowski, T. (1987). Host–parasite interactions in natural holes: marsh tits (*Parus palustris*) and blow flies (*Protocalliphora falcozi*). *Journal of Zoology, 255*(4), 495-503. doi: 10.1017/S0952836901001571

Whitworth, T.L., Dawson, R.D., Magalon, H. & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society. B, Biological sciences, 274*(1619), 1731-1739. doi: 10.1098/rspb.2007.0062

Cecidomyiidae

Aggarwal, R., Subramanyam, S., Zhao, C., Chen, M., Harris, M.O., Stuart, J.J. & Castagnone-Sereno, P. (2014). Avirulence effector discovery in a plant galling and plant parasitic arthropod, the Hessian fly (*Mayetiola destructor*). *PLoS One, 9*(6), p.e100958. doi: 10.1371/journal.pone.0100958

Chen, M., Liu, S., Wang, H., Cheng, X., El Bouhssini, M., Whitworth, R. J. & Escriva, H. (2016). Massive shift in gene expression during transitions between developmental stages of the gall midge, *Mayetiola destructor. PLoS One, 11*(5), p.e0155616. doi: 10.1371/journal.pone.0155616

Dorchin, N., Harris, K.M. & Stireman, J.O. (2019). Phylogeny of the gall midges (Diptera, Cecidomyiidae, Cecidomyiinae): Systematics, evolution of feeding modes and diversification rates. *Molecular Phylogenetics and Evolution, 140,* 106602. doi: 10.1016/j.ympev.2019.106602

Heath, J.J. & Stireman, J.O. III (2010). Dissecting the association between a gall midge, *Asteromyia carbonifera*, and its symbiotic fungus, *Botryosphaeria dothidea, 137*(1), 36-49. doi: 10.1111/j.1570-7458.2010.01040.x

Johnson, A. J., Shukle, R. H., Chen, M.S., Srivastava, S., Subramanyam, S., Schemerhorn, B. J., Weintraub, P. G., Abdel Moniem, H. E. M., Flanders, K. L., Buntin, G. D. & Williams, C. E. (2015). Differential expression of candidate salivary effector proteins in field collections of hessian fly, *Mayetiola destructor. Insect Molecular Biology, 24*(2), 191-202. doi: 10.1111/imb.12148

Sá, C.E.M., Silveira, F.A.O., Santos, J.C., Isaias, R.M.S. & Fernandes, G.W. (2009). Anatomical and developmental aspects of leaf galls induced by *Schizomyia macrocapillata* Maia (Diptera: Cecidomyiidae) on *Bauhinia brevipes* Vogel (Fabaceae). *Revista Brasileira de Botânica, 32*(2), 319-327. doi: 10.1590/S0100-84042009000200011

Sadeghi, R., Odubiyi, S., Nikoukar, A., Schroeder, K.L. & Rashed, A. (2021). *Mayetiola destructor* (Diptera: Cecidmyiidae) host preference and survival on small grains with respect to leaf reflectance and phytohormone concentrations. *Scientific Reports, 11*(1), 4761. doi: 10.1038/s41598-021-84212-x

Stireman, J.O., Janson, E.M., Carr, T.G., Devlin, H. & Abbot, P. (2008). Evolutionary radiation of *Asteromyia carbonifera* (Diptera: Cecidomyiidae) gall morphotypes on the goldenrod *Solidago altissima* (Asteraceae). *Biological Journal of the Linnean Society, 95*(4), 840-858. doi: 10.1111/j.1095-8312.2008.01101.x

Subramanyam, S., Shreve, J.T., Nemacheck, J.A., Johnson, A.J., Schemerhorn, B., Shukle, R.H. & Williams, C.E. (2018). Modulation of nonessential amino acid biosynthetic pathways in virulent hessian fly larvae (*Mayetiola destructor),* feeding on susceptible host wheat (*Triticum aestivum*). *Journal of Insect Physiology, 105*, 54-63. doi: 10.1016/j.jinsphys.2018.01.001

Wearing, C. H., Marshall, R. R., Attfield, B. & Colhoun, C. (2013). Phenology and distribution of the apple leafcurling midge (*Dasineura mali* (Kieffer)) (Diptera: Cecidomyiidae) and its natural enemies on apples under biological and integrated pest management in Central Otago, New Zealand. *New Zealand Entomologist, 36*(2), 87-106. doi: 10.1080/00779962.2012.712887

Ceratopogonidae

Bakhoum, M.T., Fall, A.G., Fall, M., Bassene, C.K., Baldet, T., Seck, M.T., Bouyer, J. Garros, C. & Gimonneau, G. (2016). Insight on the larval habitat of Afrotropical *Culicoides Latreille* (Diptera: Ceratopogonidae) in the Niayes area of Senegal, West Africa. *Parasites & Vectors, 9*(1), 462. doi: 10.1186/s13071-016-1749-1

Barceló, C. & Miranda, M.A. (2021). Development and lifespan of *Culicoides obsoletus* s.s. (Meigen) and other livestock-associated species reared at different temperatures under laboratory conditions. *Medical and Veterinary Entomology, 35*(2), 187-201. doi: 10.1111/mve.12487

Chen, H., Chou, J., Lin, C., Wen, Y. & Wang, W. (2016). Seasonal yeast compositions in *Forcipomyia taiwana* (Diptera: Ceratopogonidae). *Journal of Asia-Pacific Entomology, 19*(2), 509-514. doi: 10.1016/j.aspen.2016.04.020

Díaz, F., Mangudo C., Gleiser, R.M. & Ronderos, M.M. (2019). Redescription of immatures of *Dasyhelea flavifrons* Guérin-Méneville (Culicomorpha: Ceratopogonidae) and new contribution to the knowledge of its larval habitats. *Anais da Academia Brasileira de Ciências, 91*(1). doi: 10.1590/0001-3765201920180047

Erram, D., Blosser, E.M. & Burkett-Cadena, N. (2019). Habitat associations of *Culicoides* species (Diptera: Ceratopogonidae) abundant on a commercial cervid farm in Florida, USA. *Parasites & Vectors, 12*(1), 367. doi: 10.1186/s13071-019-3626-1

Erram, D. & Burkett-Cadena, N. (2021). Oviposition of *Culicoides insignis* (Diptera: Ceratopogonidae) under laboratory conditions with notes on the developmental life history traits of its immature stages. *Parasites & Vectors, 14*(1), 522. doi: 10.1186/s13071-021-05025-5

Erram, D., Burkett-Cadena, N. & Foil, L. (2019). Laboratory rearing of *Culicoides stellifer* (Diptera: Ceratopogonidae), a suspected vector of orbiviruses in the United States. *Journal of Medical Entomology, 57*(1), 25-32. doi: 10.1093/jme/tjz154

Erram, D. & Zurek, L. (2017). Larval development of *Culicoides sonorensis* (Diptera: Ceratopogonidae) in mud supplemented with manure of various farm animals. *Journal of Medical Entomology, 55*(1), 43-50. doi: 10.1093/jme/tjx197

Kameke, D., Kampen, H., Wacker, A. & Werner, D. (2021). Field studies on breeding sites of *Culicoides* LATREILLE (Diptera: Ceratopogonidae) in agriculturally used and natural habitats. *Scientific Reports, 11*(1), 10007. doi: 10.1038/s41598-021-86163-9

Liu, Y., Tao, H., Yu, Y., Yue, L., Xia, W., Zheng, W., Ma, H., Liu, X. & Chen, H. (2018). Molecular differentiation and species composition of genus *Culicoides* biting midges (Diptera: Ceratopogonidae) in different habitats in southern China. *Veterinary Parasitology, 254*, 49-57. doi: 10.1016/j.vetpar.2018.02.035

Riddin, M. A., Venter, G. J., Labuschagne, K. & Villet, M. H. (2019). Bloodmeal analysis in *Culicoides* midges collected near horses, donkeys and zebras in the Eastern Cape, South Africa. *Medical and Veterinary Entomology, 33*(4), 467-475. doi: 10.1111/mve.12381

Steinke, S., Lühken, R., Balczun, C. & Kiel, E. (2016). Emergence of *Culicoides obsoletus* group species from farm-associated habitats in Germany: Emergence of *Culicoides obsoletus* group species. *Medical and Veterinary Entomology, 30*(2), 174-184. doi: 10.1111/mve.12159

Urbanek, A. & Rost-Roszkowska, M.M. (2015). Ultrastructural studies on the midgut of biting midge *Forcipomyia nigra* (Winnertz) (Diptera: Ceratopogonidae). *Micron, 69,* 25-34. doi: 10.1016/j.micron.2014.11.003

Urbanek, A., Szadziewski, R., Stepnowski, P., Boros-Majewska, J., Gabriel, I., Dawgul, M., Kamysz, W., Sosnowska, D. & Gołębiowski, M. (2012). Composition and antimicrobial activity of fatty acids detected in the hygroscopic secretion collected from the secretory setae of larvae of the biting midge *Forcipomyia nigra* (Diptera: Ceratopogonidae). *Journal of Insect Physiology*, *58*(9), 1265-1276. doi: 10.1016/j.jinsphys.2012.06.014

Van den Eynde, C., Sohier, C., Matthijs, S. & De Regge, N. (2021). Temperature and food sources influence subadult development and blood-feeding response *of Culicoides obsoletus* (sensu lato) under laboratory conditions. *Parasites & Vectors, 14*(1), 1-300. doi: 10.1186/s13071-021-04781-8

Veiga, J., Martínez-de la Puente, J., Václav, R., Figuerola, J. & Valera, F. (2018*). Culicoides paolae* and *C. circumscriptus* as potential vectors of avian haemosporidians in an arid ecosystem. *Parasites & Vectors, 11*(1), 524. doi: 10.1186/s13071-018-3098-8

Werner, D., Groschupp, S., Bauer, C. & Kampen, H. (2020). Breeding habitat preferences of major Culicoides species (Diptera: Ceratopogonidae) in Germany. *International Journal of Environmental Research and Public Health, 17*(14), 5000. Doi:

Yeh, W.-B., Lee, H.-M, Tu, W.-C, Tang, L.-C & Lee, P.-Y. (2009). Molecular differentiation and diversity of *Forcipomyia taiwana* (Diptera: Ceratopogonidae) based on the mitochondrial cytochrome oxidase II sequence. *Journal of Medical Entomology, 46*(2), 249-256. doi: 10.1603/033.046.0209

Chaoboridae

Barth, L.E., Sprules, W.G., Wells, M., Coman,M. & Prairie, Y. (2014). Seasonal changes in the diel vertical migration of *Chaoborus punctipennis* larval instars. *Canadian Journal of Fisheries and Aquatic Sciences, 71*(5), 665-674. doi: 10.1139/cjfas-2013-0440

Berendonk, T.U. & Spitze, K. (2006). Linking lakes? The population genetic structure of *Chaoborus flavicans. Journal of Plankton Research, 28*(11), 1015-1025. doi: 10.1093/plankt/fbl033

Härkönen, L., Pekcan-Hekim, Z., Hellén, N. & Horppila, J. (2013). Feeding efficiency of *Chaoborus flavicans* (Insecta, Diptera) under turbulent conditions. *Hydrobiologia, 722*(1), 9-17. doi: 10.1007/s10750-013-1670-y

Jäger, I.S., Hölker, F., Flöder, S. & Walz, N. (2011). Impact of *Chaoborus flavicans* - predation on the zooplankton in a mesotrophic lake - a three year study. *International Review of Hydrobiology, 96*(1), 191-208. doi: 10.1002/iroh.201011253

Moore, M.V. (1988). Differential use of food resources by the instars of *Chaoborus punctipennis. Freshwater Biology, 19*(2), 249-268. doi: 10.1111/j.1365-2427.1988.tb00346.x

Schroder, A. & Dam, H.G. (2013). Density- and size-dependent winter mortality and growth of late *Chaoborus flavicans* larvae. *PLoS One, 8*(10), p.e75839. doi: 10.1371/journal.pone.0075839

Chironomidae

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

Baniszewski, J., Miller, N., Kariuki, E.M., Cuda, J.P. & Weeks, E.N.I. (2020). *Cricotopus lebetis* intraspecific competition and damage to *Hydrilla. The Florida Entomologist, 103*(1), 32-37. doi: 10.1653/024.103.0405

Baniszewski, J., Weeks, E.N.I. & Cuda, J.P. (2016). *Bacillus thuringiensis* subspecies Kurstaki reduces competition by *Parapoynx diminutalis* (Lepidoptera: Crambidae) in colonies of the *Hydrilla* biological control agent *Cricotopus lebetis* (Diptera: Chironomidae). *The Florida Entomologist, 99*(4), 644-647. doi: 10.1653/024.099.0410

Belle, S., Millet, L., Gillet, F., Verneaux, V. & Magny, M. (2015). Assemblages and paleo-diet variability of subfossil Chironomidae (Diptera) from a deep lake (Lake Grand Maclu, France). *Hydrobiologia, 755*(1), 145-160. doi: 10.1007/s10750-015-2222-4

Boggero, A., Füreder, L., Lencioni, V., Simcic, T., Thaler, B., Ferrarese, U., Lotter, A. F. & Ettinger, R. (2006). Littoral chironomid communities of Alpine lakes in relation to environmental factors. *Hydrobiologia, 562*(1), 145-165. doi: 10.1007/s10750-005-1809-6

Boonsoong, B. (2016). Phoretic associations between *Nanocladius asiaticus* (Diptera, Chironomidae) and its hosts *Gestroiella* (Heteroptera, Naucoridae) and *Euphaea masoni* (Odonata, Euphaeidae) in streams in Western Thailand. *Annales de Limnologie, 52*, 163-169. doi: 10.1051/limn/2015025

Brown, N.E., Mitchell, S.C. & Garbary, D.J. (2013). Host selection by larvae of a marine insect *Halocladius variabilis*: nutritional dependency or escape from predation? *Journal of the Marine Biological Association of the United Kingdom, 93*(5), 1373-1379. doi: 10.1017/S0025315412001634

Cranston, P.S. & Krosch, M.N. (2015). DNA sequences and austral taxa indicate generic synonymy of *Paratrichocladius* Santos-Abreu with *Cricotopus* Wulp (Diptera: Chironomidae). *Systematic Entomology, 40*(4), 719-732. doi: 10.1111/syen.12130

Delettre, Y.R. (2005). Short-range spatial patterning of terrestrial Chironomidae (Insecta: Diptera) and farmland heterogeneity. *Pedobiologia, 49*(1), 15-27. Doi: 10.1016/j.pedobi.2004.06.010

Diserud, O.H., Stur, E. & Aagaard, K. (2012). How reliable are malaise traps for biomonitoring? -A bivariate species abundance model evaluation using alpine Chironomidae (Diptera). *Insect Conservation and Diversity, 6*(5), 561-571. doi: 10.1111/icad.12012

Dosdall, L.M. & Parker, D.W. (1998). First report of a symphoretic association between *Nanocladius branchicolus* Saether (Diptera: Chironomidae) and *Argia moesta* (Hagen) (Odonata: Coenagrionidae). *The American Midland Naturalist, 139*(1), 181-185. Doi: 10.1674/0003-0031(1998)139[0181:FROASA]2.0.CO2

Durnova, N. A., Demin, A. G., Polukonova, N. V. & Mugue, N. S. (2014). Time of origin of the capacity to accrete and mine submerged substrates in the midge subfamilies Chironominae Macquart, 1838 and Orthocladiinae Lenz, 1921 (Diptera, Chironomidae): Analysis of mitochondrial genes COI and COII. *Entomological Review, 94*(7), 949-958. doi: 10.1134/S0013873814070045

Earle, W., Mangan, R., O'Brien, M. & Baars, J. (2013). Biology of *Polypedilum* n. sp. (Diptera: Chironomidae), a promising candidate agent for the biological control of the aquatic weed *Lagarosiphon* major (Hydrocharitaceae) in Ireland. *Biocontrol Science and Technology, 23*(11), 1267-1283. doi: 10.1080/09583157.2013.826344

Egan, A.T. & Ferrington, L.C. (2015). Chironomidae (Diptera) in freshwater coastal rock pools at Isle Royale, Michigan. *Transactions of the American Entomological Society, 141*(1), 1-25. doi: 10.3157/061.141.0102

Egan, A.T. & Ferrington, L.C. (2019). Chironomidae of the upper Saint Croix river, Wisconsin. *Transactions of the American Entomological Society, 145*(3), 353-384. doi: 10.3157/061/145/0307

Enushchenko, I.V. (2018). Most interesting findings from remnants of Chironomid (Insecta: Diptera: Chironomidae) larvae in layers of a 200-year-old stratum of lake Oron bottom sediments. *Russian Journal of Ecology, 49*(1), 69-74. doi: 10.1134/S1067413618010058

Ghonaim, M., Ali, A. & Salem, M. (2004). *Tanytarsus* (Diptera: Chironomidae) from Egypt with description of a new species. *The Florida Entomologist, 87*(4), 571-575. doi: 10.1653/0015-4040(2004)087[0571:TDCEWD]2.0.CO2

Goffová, K., Čiamporová-Zat'ovičová, Z. & Čiampor, F. (2015). Twenty-one new microsatellite markers for the ecologically important midge *Heterotrissocladius marcidus*, and their use in studies of alpine lakes. *Biochemical Systematics and Ecology, 61*, 329-335. doi: 10.1016/j.bse.2015.06.037

Hänel, C. & Chown, S.L. (1997). The impact of a small, alien invertebrate on a sub-Antarctic terrestrial ecosystem: *Limnophyes minimus* (Diptera, Chironomidae) at Marion Island. *Polar Biology, 20*(2), 99-106. doi: 10.1007/s003000050282

Hazra, N., Saha, G.K. & Chaudhuri, P.K. (2002). Records of Orthoclad species from the Darjeeling-Sikkim Himalayas of India (Diptera: Chironomidae), with notes on their ecology. *Hydrobiologia, 474*(1), 41-55. doi: 10.1023/A:1016511702944

Hayford, B. (2012). Parochlus Kiefferi (Garrett, 1925) in Nebraska (Diptera: Chrionomidae). *Great Plains Research, 22*(1), 27-33.

Hershey, A.E. (1986). Selective predation by *Procladius* in an Arctic Alaskan lake. *Canadian Journal of Fisheries and Aquatic Sciences, 43*(12), 2523-2528. doi: 10.1139/f86-312

Ingvason, H.R., Ólafsson, J.S. & Gardarsson, A. (2004). Food selection of *Tanytarsus gracilentus* larvae (Diptera: Chironomidae): An analysis of instars and cohorts. *Aquatic Ecology, 38*(2), 231-237. doi: 10.1023/B:AECO.0000032053.67992.03

Johnson, R.K. (1987). Seasonal variation in diet of *Chironomus plumosus* (L.) and *C. anthracinus* Zett. (Diptera: Chironomidae) in mesotrophic Lake Erken. *Freshwater Biology, 17*(3), 525-532. doi: 10.1111/j.1365-2427.1987.tb01073.x

Lencioni, V. & Bernabò, P. (2015). Thermal survival limits of young and mature larvae of a cold stenothermal chironomid from the Alps (Diamesinae: Pseudodiamesa branickii [Nowicki, 1873]. *Insect Science, 24*(2), 314-324. doi: 10.1111/1744-7917.12278

Li, X. & Wang, X. (2014). New species and records of *Metriocnemus* van der Wulp s. str. from China (Diptera, Chironomidae). *ZooKeys, 387*(387), 73-87. doi: 10.3897/zookeys.387.6408

Lilley, T.M., Ruokolainen, L., Pikkarainen, A., Laine, V.N., Kilpimaa, J., Rantala, M.J., Nikinmaa, M. (2012). Impact of Tributyltin on immune response and life history traits of *Chironomus riparius*: single and multigeneration effects and recovery from pollution. *Environmental Science & Technology, 46*(13), 7382-7389. doi: 10.1021/es300536t

Malchow, D.E., Knight, A.W. & Maier, K.J. (1995). Bioaccumulation and toxicity of selenium in *Chironomus decorus* larvae fed a diet of seleniferous *Selenastrum capricornutum. Archives of Environmental Contamination and Toxicology, 29*(1), 104-109. doi: 10.1007/BF00213094

McLachlan, A.J. (1986). Sexual dimorphism in midges: strategies for flight in the rain-pool dweller *Chironomus imicola* (Diptera: Chironomidae). *The Journal of Animal Ecology, 55*(1), 261-267. doi: 10.2307/4706

Medeiros, A.S., Milošević, D., Francis, D.R., Maddison, E., Woodroffe, S., Long, A., Walker, I.R., Hamerlík, L., Quinlan, R., Langdon, P., Brodersen, K.P. & Axford, Y. (2020). Arctic chironomids of the northwest North Atlantic reflect environmental and biogeographic gradients. *Journal of Biogeography, 48*(3), 511-525. doi: 10.1111/jbi.14015

Medeiros, A.S. & Quinlan, R. (2011). The distribution of the Chironomidae (Insecta: Diptera) along multiple environmental gradients in lakes and ponds of the eastern Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences, 68*(9), 1511-1527. doi: 10.1139/F2011-076

Moubayed-Breil, J. & Bitušík, P. (2019). Taxonomic notes on the genus *Chaetocladius* (laminatus-group) (Diptera: Chironomidae, Orthocladiinae) II. Descriptions of *C. bitusiki* sp. n. and *C. mantetensis* sp. n., two relict species inhabiting cold alpine springs and streams. *Biologia, 74*(11), 1489-1500. doi: 10.2478/s11756-019-00253-8

Namayandeh, A. & Beresford, D.V. (2019). Chironomidae of Fosheim Peninsula, Ellesmere Island: first female description of *Chaetocladius* (Chaetocladius) *glacialis* (Lundström, 1915) (Diptera: Chironomidae: Orthocladiinae) and new faunistic records from the Nunavut (Canada) and the Nearctic Region. *Aquatic Insects, 40*(1), 1-18. doi: 10.1080/01650424.2018.1523435

Niedrist, G.H. & Füreder, L. (2018). When the going gets tough, the tough get going: The enigma of survival strategies in harsh glacial stream environments. *Freshwater Biology, 63*(10), 1260-1272. doi: 10.1111/fwb.13131

Nieto, E., Corada-Fernández, C., Hampel, M., Lara-Martín, P.A., Sánchez-Argüello, P. & Blasco, J. (2017). Effects of exposure to pharmaceuticals (diclofenac and carbamazepine) spiked sediments in the midge, *Chironomus riparius* (Diptera, Chironomidae). *The Science of the Total Environment, 609*, 715-723. doi: 10.1016/j.scitotenv.2017.07.171

Nolte, U. & Hoffmann, T. (1992). Fast life in cold water*: Diamesa incallida* (Chironomidae). *Ecography, 15*(1), 25-30. doi: 10.1111/j.1600-0587.1992.tb00004.x

Paasivirta, L. (2014). Checklist of the family Chironomidae (Diptera) of Finland. *ZooKeys, 441*(441), 63-90. doi: 10.3897/zookeys.441.7461

Prat, N. & García-Roger, E.M. (2018). Are there so many congeneric species of chironomid larvae in a small stream? *Journal of Limnology, 77*(s1), 65-80. doi: 10.4081/jlimnol.2018.1757

Ricciardi, A. (1994). Occurrence of chironomid larvae (*Paratanytarsus* sp.) as commensals of dreissenid mussels (*Dreissena polymorpha* and *D. bugensis*). *Canadian Journal of Zoology, 72*(6), 1159-1162. doi: 10.1139/z94-155

Silva, C.J.M., Silva, A.L.P., Gravato, C. & Pestana, J.L.T. (2019). Ingestion of small-sized and irregularly shaped polyethylene microplastics affect *Chironomus riparius* life-history traits. *The Science of the Total Environment, 672*, 862-868. doi: 10.1016/j.scitotenv.2019.04.017

Som, D.K., Das, N. & Hazra, N. (2013). Systematics and biology of Metriocnemus clarivirgulus sp. n. (Diptera: Chrionomidae) from Darjeeling, India with revised keys to male and female adults of Metriocnemus van der Wulp. *Deutsche Entomologische Zeitschrift*, *60*(1), 111-121. doi: 10.1002/mmnd.201300014

Stoaks, R. (2009). Rare and unusual arthropod ecology, behavior and occurrence at hummingbird feeders in northern California. *The Pan-Pacific Entomologist, 85*(4), 194-198. doi: 10.3956/2009-18.1

Stur, E. & Ekrem, T. (2006). A revision of West Palaearctic species of the *Micropsectra atrofasciata* species group (Diptera: Chironomidae). *Zoological Journal of the Linnean Society, 146*(2), 165-225. doi: 10.1111/j.1096-3642.2006.00198.x

Stur, E. & 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*(508), 127-142. doi: 10.3897/zookeys.508.9874

Syrjämäki, J. (1968). Diel patterns of swarming and other activities of two Arctic Dipterans (Chironomidae and Trichoceridae) on Spitsbergen. *Oikos, 19*(2), 250-258. doi: 10.2307.3565012

Syrovátka, V. (2018). The predatory behaviour of *Monopelopia tenuicalcar* (Kieffer, 1918) larvae in a laboratory experiment. *Journal of Limnology, 77.* doi: 10.4081/jlimnol.2018.1792

Tarakhovskaya, E.R. & Garbary, D.J. (2009). *Halocladius variabilis* (Diptera: Chironomidae): a marine insect symbiotic with seaweeds from the White Sea, Russia. *Journal of the Marine Biological Association of the United Kingdom, 89*(7), 1381-1385. doi: 10.1017/S0025315409000071

Tarkowska-Kukuryk, M. (2014). Spatial distribution of epiphytic chironomid larvae in a shallow macrophyte-dominated lake: effect of macrophyte species and food resources. *Limnology, 15*(2), 141-153. doi: 10.1007/s10201-014-0425-4

Torrisi, G.J., Hoback, W.W., Foster, J.E., Heinrichs, E.A. & Higley, L.G. (2015). Predation of introduced mosquito larvae by the midge *Metriocnemus knabi* in the Phytotelma of the pitcher plant Sarracenia purpurea and colonization following dry conditions. *Northeastern Naturalist, 22*(3), 513-520.

Tucker, T.R., Hudson, P.L. & Riley, S.C. (2016). Observations of cocooned *Hydrobaenus* (Diptera: Chironomidae) larvae in Lake Michigan. *Journal of Great Lakes Research*, *42*(5), 1129-1135. doi: 10.1016/j.jglr.2016.07.013

Vinogradova, E.M., Riss, H.W. & Spies, M. (2009). New species of *Tanytarsus* van der Wulp, 1874 (Diptera: Chironomidae) from Central America. *Aquatic Insects, 31*(1), 11-17. doi: 10.1080/01650420802502477

Wrubleski, D.A. & Roback, S.S. (1987). Two species of *Procladius* (Diptera:Chironomidae) from a Northern prairie marsh: descriptions, phenologies and mating behaviour. *Journal of the North American Benthological Society, 6*(3), 198-212. doi: 10.2307/1467511

Younc, R.M. (1969). Field observations on a midwinter breeding flight of Diamesa arctica (Diptera: Chironomidae). *Annals of the Entomological Society of America, 62*(5), 1204. doi: 10.1093/aesa/62.5.1204

Chloropidae

Allen, W.A. & Pienkowski, R.L. (1974). The biology and seasonal abundance of the fruit fly, *Oscinella* in reed canarygrass in Virginia. *Annals of the Entomological Society of America, 67*(4), 539-544. doi: 10.1093/aesa/67.4.539

Jones, M.G. (1969). Effect of weather on frit fly (*Oscinella* frit L.) and its predators. *The Journal of Applied Ecology, 6*(3), 425-441. doi: 10.2307/2401509

Nartshuk, E.P. (2014). Grass-fly larvae (Diptera, Chloropidae): Diversity, habitats, and feeding specializations. *Entomological Review, 94*(4), 514-525. doi: 10.1134/S001387381404006X

Nartshuk, E.P. & Khruleva, O.A. (2011). Plant-feeding dipterans (Diptera, Chloropidae, Agromyzidae) from Wrangel Island (the Chukchi Sea). *Entomological Review, 91*(7), 849-854. doi: 10.1134/S0013873811070062

Riccardi, P.R. & Amorim, D.S. (2020). Phylogenetic relationships and classification of the Chloropinae of the world (Diptera: Chloropidae). *Zoological Journal of the Linnean Society, 190*(3), 889-941. doi: 10.1093/zoolinnean/zlaa007

Triseleva, T.A., Petrosyan, V.G., Yatsuk, A.A. & Safonkin, A.F. (2021). The role of plants in the formation of species-specific features in grass flies (Diptera, Chloropidae, Meromyza). *Biodiversity Data Journal, 9*(20), p.e78017. doi: 10.3897/BDJ.9.e78017

Vandevyvere, I. & Bruyn, L. de. Morphological and histochemical analysis of galls of *Lipara lucens* (Diptera, Chloropidae) on *Phragmites australis* (Poaceae). *Canadian Journal of Botany, 76*(8), 1374-1384. doi: 10.1139/b98-143

Clusiidae

Roháček, J. (2012). The fauna of the opomyzoid families Clusiidae, Acartophthalmidae, Anthomyzidae, Opomyzidae, Stenomicridae, Periscelididae, Asteiidae (Diptera) in the Gemer area (Central Slovakia). Časopis Slezského Zemského Muzea, 61(2), 97-111. doi: 10.2478/v10210-012-0011-5

Coelopidae

Edward, D.A., Blyth, J.E., Mckee, R. & Gilburn, A.S. (2007). Change in the distribution of a member of the strand line community: the seaweed fly (Diptera: Coelopidae). *Ecological Entomology, 32*(6), 741-746. doi: 10.1111/j.1365-2311.2007.00919.x

Culicidae

Andrew, E.S., Xu, G. & Rich, S.M. (2014). Microbial communities within field-collected *Culiseta melanura* and *Coquillettidia perturbans. Medical and Veterinary Entomology, 28*(2), 125-132. doi: 10.1111/mve.12017

Anderson, I.H. (1992). Effect of sugar meals and body size on fecundity and longevity of female *Aedes communis* (Diptera: Culicidae). *Physiological Entomology, 17*(3), 203-207. Doi: 10.1111/j.1365-3032.1992.tb01011.x

Bartlett-Healy, K., Crans, W. & Gaugler, R. (2009). Vertebrate hosts and phylogenetic relationships of amphibian trypanosomes from a potential invertebrate vector, *Culex territans* Walker (Diptera: Culicidae). *The Journal of Parasitology, 95*(2), 381-387. doi: 10.1645/GE-1793.1

Benedict, M.Q., Hunt, C.M., Vella, M.G., Gonzalez, K.M., Dotson, E.M., Collins, C.M. & Hansen, I.A. (2020). Pragmatic selection of larval mosquito diets for insectary rearing of *Anopheles gambiae* and *Aedes aegypti. PLoS One, 15*(3), p.e0221838. doi: 10.1371/journal.pone.0221838

Friend, W.G., Smith, J.J.B., Schmidt, J.M. & Tanner, R.J. (1989). Ingestion and diet destination in *Culiseta inornata*: responses to water, sucrose and cellobiose. *Physiological Entomology, 14*(2), 137-146. doi: 10.1111/j.1365-3032.1989.tb00945.x

Gadawski, R.M. & Smith, S.M. (1992). Nectar sources and age structure in a population of *Aedes provocans* (Ditpera: Culicidae). *Journal of Medical Entomology, 29*(5), 879-886. doi: 10.1093/jmedent/29.5.879

Greenberg, J.A., Lujan, D.A., DiMenna, M.A., Wearing, H.J. & Hofkin, B.V. (2013). Identification of blood meal sources in *Aedes vexans* and *Culex quinquefasciatus* in Bernalillo County, New Mexico. *Journal of Insect Science, 13*(75), 1-12. Doi: 10.1673/031.013.7501

Ribeiro, J.M.C., Martin-Martin, I., Arcà, B., Calvo, E. & Hansen, I.A. (2016). A deep insight into the sialome of male and female *Aedes aegypti* mosquitoes. *PLoS One, 11*(3), p.e0151400. doi: 10.1371/journal.pone.0151400

Saab, S.A., zu Dohna, H., Nilsson, L.K.J., Onorati, P., Nakhleh, J., Terenius, O. & Osta, M.A. (2020). The environment and species affect gut bacteria composition in laboratory co-cultured *Anopheles gambiae* and *Aedes albopictus* mosquitoes. *Scientific Reports, 10*(1), 3352. doi: 10.1038/s41598-020-60075-6

Schaefer, M.L. & Lundstroem, J.O. (2006). Different responses of two floodwater mosquito species, *Aedes vexans* and *Ochlerotatus sticticus* (Diptera: Culicidae), to larval habitat drying. *Journal of Vector Ecology, 31*(1), 123-128.

Sherwood, J. A., Stehman, S., Howard, J. J. & Oliver, J. (2020). Cases of Eastern equine encephalitis in humans associated with *Aedes canadensis*, *Coquillettidia perturbans* and *Culiseta melanura* mosquitoes with the virus in New York State from 1971 to 2012 by analysis of aggregated published data. *Epidemiology and Infection, 148*, p.e72. doi: 10.1017/S0950268820000308

Smith, S.M. & Gadawski, R.M. (1994). Nectar feeding by the early-spring mosquito *Aedes provocans. Medical and Veterinary Entomology, 8*(3), 201-213. doi: 10.1111/j.1365-2915.1994.tb00499.x

Yee, W.L., Foster, W.A., Howe, M.J. & Hancock, R.G. (1992). Simultaneous field comparison of evening temporal distributions of nectar and blood feeding by

Aedes vexans and Ae. trivittatus (Diptera: Culicidae) in Ohio. *Journal of Medical Entomology, 29*(2), 356-360. doi: 10.1093/jmedent/29.2.356

Dolichopodidae

Bernasconi, M.V., Pollet, M., Varini-Ooijen, M., Irvine Ward, P. (2007). Phylogeny of European *Dolichopus* and *Gymnopternus* (Diptera: Dolichopodidae) and the significance of morphological characters inferred from molecular data. *European Journal of Entomology, 104*(3), 601-617. doi: 10.14411/eje.2007.075

Chursina, M.A. & Negrobov, O.P. (2020). Morphometric variations in leg segments' length in the family Dolichopodidae (Diptera). *Moscow University Biological Sciences Bulletin, 75*(3), 110-116. doi: 10.3103/S0096392520030049

Gelbic, I. & Olejnicek, J. (2011). Ecology of Dolichopodidae (Diptera) in a wetland habitat and their potential role as bioindicators. *Central European Journal of Biology, 6*(1), 118-129. doi: 10.2478/s11535-010-0098-x

Hulcr, J., Pollet, M., Ubik, K. & Vrkoč, J. (2005). Exploitation of kairomones and synomones by *Medetera* spp. (Diptera: Dolichopodidae), predators of spruce bark beetles. *European Journal of Entomology, 102*(4), 655-662. Doi: 10.14411/eje.2005.093

Kautz, A.R. & Gardiner, M.M. (2019). Agricultural intensification may create an attractive sink for Dolichopodidae, a ubiquitous but understudied predatory fly family. *Journal of Insect Conservation, 23*(3), 453-465. doi: 10.1007/s10841-018-0116-2

Martin, O.Y., Puniamoorthy, N., Gubler, A., Wimmer, C. & Bernasconi, M.V. (2013). Infections with Wolbachia, Spiroplasma, and Rickettsia in the Dolichopodidae and other Empidoidea. *Infection, Genetics and Evolution, 13*, 317-330. doi: 10.1016/j.meegid.2012.11.005

Mcghehey, J.H. & Nagel, W.P. (1966). A technique for rearing larvae of *Medetera aldrichii* (Diptera: Dolichopodidae). *Annals of the Entomological Society of America, 59*(2), 290-292.

Drosophilidae

Band, H.T., Bachli, G. & Band, R.N. Behavioral constancy for interspecies dependency enables Nearctic *Chymomyza amoena* (Loew) (Diptera: Drosophilidae) to spread in orchards and forests in Central and Southern Europe. *Biological Invasions, 7*(3), 509-530. doi: 10.1007/s10530-004-6352-2

Church, S.H., Extavour, C.G. & Rosenberg, M. (2022). Phylotranscriptomics reveals discordance in the phylogeny of Hawaiian *Drosophila* and *Scaptomyza* (Diptera: Drosophilidae). *Molecular Biology and Evolution, 39*(3). doi: 10.1093/molbev/msac012

Dyer, K.A. (2012). Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution, 66*(4), 973-984. doi: 10.1111/j.1558-5646.2011.01497.x

Dyer, KA., White, B.E., Sztepanacz, J.L., Bewick, E.R. & Rundle, H.D. (2014). Reproductive character displacement of epicuticular compounds and their contribution to mate choice in *Drosophila subquinaria* and *Drosophila recens*. *Evolution, 68*(4), 1163-1175. doi: 10.1111/evo.12335

Jaenike, J. (1985). Genetic and environmental determinants of food preference in *Drosophila tripunctata. Evolution, 39*(2), 362-369. doi: 10.1111/j.1558-5646.1985.tb05673.x

James, A.C., Jakubczak, J., Riley, M.P. & Jaenike, J. (1988). On the causes of monophagy in *Drosophila quinarian. Evolution, 42*(3), 626-630. doi: 10.1111/j.1558-5646.1988.tb04166.x

Jaramillo, S.L., Mehlferber, E. & Moore, P.J. (2015). Life-history trade-offs under different larval diets in *Drosophila suzukii* (Diptera: Drosophilidae). *Physiological Entomology, 40*(1), 2-9. doi: 10.1111/phen.12082

Koštál, V., Mollaei, M. & Schöttner, K. (2016). Diapause induction as an interplay between seasonal token stimuli, and modifying and directly limiting factors: hibernation in *Chymomyza costata*: Environmental factors inducing diapause. *Physiological Entomology, 41*(4), 344-357. doi: 10.1111/phen.12159

Stehlík, J., Závodská, R., Shimada, K., Šauman, I. & Koštál, V. (2008). Photoperiodic induction of diapause requires regulated transcription of timeless in the larval brain of *Chymomyza costata. Journal of Biological Rhythms, 23*(2), 129-139. doi: 10.1177/0748730407313364

Worthen, W.B., Jones, M.T. & Jetton, R.M. (1998). Community structure and environmental stress: desiccation promotes nestedness in mycophagous fly communities [*Drosophila tripunctata, Drosophila putrida*, nested-subset pattern, depauperate assemblage]. *Oikos, 81*(1), 45-54. doi: 10.2307/3546466

<u>Dryomyzidae</u>

Gibson, J.F. & Choong, H.H.C. (2021). New range records and life history observations of insects (Diptera: Dryomyzidae, Chironomidae; Coleoptera: Staphylinidae) associated with barnacles (Balanomorpha: Balanidae, Chthamalidae) on the Pacific coasts of North America and Japan. *Canadian Entomologist, 153*(2), 196-210. doi: 10.4039/tce.2020.69

Harley, C.D.G. & Lopez, J.P. (2003). The natural history, thermal physiology, and ecological impacts of intertidal mesopredators, *Oedoparena* spp. (Diptera: Dryomyzidae). *Invertebrate Biology, 122*(1), 61-73. doi: 10.1111/j.1744-7410.2003.tb00073.x

Empididae

Alcock, J. (2016). The mating behavior of an undescribed species of *Rhamphomyia* (Diptera: Empididae). *Journal of Insect Behavior, 29*(2), 153-161. doi: 10.1007/s10905-016-9545-5

Bussière, L.F., Gwynne, D.T. & Brooks, R. (2008). Contrasting sexual selection on males and females in a role-reversed swarming dance fly, *Rhamphomyia longicauda* Loew (Diptera: Empididae). *Journal of Evolutionary Biology, 21*(6), 1683-1691. doi: 10.1111/j.1420-9101.2008.01580.x

Evans, H.E. (1988). Observations on swarms of *Rhamphomyia sociabilis* (Williston) (Diptera: Empididae). *Journal of the New York Entomological Society, 96*(3), 316-322.

Gwynne, D.T. & Bussière, L.F. (2002). Female mating swarms increase predation risk in a 'role-reversed' dance fly (Diptera: Empididae: *Rhamphomyia longicauda* Loew). *Behaviour, 139*(11-12), 1425-1430. doi: 10.1163/15685390260514690

Ivkovic, M., Perovic, M., Grootaert, P. & Pollet, M. (2021). High endemicity in aquatic dance flies of Corsica, France (Diptera, Empididae, Clinocerinae and Hemerodromiinae), with the description of a new species of *Chelipoda. ZooKeys, 1039*(1039), 177-197. doi: 10.3897/zookeys.1039.66493

Khruleva, O. A., Shamshev, I. V. & Sinclair, B. J. (2021). The Empidoid Flies (Diptera: Brachystomatidae, Empididae, Hybotidae) of Wrangel Island (Chukotka Autonomous Okrug): composition and distribution of the fauna. *Entomological Review, 101*(6), 792-819. doi: 10.1134/S0013873821060063

Poinar, G.O. & Vega, F.E. (2021). A new genus of Empididae (Diptera) with enlarged postpedicels in mid-Cretaceous Burmese amber. *Historical Biology, 33*(9), 1830-1835. doi: 10.1080/08912963.2020.1743700

Shamshev, I.V. & Selitskaya, O.G. (2016). Methyl salicylate as an attractant for the dance fly *Rhamphomyia gibba* (Fallen) (Diptera, Empididae). *Entomological Review, 96*(8), 1003-1007. doi: 10.1134/S0013873816080054

Shamshev, I.V. & Sinclair, B.J. (2009). Revision of the *Iteaphila setosa* group (Diptera: Empididae). *European Journal of Entomology, 106*(3), 441-450. doi: 10.14411/eje.2009.055

Ephydridae

Bownes, A. (2018). Effects of *Hydrilla verticillata* (L.f.) Royle (Hydrocharitaceae) growing conditions and nutrient content on the performance of a leaf-mining fly, *Hydrellia purcelli* Deeming (Diptera: Ephydridae). *Biocontrol Science and Technology, 28*(3), 278-292. doi: 10.1080/09583157.2018.1441370

Cabrera Walsh, G., Magalí Dalto, Y., Mattioli, Federico M., Carruthers, Raymond I. & Anderson, L.W. (2013). Biology and ecology of Brazilian elodea *(Egeria densa*) and its specific herbivore*, Hydrellia* sp., in Argentina. *BioControl, 58*(1), 133-147. doi: 10.1007/s10526-012-9475-x

Prado e Castro, C., Santos, V. & Ameixa, O.M.C.C. (2022). Shore fly (Diptera: Ephydridae) communities from Ria de Aveiro, with new records for Portugal. *Fragmenta Entomologica, 54*(1). doi: 10.13133/2284-4880/715

Purcell, M.F., Brown, B.T., Harms, N.E., Sun-Hee, H. & McCulloch, G.A. (2021). Biology and preliminary host range of a Korean leaf-mining *Hydrellia* sp. (Diptera: Ephydridae) rejected as a potential biological control agent for monoecious *Hydrilla verticillata* in the United States. *Biocontrol Science and Technology, 31*(4), 343-356. doi: 10.1080/09583157.2020.1853051

Kubatova-Hirsova, H. (2005). Temporal patterns in shore fly (Diptera, Ephydridae) community structure in a salt marsh habitat. *Ecological Entomology, 30*(2), 234-240. doi: 10.1111/j.0307-6946.2005.00602.x

Mangan, R., Dirilgen, T. & Baars, J. (2015). Responses of adult *Hydrellia lagarosiphon* to a revised diet: implications for life cycle studies and laboratory culturing techniques. *Entomologia Experimentalis et Applicata, 157*(2), 164-169. doi: 10.1111/eea.12350

Morton, A. & García del Pino, F. (2007). Susceptibility of shore fly *Scatella stagnalis* to five entomopathogenic nematode strains in bioassays. *BioControl, 52*(4), 533-545. doi: 10.1007/s10526-006-9047-z

Scheiring, J.F. (1974). Diversity of shore flies (Diptera: Ephydridae) in inland freshwater habitat. *Journal of the Kansas Entomological Society, 47*(4), 485-491.

Scheiring, J.F. (1975). A multivariate analysis of similarity in shore fly habitats (Diptera: Ephydridae). *Journal of the Kansas Entomological Society, 48*(2), 232-243.

Scheiring, J.F. (1977). Geographic variation in *Scatella stagnalis* (Diptera: Ephydridae) [North America, insects]. *Annals of the Entomological Society of America, 70*(4), 511-523. doi: 10.1093/aesa/70.4.511

Smith, R., Mangan, R. & Coetzee, J.A. (2019). Risk assessment to interpret the physiological host range of *Hydrellia egeriae*, a biocontrol agent for *Egeria densa. BioControl, 64*(4), 447-456. Doi: 10.1007/s10526-019-09942-4

Vanninen, I. (2001). Biology of the shore fly *Scatella stagnalis* in rockwool under greenhouse conditions. *Entomologia Experimentalis et Applicata, 98*(3), 317-328. doi: 10.1046/j.1570-7458.2001.00788.x

Fanniidae

Velásquez, Y., Martínez-Sánchez, A. & Rojo, S. (2013). First record of *Fannia leucosticta* (Meigen) (Diptera: Fanniidae) breeding in human corpses. *Forensic Science International, 229*(1), p.e13-e15. doi: 10.1016/j.forsciint.2013.03.022

Wang, M., Li, W., Zhu, W. & Zhang, D. (2016). Review of the *Fannia postica*-group Chillcott, 1961 of the genus *Fannia* Robineau-Desvoidy, 1830, with description of two new species from the Palearctic and Oriental regions (Diptera, Fanniidae). *ZooKeys, 598*(598), 113-128. doi: 10.3897/zookeys.598.7983

Heleomyzidae

Rotheray, G.E. (2012). Morphology of the puparium and breeding sites of eight species of Heleomyzidae (Diptera). *Journal of Natural History, 46*(33-34), 2075-2102. doi: 10.1080/00222933.2012.707241

Soszynska-Maj, A. & Woznica, A.J. (2016). A case study of Heleomyzidae (Diptera) recorded on snow in Poland with a review of their winter activity in Europe. *European Journal of Entomology, 113*(1), 279-294. doi: 10.14411/eje.2016.035

Tuno, N., Sagara, N. & Okadome, T. (2003). Territorial behaviour of *Suillia* males on basidiocarps of *Hebeloma radicosum* in central Japan. *Mycologist, 17*(3), 122-125. doi: 10.1017/S0269915X03002143

Hybotidae

Bartak, M. & Kubik, S. (2015). Three new species of European *Platypalpus* (Diptera, Hybotidae). *ZooKeys, 470*(470), 145-155. doi: 10.3897/zookeys.470.8967

Jaume-Schinkel, S., Machado Barros, L., Breno Graça, M. & Mota Soares, M.M. (2022). In natura sit-and-wait behaviour and predation success of a Neotropical dance fly (Diptera: Hybotidae). *Journal of Natural History, 56*(1-4), 1-14. doi: 10.1080/00222933.2022.2032443

Khruleva, O. A., Shamshev, I. V. & Sinclair, B. J. (2021). The Empidoid Flies (Diptera: Brachystomatidae, Empididae, Hybotidae) of Wrangel Island (Chukotka Autonomous Okrug): composition and distribution of the fauna. *Entomological Review, 101*(6), 792-819. doi: 10.1134/S0013873821060063

Keroplatidae

Sivinski, J. (1982). Prey attraction by luminous larvae of the fungus gnat *Orfelia fultoni. Ecological Entomology, 7*(4), 443-446. doi: 10.1111/j.1365-2311.1982.tb00686.x

Viviani, V.R., Silva, J.R., Amaral, D.T., Bevilaqua, V.R., Abdalla, F.C., Branchini, B.R. & Johnson, C.H. (2020). A new brilliantly blue-emitting luciferin-luciferase system from *Orfelia fultoni* and Keroplatinae (Diptera). *Scientific Reports, 10*(1), 9608. doi: 10.1038/s41598-020-66286-1

Lauxaniidae

Davies, G.B.P. & Miller, R.M. (2009). Revision of Afrotropical *Pachycerina* Macquart (Diptera: Lauxaniidae). *African Invertebrates, 50*(2), 295-320. doi: 10.10520/EJC84650

Hesler, L.S. & Fernandes, G.W. (2016). Capture of nontarget flies (Diptera: Lauxaniidae, Chloropidae, and Anthomyiidae) on traps baited with volatile chemicals in field-crop habitats. *Psyche, 2016,* 1-8. doi: 10.1155/2016/6938368

Limoniidae

Bilalli, A., Ibrahimi, H., Musliu, M., Grapci-Kotori, L., Geci, D., Slavevska-Stamenkovic, V., Hinic, J., Mitic-Kopanja, D. & Keresztes, L. (2021). New records of the craneflies (Diptera: Limoniidae, Tipulidae) from the Western Balkans. *Journal of Entomological Research Society, 23*(2), 141-152. Doi: 10.51963/jers.v23i2.1929

Ferreira, S., Oosterbroek, P., Stary, J., Sousa, P., Mata, V.A., da Silva, L.P., Pauperio, J. & Beja, P. (2021). The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese Diptera 02-Limoniidae, Pediciidae and Tipulidae. *Biodiversity Data Journal, 9*(4), p.e69841. doi: 10.3897/BDJ.9.e69841

Kolcsár, L., Ivković, M. & Ternjej, I. (2015). New records of Limoniidae and Pediciidae (Diptera) from Croatia. *ZooKeys, 513*(513), 23-37. Doi: 10.3897/zookeys.513.10066

Ozgul, O. & Koc, H. (2014). Four new species of Limoniidae (Diptera, Nematocera) from the Inner-West Anatolian Subregion of Turkey. *The Florida Entomologist, 97*(2), 620-625. doi: 10.1653/024.097.0238

Petersen, M.J. (2008). A systematic investigation of the crane fly subfamily Limoniinae (Tipuloidea: Limoniidae). Iowa State University. doi: 10.31274/rtd-180813-6702

Podenas, S. & Gelhaus, J. (2011). Three new species of Chioneinae crane flies (Diptera: Limoniidae) from North-Central Mongolia. *Proceedings of the Academy of Natural Sciences of Philadelphia, 161*(1), 73-86. doi: 10.1635/053.161.0105

Podeniene, V., Podenas, S. & Gelhaus, J.K. (2002). The first description of the larva of the crane fly genus *Gonempeda* Alexander, 1924 (Limoniidae: Chioneinae), with new information for understanding the phylogenetic relationships of the genus. *Proceedings of the Academy of Natural Sciences of Philadelphia, 152*(1), 67-73. doi: 10.1635/0097-3157(2002)152[0067:TFDOTL]2.0.CO2

Podeniene, V., Podenas, S. & Gelhaus, J.K. (2004). First record of a crane fly larva (Diptera, Limoniidae: Chioneinae) from Baltic amber. *Annals of the Entomological Society of America, 97*(6), 1126-1128. doi: 10.1603/0013-8746(2004)097[1126:FROACF]2.0.CO2

Suetsugu, K., Tetsu, S. & Sueyoshi, M. (2019). Fruit- and seed-feeding habit of the crane fly *Libnotes puella* (Diptera: Limoniidae) in Ryukyu Islands, Japan. *Entomological Science, 22*(4), 413-417. doi: 10.1111/ens.12387

<u>Lonchopteridae</u>

Stalker, H.D. (1956). On the evolution of parthenogenesis in *Lonchoptera* (Diptera). *Evolution, 10*(4), 345-359. doi: 10.1111/j.1558-5646.1956.tb02862.x

Micropezidae

Barnes, J.K. (2015). Biology and immature stages of *Compsobata univitta* (Walker, 1849) (Diptera: Micropezidae: Calobatinae). *Proceedings of the Entomological Society of Washington, 117*(4), 421-434. doi: 10.4289/0013-8797.117.4.421

Milichiidae

Giordani, G., Tuccia, F., Zoppis, S., Vecchiotti, C. & Vanin, S. (2019). Record of *Leptometopa latipes* (Diptera: Milichiidae) from a human cadaver in the Mediterranean area. *Forensic Sciences Research, 4*(4), 341-347. doi: 10.1080/20961790.2018.1490473

Xi, Y., Yang, D. & Yin, X. (2018). The genus *Phyllomyza* Fallén from China, with descriptions of three new species (Diptera, Milichiidae). *ZooKeys, 760*(760), 143-157. doi: 10.3897/zookeys.760.22595

Yusah, K.M. & Fayle, T.M. (2014). The first record of a fly of the family Milichiidae (Diptera) interacting with an ant of the genus *Polyrhachis* Smith, 1857 (Hymenoptera: Formicidae). *Biodiversity Data Journal, 2*(2), p.e4168. doi: 10.3897/BDJ.2.e4168

Muscidae

Achint, R. & Singh, D. (2021). Application of COI gene for identification of some economically and forensically important muscid flies of India (Diptera: Muscidae). *International Journal of Tropical Insect Science, 41*(4), 3023-3029. doi: 10.1007/s42690-021-00494-8

Baleba, S.B.S., Torto, B., Masiga, D., Weldon, C.W. & Getahun, M.N. (2019). Egg-laying decisions based on olfactory cues enhance offspring fitness in *Stomoxys calcitrans* L. (Diptera: Muscidae). *Scientific Reports, 9*(1), 3850. doi: 10.1038/s41598-019-40479-9

Bautista-Martínez, N., Illescas-Riquelme, C.P. & García-Ávila, C., de J. (2017). First report of “hunter-fly” *Coenosia attenuate* (Diptera: Muscidae) in Mexico. *The Florida Entomologist, 100*(1), 174-175. doi: 10.1653/024.100.0126

Butterworth, N.J. & Wallman, J.F. (2022). Flies getting filthy: The precopulatory mating behaviours of three mud-dwelling species of Australian *Lispe* (Diptera: Muscidae). *Ethology, 128*(4), 369-377. doi: 10.1111/eth.13236

Fowler, F.E. & Mullens, B.A. (2016). Dividing the pie: differential dung pat size utilization by sympatric *Haematobia irritans* and *Musca autumnalis*: Dung pat size use by horn flies and face flies. *Medical and Veterinary Entomology, 30*(2), 185-192. doi: 10.1111/mve.12166

Getahun, M.N., Ahuya, P., Ngiela, J., Orone, A., Masiga, D. & Torto, B. (2020). Shared volatile organic compounds between camel metabolic products elicits strong *Stomoxys calcitrans* attraction. *Scientific Reports, 10*(1), 21454. doi: 10.1038/s41598-020-78495-9

Grzywacz, A., Hall, M.J.R. & Pape, T. (2015). Morphology successfully separates third instar larvae of *Muscina. Medical and Veterinary Entomology, 29*(3), 314-329. doi: 10.1111/mve.12117

Gunn, A. (2016). The colonisation of remains by the muscid flies *Muscina stabulans* (Fallén) and *Muscina prolapsa* (Harris) (Diptera: Muscidae). *Forensic Science International, 266*, 349-356. doi: 10.1016/j.forsciint.2016.06.013

Hieu, T.T., Jung, J., Kim, S.., Ahn, Y. & Kwon, H.W. (2014). Behavioural and electroantennogram responses of the stable fly (*Stomoxys calcitrans* L.) to plant essential oils and their mixtures with attractants. *Pest Management Science, 70*(1), 163-172. doi: 10.1002/ps.3547

Ivkovic, M. & Pont, A.C. (2016). Long-time emergence patterns of *Limnophora* species (Diptera, Muscidae) in specific karst habitats: tufa barriers. *Limnologica, 61*, 29-35. doi: 10.1016/j.limno.2016.09.003

Krafsur, E.S. & Moon, R.D. (1997). Bionomics of the face fly, *Musca autumnalis*. *Annual Review of Entomology, 42*(1), 503-523. doi: 10.1146/annurev.ento.42.1.503

Krivosheina, N.P. (2013). On the ecology of *Phaonia* larvae (Diptera, Muscidae). *Entomological Review, 93*(3), 324-333. doi: 10.1134/S0013873813030068

Krivosheina, N.P. (2014). Morpho-ecological characteristic of immature stages of the genus *Mydaea* (Diptera, Muscidae). *Entomological Review, 94*(5), 675-686. doi: 10.1134/S0013873814050042

Krivosheina, N.P. (2019). Morpho-ecological characteristics of immature stages of Xytobiont species of the genus *Phaonia* Robineau-Desvoidy, 1830 (Muscidae, Diptera) inhabiting tree holes and sap accumulations. Entomological Review, 99(2), 250-261. doi: 10.1134/S0013873819020118

Kutty, S.N., Pont, A.C., Meier, R. & Pape, T. (2014). Complete tribal sampling reveals basal split in Muscidae (Diptera), confirms saprophagy as ancestral feeding mode, and reveals an evolutionary correlation between instar numbers and carnivory. *Molecular Phylogenetics and Evolution, 78*, 349-364. doi: 10.1016/j.ympev.2014.05.027

Liénard, E., Salem, A., Jacquiet, P., Grisez, C., Prévot, F., Blanchard, B., Bouhsira, E. & Franc, M. (1987). Development of a protocol testing the ability of *Stomoxys calcitrans* (Linnaeus, 1758) (Diptera: Muscidae) to transmit *Besnoitia besnoiti* (Henry, 1913) (Apicomplexa: Sarcocystidae). *Parasitology Research, 112*(2), 479-486. doi: 10.1007/s00436-012-3157-6

Liu, G., Wang, Q., Liu, X., Li, X., Pang, X. & Zhang, D. (2021). Antennal and palpal sensilla of three predatory *Lispe* species (Diptera: Muscidae): an ultrastructural investigation. *Scientific Reports, 11*(1), 18357. doi: 10.1038/s41598-021-97677-7

Olafson, PU., Aksoy, S., Attardo, G.M., Buckmeier, G., Chen, X., Coates, C.J., Davis, M., Dykema, J., Emrich, S.J., Friedrich, M., Holmes, C.J., Ioannidis, P., Jansen, E.N., Jennings, E.C., Lawson, D., Martinson, E.O., Maslen, G.L., Meisel, R.P., Murphy, T.D., Nayduch, D., Nelson, D.R., Oyen, K.J., Raszick, T.J., Ribeiro, J.M.C., Robertson, H.M., Rosendale, A.J., Sackton, T.B., Saelao, P., Swiger, S.L., Sze, S., Tarone, A.M., Taylor, D.B., Warren, W.C., Waterhouse, R.M., Weirauch, M.T., Werren, J.H., Wilson, R.K., Zdobnov, E.M. & Benoit, J.B. (2021). The genome of the stable fly, *Stomoxys calcitrans*, reveals potential mechanisms underlying reproduction, host interactions, and novel targets for pest control. *BMC Biology, 19*(1), 41. doi: 10.1186/s12915-021-00975-9

Olesen, A.S. Lohse, L., Hansen, M.F., Boklund, A., Halasa, T., Belsham, G.J., Rasmussen, T.B., Bøtner, A. & Bødker, R. (2018). Infection of pigs with African swine fever virus via ingestion of stable flies (*Stomoxys calcitrans*). *Transboundary and Emerging Diseases, 65*(5), 1152-1157. doi: 10.1111/tbed.12918

Phillips, H.N., Moon, R.D., Sorge, U.S. & Heins, B.J. (2020). Efficacy of broilers as a method of face fly (*Musca autumnalis* De Geer) larva control for organic dairy production. *Animals, 10*(12), 2429. doi: 10.3390/ani10122429

Pont, A.C. & Ivković, M. (2013). Hunter-flies of Croatia (Diptera: Muscidae: genus *Limnophora* Robineau-Desvoidy). *Journal of Natural History, 47*(15-16), 1069-1082. doi: 10.1080/00222933.2012.750775

Rotheray, G.E. & Wilkinson, G. (2015). Trophic structure and function in the larvae of predatory muscid flies (Diptera, Muscidae). Zoomorphology, 134(4), 553-563. doi: 10.1007/s00435-015-0284-5

Savage, J. & Sorokina, V.S. (2021). Review of the North American fauna of *Drymeia* Meigen (Diptera, Muscidae) and evaluation of DNA barcodes for species-level identification in the genus. *ZooKeys, 1024*(5), 31-89. doi: 10.3897/zookeys.1024.60393

Savage, J., Wheeler, T.A. & Wiegmann, B.M. (2004). Phylogenetic analysis of the genus *Thricops Rondani* (Diptera: Muscidae) based on molecular and morphological characters. *Systematic Entomology, 29*(3), 395-414. doi: 10.1111/j.0307-6970.2004.00252.x

Seabra, S.G., Martins, J., Bras, P., Tavares, A.M., Freitas, I., Barata, A., Rebelo, M.T., Mateus, C., Paulo, O.S. & Figueiredo, E. (2021). PCR-based detection of prey DNA in the gut contents of the tiger-fly, *Coenosia attenuata* (Diptera: Muscidae), a biological control agent in Mediterranean greenhouses. *European Journal of Entomology, 118*(1), 335-343. doi: 10.14411/eje.2021.035

Sommer, C., Jensen, K.-M.V. & Jespersen, J.B. (2001). Topical treatment of calves with synthetic pyrethroids: effects on the non-target dung fly *Neomyia cornicina* (Diptera: Muscidae). *Bulletin of Entomological Research, 91*(2), 131-137. doi: 10.1079/BER200079

Sorokina, V.S. & Ovtshinnikova, O.G. (2020). The position of the Azeliinae in the Muscidae (Diptera) based on musculature of the male terminalia. *ZooKeys, 975*(975), 87-110. doi: 10.3897/zookeys.975.55502

Sowig, P., Himmelsbach, R. & Himmelsbach, W. (1997). Predator-prey relationship between insect larvae: growth of *Sphaeridium* larvae (Coleoptera: Hydrophilidae) under time constraints through predation on *Musca autumnalis* maggots (Diptera: Muscidae). *Canadian Journal of Zoology, 75*(12), 2069-2076. doi: 10.1139/z97-841

Xue, W., Zhang, L. & Wang, M. (2008). A key to the species of *Spilogona* Schnabl from China, with descriptions of four new species from the Qinghai-Xizang (Tibetan) Plateau (Diptera: Muscidae). *Entomological Science, 11*(1), 87-95. doi: 10.1111/j.1479-8298.2007.00257.x

Vikhrev, N.E. (2011). Review of the Palaearctic members of the *Lispe tentaculata* species-group (Diptera, Muscidae): revised key, synonymy and notes on ecology. *ZooKeys, 84*(84), 59-70. doi: 10.3897/zookeys.84.819

Vikhrev, N.E. (2012). Revision of the *Lispe longicollis*-group (Diptera, Muscidae). *ZooKeys, 235*(235), 23-39. doi: 10.3897/zookeys.235.3306

Wang, Y., Zhang, Y., Wang, M., Hu, G., Fu, Y., Zhi, R., Wang, J. & Lysyk, T. (2020). Development of *Hydrotaea spinigera* (Diptera: Muscidae) at constant

temperatures and its significance for estimating postmortem interval. *Journal of Medical Entomology, 58*(1), 56-63. doi: 10.1093/jme/tjaa162

Zou, D., Coudron, T.A., Xu, W., Xu, J. & Wu, H. (2020). Performance of the tiger-fly *Coenosia attenuata* Stein reared on the alternative prey, *Chironomus plumosus* (L.) larvae in coir substrate. *Phytoparasitica, 49*(1), 83-92. doi: 10.1007/s12600-020-00866-9

Zou, D., Coudron, T.A., Zhang, L., Xu, W., Xu, J., Wang, M., Xiao, X. & Wu, H. (2021). Effect of prey species and prey densities on the performance of adult *Coenosia attenuate. Insects, 12*(8), 669. doi: 10.3390/insects12080669

Mycetophilidae

Bauernfeind, R., Schneeberg, K. & Beutel, R.G. (2015). The larval head of *Exechia* (Mycetophilidae) and *Bibio* (Bibionidae) (Diptera). *Arthropod Structure and Development, 44*(4), 326-345. doi: 10.1016/j.asd.2015.04.005

Bouju, V., Rosse-Guillevic, S., Griffon, M., Bojarski, B., Szwedo, J. & Perrichot, V. (2021). The genus *Allodia* (Diptera: Mycetophilidae) in Miocene Ethiopian amber. *Mitteilungen aus dem Museum für Naturkunde in Berlin, 24*(2), 339-346. doi: 10.5194/fr-24-339-2021

Jürgenstein, S., Kurina, O. & Põldmaa, K. (2015). The *Mycetophila ruficollis* Meigen (Diptera, Mycetophilidae) group in Europe: elucidating species delimitation with COI and ITS2 sequence data. *ZooKeys, 508*(508), 15-51. doi: 10.3897/zookeys.508.9814

Kaspřák, D., Kerr, P., Sýkora, V., Tóthová, A. & Ševčík, J. (2019). Molecular phylogeny of the fungus gnat subfamilies Gnoristinae and Mycomyinae, and their position within Mycetophilidae (Diptera): Molecular phylogeny of Gnoristinae and Mycomyinae. *Systematic Entomology, 44*(1), 128-138. doi: 10.1111/syen.12312

Kurina, O. (2021). A contribution towards checklist of fungus gnats (Diptera, Diadocidiidae, Ditomyiidae, Bolitophilidae, Keroplatidae, Mycetophilidae) in Georgia, Transcaucasia. *ZooKeys, 1026*(1026), 69-142. doi: 10.3897/zookeys.1026.63749

Martinsson, S., Kjærandsen, J. & Sundberg, P. (2011). Towards a molecular phylogeny of the fungus gnat genus *Boletina* (Diptera: Mycetophilidae). *Zoologica Scripta, 40*(3), 272-281. doi: 10.1111/j.1463-6409.2011.00474.x

Opomyzidae

Roháček, J. (2012). The fauna of the opomyzoid families Clusiidae, Acartophthalmidae, Anthomyzidae, Opomyzidae, Stenomicridae, Periscelididae, Asteiidae (Diptera) in the Gemer area (Central Slovakia). *Časopis Slezského Zemského Muzea, 61*(2), 97-111. doi: 10.2478/v10210-012-0011-5

Oestridae

Martín-Vega, D., Clark, B., Ferrer, L. M., López-Tamayo, S., Panadero, R., Cepeda-Palacios, R., Colwell, D. D. & Hall, M. J. R. (2021). Major differences in the larval anatomy of the digestive and excretory systems of three Oestridae species revealed by micro-CT. *Medical and Veterinary Entomology, 35*(1), 106-120. doi: 10.1111/mve.12476

Pediciidae

Ferreira, S., Oosterbroek, P., Stary, J., Sousa, P., Mata, V.A., da Silva, L.P., Pauperio, J. & Beja, P. (2021). The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese Diptera 02-Limoniidae, Pediciidae and Tipulidae. *Biodiversity Data Journal*, *9*(4), p.e69841. doi: 10.3897/BDJ.9.e69841

Krivosheina, N.P. (2011). New data on the larval morphology of limoniid flies of the genus *Ula* (Diptera, Pediciidae). *Entomological Review, 91*(4), 432-443. doi: 10.1134/S001387381104004X

Phoridae

Castillo-Alanis, L.A., González-Hernández, A.E., Quijano-Mateos, A., Pedraza-Lara, C.S., Villavicencio-Queijeiro, A. & Bravo-Gómez, M.E. (2020). Standardization of a culture medium for *Megaselia scalaris* (Diptera: Phoridae) for entomotoxicological studies. *Journal of Medical Entomology, 57*(5), 1421-1431. doi: 10.1093/jme/tjaa088

Disney, R.H.L. (2007). Natural history of the scuttle fly, *Megaselia scalaris*. *Annual Review of Entomology, 53*(1), 39-60. doi: 10.1146/annurev.ento.53.103106.093415

Feng, D. & Liu, G. (2012). A study on morphology of immature stages of *Diplonevra peregrina* (Wiedemann) (Diptera: Phoridae). *Microscopy Research and Technique, 75*(10), 1425-1431. doi: 10.1002/jemt.22085

Giangaspero, A., Barlaam, A., Pane, S., Marchili, M.R., Muda, A.O., Putignani, L. & Hall, M.J.R. (2020). Accidental nasal Myiasis caused by *Megaselia rufipes* (Diptera: Phoridae) in a child. *Journal of Medical Entomology, 58*(1), 121-124. doi: 10.1093/jme/tjaa184

Pennington, M.J., Rothman, J.A., Jones, M.B., McFrederick, Q.S., Gan, J. & Trumble, J.T. (2017). Effects of contaminants of emerging concern on *Megaselia scalaris* (Lowe, Diptera: Phoridae) and its microbial community. *Scientific Reports, 7*(1), 1-12. doi: 10.1038/s41598-017-08683-7

Sorensen, M.A., Chase-Dunn, C.M. & Trumble, J.T. (2009). Chronic exposure to elevated levels of manganese and nickel is not harmful to a cosmopolitan detritivore, Megaselia scalaris (Diptera: Phoridae). *Insect Science, 16*(1), 73-79. doi: 10.1111/j.1744-7917.2009.00256.x

Stoaks, R. (2009). Rare and unusual arthropod ecology, behavior and occurrence at hummingbird feeders in northern California. *The Pan-Pacific Entomologist, 85*(4), 194-198. doi: 10.3956/2009-18.1

Piophilidae

Martín-Vega, D. & Baz, A. (2011). Could the ‘vulture restaurants’ be a lifeboat for the recently rediscovered bone-skippers (Diptera: Piophilidae)? *Journal of Insect Conservation, 15*(5), 747-753. doi: 10.1007/s10841-011-9429-0

Oudin, M.J., Bonduriansky, R. & Rundle, H.D. (2015). Experimental evidence of condition-dependent sexual dimorphism in the weakly dimorphic antler fly *Protopiophila litigata* (Diptera: Piophilidae). *Biological Journal of the Linnean Society, 116*(1), 211-220. doi: 10.1111/bij.12549

Taleb, M., Tail, G. & Açıkgöz, H.N. (2019). DNA barcoding of *Stearibia nigriceps* (Meigen) and *Piophila casei* (Linnaeus) (Diptera: Piophilidae) from Algeria and the first African report of *Stearibia nigriceps. International Journal of Legal Medicine, 134*(3), 895-902. doi: 10.1007/s00414-019-02223-w

Pipunculidae

Jervis, M.A. (1992). A taxonomic revision of the pipunculid fly genus *Chalarus* Walker, with particular reference to the European fauna. *Zoological Journal of the Linnean Society, 105*(3), 243-352. doi: 10.1111/j.1096-3642.1992.tb01232.x

Kehlmaier, C. (2014). Checklist of the family Pipunculidae (Diptera) of Finland. *ZooKeys, 441*(441), 225-232. doi: 10.3897/zookeys.441.7278

Molinatto, G., Demichelis, S., Bodino, N., Giorgini, M., Mori, N. & Bosco, D. (2020). Biology and prevalence in Northern Italy of *Verrallia aucta* (Diptera, Pipunculidae), a parasitoid of *Philaenus spumarius* (Hemiptera, Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe. *Insects, 11*(9), 607. doi: 10.3390/insects11090607

Platystomatidae

McMichael, B.A., Foote, B.A. & Bowker, B.D. (1990). Biology of *Rivellia melliginis* (Diptera: Platystomatidae), a consumer of the nitrogen-fixing root nodules of black locust (Leguminosae). *Annals of the Entomological Society of America, 83*(5), 967-974. doi: 10.1093/aesa/83.5.967

Polleniidae

Taleb, M., Tail, G. & Açıkgöz, H.N. (2022). Molecular identification of the potentially forensically relevant cluster flies *Pollenia rudis* (Fabricius) and *Pollenia vagabunda* (Meigen) (Diptera: Polleniidae) - non-recorded species in Algeria. *Forensic Sciences Research, 7*(1), 69-77. doi: 10.1080/20961790.2020.1857937

Vezsenyi, K.A., Langer, S.V., Samkari, B.A. & Beresford, D.V. (2022). The history and current state of cluster flies (Diptera: Polleniidae: *Pollenia*) in North America, with new Canadian provincial records. *Canadian Entomologist, 154*(1). doi: 10.4039/tce.2022.11

Psilidae

Degen, T., Städler, E. & Ellis, P.R. (1999). Host-plant susceptibility to the carrot fly, *Psila rosae*. 2. Suitability of various host species for larval development. *Annals of Applied Biology, 134*(1), 13-26. doi: 10.1111/j.1744-7348.1999.tb05231.x

Psychodidae

Ali, A. & Kok-Yokomi, M.L. (1991). Preliminary population assessment of *Psychoda alternata* (Diptera: Psychodidae) in soil irrigated with wastewater for turf cultivation. *The Florida Entomologist, 74*(4), 591-596. doi: 10.2307/3495415

Azmiera, N., Low, V.L. & Heo, C.C. (2021). Colonization of rabbit carcasses by drain fly larvae, *Psychoda* sp. (Diptera: Psychodidae): The first report. *Acta Parasitologica, 66*(2), 706-709. doi: 10.1007/s11686-020-00313-z

Ptychopteridae

Wolf, B., Zwick, P. & Marxsen, J. (1997). Feeding ecology of the freshwater detritivore *Ptychoptera paludosa* (Diptera, Nematocera). *Freshwater Biology, 38*(2), 375-386. doi: 10.1046/j.1365-2427.1997.00250.x

Rhagionidae

Imada, Y., Kato, M. & Brady, S. (2016). Bryophyte-Feeders in a basal Brachyceran lineage (Diptera: Rhagionidae: Spaniinae): Adult oviposition behavior and changes in the larval mouthpart morphology accompanied with the diet shifts. *PLoS One, 11*(11), p.e0165808. doi: 10.1371/journal.pone.0165808

Roberts, M.J. (1987). Structure of the mouthparts of the larvae of the flies *Rhagio* and *Sargus* in relation to feeding habits. *Journal of Zoology, 159*(3), 381-398. doi: 10.1111/j.1469-7998.1969.tb08453.x

Sarcophagidae

Blanckenhorn, W.U., Pemberton, A.J., Bussière, L.F., Roembke, J. & Floate, K.D. (2010). A review of the natural history and laboratory culture methods for the yellow dung fly, *Scathophaga stercoraria. Journal of Insect Science, 10*(11), 1-17. doi: 10.1673/031.010.1101

Buenaventura, E. (2021). Museomics and phylogenomics with protein-encoding ultraconserved elements illuminate the evolution of life history and phallic morphology of flesh flies (Diptera: Sarcophagidae). *BMC Ecology and Evolution, 21*(1), 70. doi: 10.1186/s12862-021-01797-7

Buenaventura, E., Szpila, K., Cassel, B.k., Wiegmann, B.M. & Pape, T. (2020). Anchored hybrid enrichment challenges the traditional classification of flesh flies (Diptera: Sarcophagidae). *Systematic Entomology, 45*(2), 281-301. doi: 10.1111/syen.12395

Buenaventura, E., Whitmore, D. & Pape, T. (2017). Molecular phylogeny of the hyperdiverse genus *Sarcophaga* (Diptera: Sarcophagidae), and comparison between algorithms for identification of rogue taxa. *Cladistics, 33*(2), 109-133. doi: 10.1111/cla.12161

Danyk, T., Mackauer, M. & Johnson, D.L. (2005). The influence of host suitability on the range of grasshopper species utilized by *Blaesoxipha atlanis* (Diptera: Sarcophagidae) in the field. *Bulletin of Entomological Research, 95*(6), 571-578. doi: 10.1079/BER2005388

Dufek, M.I., Mello-Patiu, C.A. & Mulieri, P.R. (2020). Inventory of Sarcophaginae (Diptera: Sarcophagidae) for the Humid Chaco, a poorly surveyed ecoregion of South America. *Journal of Natural History, 54*(5-6), 367-403. doi: 10.1080/00222933.2020.1764646

Miura, K. & Ohsaki, N. (2013). Mortality effects of the parasitoid flesh fly *Blaesoxipha japonensis* (Diptera: Sarcophagidae) in relation to body size of the adult grasshopper *Parapodisma tanbaensis* (Orthoptera: Catantopidae). *Applied Entomology and Zoology, 49*(1), 171-176. doi: 10.1007/s13355-013-0236-9

Miura, K. & Ohsaki, N. (2014). The cost of autotomy caused by the parasitoid fly *Blaesoxipha japonensis* (Diptera: Sarcophagidae): an interspecific comparison between two sympatric grasshopper host species. *Ecological Research, 30*(1), 33-39. doi: 10.1007/s11284-014-1202-x

Otronen, M. (1995). Energy reserves and mating success in males of the yellow dung fly, *Scathophaga stercoraria. Functional Ecology, 9*(4), 683-688. doi: 10.2307/2390161

Rees, N.E. (1970). Suitability of selected North American grasshopper species as hosts for the Eurasian parasites *Acridomyia sacharovi1* and *Blaesoxipha lineata2* (Diptera). *Annals of the Entomological Society of America, 63*(3), 901-903. doi: 10.1093/aesa/63.3.901

Ren, L., Shang, Y., Yang, L., Wang, S., Wang, X., Chen, S., Bao, Z., An, D., Meng, F., Cai, J. & Guo, Y. (2021). Chromosome-level de novo genome assembly of *Sarcophaga peregrina* provides insights into the evolutionary adaptation of flesh flies. *Molecular Ecology Resources, 21*(1), 251-262. doi: 10.1111/1755-0998.13246

Souza, C.M., Madeira-Ott, T., Masiero, F.S., Bunde, P.R.S., Ribeiro, G.A., Thyssen, P.J. & Geden, C. (2020). Synanthropy of Sarcophaginae (Diptera: Sarcophagidae) from Southern Brazil and its sanitary implications. *Journal of Medical Entomology, 58*(2), 913-920. doi: 10.1093/jme/tjaa243

Stamper, T., Dahlem, G.A., Cookman, C. & Debry, R.W. (2013). Phylogenetic relationships of flesh flies in the subfamily Sarcophaginae based on three mtDNA fragments (Diptera: Sarcophagidae). *Systematic Entomology, 38*(1), 35-44. doi: 10.1111/j.1365-3113.2012.00646.x

Valverde Castro, C., Buenaventura, E., Sánchez-Rodríguez, J.D. & Wolff, M. (2017). Flesh flies (Diptera: Sarcophagidae: Sarcophaginae) from the Colombian Guajira biogeographic province, an approach to their ecology and distribution. *Zoologia, 34,* 1-11. doi: 10.3897/zoologia.34.e12277

Scathophagidae

Blanckenhorn, W.U., Pemberton, A.J., Bussière, L.F., Roembke, J. & Floate, K.D. (2010). A review of the natural history and laboratory culture methods for the yellow dung fly, *Scathophaga stercoraria. Journal of Insect Science, 10*(11), 1-17. doi: 10.1673/031.010.1101

Kutty, S.N., Bernasconi, M.V., Šifner, F. & Meier, R. (2007). Sensitivity analysis, molecular systematics and natural history evolution of Scathophagidae (Diptera: Cyclorrhapha: Calyptratae). *Cladistics, 23*(1), 64-83. doi: 10.1111/j.1096-0031.2006.00131.x

Ovchinnikov, A.N. & Ovtshinnikova, O.G. (2013). The muscular system of the ovipositor of *Spaziphora hydromyzina* (Fallen) (Diptera, Scathophagidae). *Entomological Review*, *93*(8), 987-990. doi: 10.1134/S001387381308006X

Ozerov, A.L. (2017). A new species of the genus *Scathophaga* Meigen, 1803 (Diptera, Scathophagidae) from Yakutia (Russia). *Entomological Review, 97*(1), 132-135. doi: 10.1134/S0013873817010134

Rohner, P.T., Blanckenhorn, W.U. & Schäfer, M.A. (2017). Critical weight mediates sex-specific body size plasticity and sexual dimorphism in the yellow dung fly *Scathophaga stercoraria* (Diptera: Scathophagidae). *Evolution & Development, 19*(3), 147-156. doi: 10.1111/ede.12223

Webb, L., Beaumont, D.J., Nager, R.G. & McCracken, D.I. (2007). Effects of avermectin residues in cattle dung on yellow dung fly *Scathophaga stercoraria* (Diptera:

Scathophagidae) populations in grazed pastures. *Bulletin of Entomological Research, 97*(2), 129-138. doi: 10.1017/S0007485307004683

Sciaridae

Badshah, K., Ullah, F., Ahmad, B., Ahmad, S., Alam, S., Ullah, M., Jamil, M. & Sardar, S. (2021). Management of *Lycoriella ingenua* (Diptera: Sciaridae) on oyster mushroom (*Pleurotus ostreatus*) through different botanicals. *International Journal of Tropical Insect Science, 41*(2), 1435-1440. doi: 10.1007/s42690-020-00341-2

Cloonan, K.R., Andreadis, S.S. & Baker, T.C. (2016). Attraction of female fungus gnats, *Lycoriella ingenua*, to mushroom-growing substrates and the green mold *Trichoderma aggressivum. Entomologia Experimentalis et Applicata, 159*(3), 298-304. doi: 10.1111/eea.12439

Cloonan, K.R., Andreadis, S.S. & Baker, T.C. (2019). Little effect of delayed mating on fecundity or fertility of female fungus gnats *Lycoriella ingenua*. *Physiological Entomology, 44*(1), 60-64. doi: 10.1111/phen.12276

Gou, Y., Guo, S., Wang, G. & Liu, C. (1986). Effects of short-term heat stress on the growth and development of *Bradysia cellarum* and *Bradysia impatiens. Journal of Applied Entomology, 144*(4), 315-321. doi: 10.1111/jen.12733

Gou, Y., Quandahor, P., Zhang, Y., Coulter, J.A., Liu, C. & Gao, Y. (2020). Host plant nutrient contents influence nutrient contents in *Bradysia cellarum* and *Bradysia impatiens. PLoS One, 15*(4), p.e0226471. doi: 10.1371/journal.pone.0226471

Heiri, O. & Lotter, A.F. (2007). Sciaridae in lake sediments: indicators of catchment and stream contribution to fossil insect assemblages. *Journal of Paleolimnology, 38*(2), 183-189. doi: 10.1007/s10933-006-9068-8

Kecskeméti, S., Fail, J. & Geösel, A. (2018). Development of *Lycoriella ingenua* and *Bradysia impatiens* on different phases of agaricus composts. *Review on Agriculture and Rural Development, 7*(1-2), 55.

Kecskeméti, S., Szelényi, M.O., Erdei, A.L., Geösel, A., Fail, J. & Molnár, B.P. (2020). Fungal volatiles as olfactory cues for female fungus gnat, *Lycoriella ingenua* in the avoidance of Mycelia colonized compost. *Journal of Chemical Ecology, 46*(10), 917-926. doi: 10.1007/s10886-020-01210-5

Sawangproh, W. & Cronberg, N. (2016). Life history traits of the liverwort herbivore *Scatopsciara cunicularius* (Diptera: Sciaridae). *Annals of the Entomological Society of America, 109*(3), 343-349. doi: 10.1093/aesa/saw004

Sawangproh, W., Ekroos, J. & Cronberg, N. (2016). The effect of ambient temperature on larvae of *Scatopsciara cunicularius* (Diptera: Sciaridae) feeding on the thallose liverwort Marchantia polymorpha. *European Journal of Entomology, 113*(1), 259-264. Doi: 10.14411/eje.2016.030

Shamshad, A. (2010). The development of integrated pest management for the control of mushroom sciarid flies, *Lycoriella ingenua* (Dufour) and *Bradysia ocellaris* (Comstock), in cultivated mushrooms. *Pest Management Science, 66*(10), 1063-1074. doi: 10.1002/ps.1987

Shin, S., Jung, S., Heller, K., Menzel, F., Hong, T. K., Shin, J. S., Lee, S. H., Lee, H. & Lee, S. (1986). DNA barcoding of *Bradysia* (Diptera: Sciaridae) for detection of the immature stages on agricultural crops. *Journal of Applied Entomology, 139*(8), 638-645. doi: 10.1111/jen.12198

Shin, S., Jung, S., Menzel, F., Heller, K., Lee, H. & Lee, S. (2013). Molecular phylogeny of black fungus gnats (Diptera: Sciaroidea: Sciaridae) and the evolution of larval habitats. *Molecular Phylogenetics and Evolution, 66*(3), 833-846. doi: 10.1016/j.ympev.2012.11.008

Vilkamaa, P., Hippa, H. & Taylor, S.J. (2011). The genus *Camptochaeta* in Nearctic caves, with the description of *C. prolixa* sp. n. (Diptera, Sciaridae). *ZooKeys, 135*(135), 69-75. doi: 10.3897/zookeys.135.1624

Xi, L., Zhou, X., Liu, Y., Li, Y., Shi, H., Wu, H. & Huang, J. (1986). Behavioural response of the fungus gnat, *Bradysia impatiens* (Diptera: Sciaridae) towards certain edible mushrooms and saprophytic fungi. *Journal of Applied Entomology, 145*(5), 458-466. doi: 10.1111/jen.12857

Yan, H., Zhang, B., Wang, E., Xu, X. & Wei, G. (2021). Combining predatory mites and film mulching to control *Bradysia cellarum* (Diptera: Sciaridae) on Chinese chives, *Allium tuberosum. Experimental & Applied Acarology, 86*(1), 117-127. doi: 10.1007/s10493-021-00681-9

Zhu, G., Luo, Y., Xue, M., Zhao, H., Sun, X. & Wang, X. (2017). Effects of feeding on different host plants and diets on *Bradysia odoriphaga* population parameters and tolerance to heat and insecticides. *Journal of Economic Entomology, 110*(6), 2371-2380. doi: 10.1093/jee/tox242

<u>Sciomyzidae</u>

Bistline-East, A., Carey, J.G.J., Colton, A., Day, M.F. & Gormally, M.J. (2018). Catching flies with honey(dew): adult marsh flies (Diptera: Sciomyzidae) utilize sugary secretions for high-carbohydrate diets. *Environmental Entomology, 47*(6), 1632-1641. doi: 10.1093/ee/nvy155

Chapman, E.G., Foote, B.A., Malukiewicz, J. & Hoeh, W.R. (2006). Parallel evolution of larval morphology and habitat in the snail-killing fly genus *Tetanocera. Journal of Evolutionary Biology, 19*(5), 1459-1474. doi: 10.1111/j.1420-9101.2006.01132.x

Chapman, E.G., Przhiboro, A.A., Harwood, J.D., Foote, B.A. & Hoeh, W.R. (2012). Widespread and persistent invasions of terrestrial habitats coincident with larval feeding behavior transitions during snail-killing fly evolution (Diptera: Sciomyzidae). *BMC Evolutionary Biology, 12*(1), 175. doi: 10.1186/1471-2148-12-175

Murphy, W.L., Knutson, L.V., Chapman, E.G. Mc Donnell, R.J., Williams, C.D., Foote, B.A. & Vala, J. (2012). Key aspects of the biology of snail-killing Sciomyzidae flies. *Annual Review of Entomology, 57*(1), 425-447. doi: 10.1146/annurev-ento-120710-100702

Sepsidae

Rohner, P.T., Haenni, J., Giesen, A., Busso, J.P., Schäfer, M.A., Püchel-Wieling, F. & Blanckenhorn, W.U. (2019). Temporal niche partitioning of Swiss black scavenger flies in relation to season and substrate age (Diptera, Sepsidae). *Alpine Entomology, 3,* 1-10. doi: 10.3897/alpento.3.28366

Roy, J., Blanckenhorn, W.U. & Rohner, P.T. (2018). Largely flat latitudinal life history clines in the dung fly *Sepsis fulgens* across Europe (Diptera: Sepsidae). *Oecologia, 187*(3), 851-862. doi: 10.1007/s00442-018-4166-7

Zeender, V., Roy, J., Wegmann, A., Schäfer, M.A., Gourgoulianni, N., Blanckenhorn, W.U. & Rohner, P.T. (2019). Comparative reproductive dormancy differentiation in European black scavenger flies (Diptera: Sepsidae). *Oecologia, 189*(4), 905-917. doi: 10.1007/s00442-019-04378-0

Simuliidae

Adler, P.H., Huang, Y., Reeves, W.K., Kim, S.K., Otsuka, Y., Takaoka, H. & Laudet, V. (2013). Macrogenomic evidence for the origin of the black fly *Simulium suzukii* (Diptera: Simuliidae) on Okinawa Island, Japan. *PLoS One, 8*(8), p.e70765. doi: 10.1371/journal.pone.0070765

Al-Shaer, L., Pierce, A.K., Larson, D. & Hancock, R. (2015). Notes on facultative predation in *Prosimulium* larvae (Diptera: Simuliidae) in alpine and subalpine streams in Colorado. *Journal of the American Mosquito Control Association, 31*(1), 113-116. doi: 10.2987/14-6460.1

Brabrand, Å., Bremnes, T., Koestler, A. G., Marthinsen, G., Pavels, H., Rindal, E., Raastad, J. E., Saltveit, S. J. & Johnsen, A. (2014). Mass occurrence of bloodsucking blackflies in a regulated river reach: localization of oviposition habitat of *Simulium truncatum* using DNA barcoding: mass occurrence bloodsucking blackflies. *River Research and Applications, 30*(5), 602-608. doi: 10.1002/rra.2669

Finn, D.S. & Alder, P.H. (2006). Population genetic structure of a rare high-elevation black fly, *Metacnephia coloradensis*, occupying Colorado lake outlet streams. *Freshwater Biology, 51*(12), 2240-2251. doi: 10.1111/j.1365-2427.2006.01647.x

Gaudreau, C., Larue, B. & Charpentier, G. (2010). Molecular comparison of Quebec and Newfoundland populations of the blackfly, *Simulium vittatum*, species complex. *Medical and Veterinary Entomology, 24*(2), 214-217. doi: 10.1111/j.1365-2915.2009.00844.

Hall, R.J., Bailey, R. C. & Findeis, J. (1988). Factors affecting survival and cation concentration in the blackflies *Prosimulium fuscum/mixtum* and the mayfly *Leptophlebia cupida* during spring snowmelt. *Canadian Journal of Fisheries and Aquatic Sciences, 45*(12), 2123. doi: 10.1139/f88-247

Hunter, F.F., Sutcliffe, J.F. & Downe, A.E.R. (1993). Blood-feeding host preferences of the isomorphic species *Simulium venustum* and *S. truncatum. Medical and Veterinary Entomology, 7*(2), 105-110.

Iburg, J.P., Gray, E.W. & Noblet, R. (2016). Feeding behavior of *Simulium vittatum* larvae in response to various suspended materials: implications for control using biological larvicides. *Entomologia Experimentalis et Applicata, 159*(1), 70-76. doi: 10.1111/eea.12414

Karjalainen, A.K., Salmelin, J., Dimock, B. & Hintelmann, H. (2020). Blackfly larvae (*Simulium* spp.) can intensify methylmercury biomagnification in boreal food webs. *Water, Air, and Soil Pollution, 231*(1). doi: 10.1007/s11270-020-04717-5

Malmqvist, B. (1999). Life history of *Metacnephia lyra*, a blackfly highly characteristic of large Swedish river rapids at the time of maximum discharge (Diptera: Simuliidae). *Aquatic Insects, 21*(2), 89-98. doi: 10.1076/aqin.21.2.89.4534

Martin, P.J.S. & Edman, J.D. (1993). Assimilation rates of different particulate foods for *Simulium verecundum* (Diptera: Simuliidae). *Journal of Medical Entomology, 30*(4), 805-809. doi: 10.1093/jmedent/30.4.805

McCreadie, J.W. & Colbo, M.H. (1991). Spatial distribution patterns of larval cytotypes of the *Simulium venustum/verecundum* complex (Diptera: Simuliidae) on the Avalon Peninsula, Newfoundland: factors associated with occurrence. *Canadian Journal of Zoology, 69*(10), 2651-2659. doi: 10.1139/z91-373

McCreadie, J.W. & Colbo, M.H. (1993). Larval and pupal microhabitat selection by *Simulium truncatum* Lundström, *S. rostratum* Lundström, and *S. verecundum* AA (Diptera: Simuliidae). *Canadian Journal of Zoology, 71*(2), 358-367. doi: 10.1139/z93-050

McCreadie, J.W. & Colbo, M.H. (1993). Seasonal succession and spatial-temporal distribution patterns of six larval cytospecies of the *Simulium venustum/verecundum* complex (Diptera: Simuliidae). *Canadian Journal of Zoology, 71*(1), 116-124. doi: 10.1139/z93-017

McCreadie, J.W., Williams, R.H., Stutsman, S., Finn, D.S. & Adler, P.H. (2018). The influence of habitat heterogeneity and latitude on gamma diversity of the Nearctic Simuliidae, a ubiquitous group of stream-dwelling insects. *Insect Science, 25*(4), 712-720. doi: 10.1111/1744-7917.12442

Pangjanda, S. & Pramual, P. (2016). Trait-based and phylogenetic community ecology of black flies (Diptera: Simuliidae) in tropical streams of Thailand. *Hydrobiologia, 763*(1), 345-356. doi: 10.1007/s10750-015-2389-8

Sitarz, M., Buczek, A.M., Buczek, W., Buczek, A. & Bartosik, K. (2022). Risk of attacks by blackflies (Diptera: Simuliidae) and occurrence of severe skin symptoms in bitten patients along the Eastern border of the European Union. *International Journal of Environmental Research and Public Health, 19*(13), 7610. doi: 10.3390/ijerph19137610

Ya'cob, Z., Low, V.L., Tan, T.K., Noor-Izwan, A., Lourdes, E.Y., Ramli, R., Takaoka, H. & Adler, P.H. (1987). Sexually anomalous individuals of the black fly *Simulium trangense* (Diptera: Simuliidae) infected with mermithid parasites (Nematoda: Mermithidae). *Parasitology Research, 120*(5), 1555-1561. doi: 10.1007/s00436-021-07087-x

Zhang, Y. (2006). Balancing food availability and hydrodynamic constraint: phenotypic plasticity and growth in *Simulium noelleri* blackfly larvae. *Oecologia, 147*(1), 39-46. doi: 10.1007/s00442-005-0243-9

Zettler, J.A., Adler, P.H. & McCreadie, J.W. (1998). Factors influencing larval color in the *Simulium vittatum* Complex (Diptera: Simuliidae). *Invertebrate Biology, 117*(3), 245-252. doi: 10.2307/3226989

Sphaeroceridae

Marshall, S.A. (1982). A revision of the Nearctic *Leptocera (Thoracochaeta Duda*) (Diptera: Sphaeroceridae). *Canadian Entomologist, 114*(1), 63-78. doi: 10.4039/Ent11463-1

Pitkin, B.R. (1986). Bait, habitat preferences and the phenology of some lesser dung flies (Diptera: Sphaeroceridae) in Britain. *Journal of Natural History, 20*(6), 1283-1295. doi: 10.1080/00222938600770851

Stratiomyidae

Fontenelle, J.C.R., Viana-Silva, F.E.C., Martins, R.P. & Del-Claro, K. (2012). Use of plant resources by *Merosargus* (Diptera, Stratiomyidae, Sarginae) larvae. *Psyche, 2012*, 1-10. doi: 10.1155/2012/690203

Gligorescu, A., Toft, S., Hauggaard-Nielsen, H., Axelsen, J.A., & Nielsen, S. A. (2018). Development, metabolism and nutrient composition of black soldier fly larvae (*Hermetia illucens*; Diptera: Stratiomyidae) in relation to temperature and diet. *Journal of Insects as Food and Feed, 4*(2), 123-133. doi: 10.3920/JIFF2017.0080

Syrphidae

Branquart, E. & Hemptinne, J. (2008). Selectivity in the exploitation of floral resources by hoverflies (Diptera: Syrphinae). *Ecography, 23*(6), 732-742. doi: 10.1111/j.1600-0587.2000.tb00316.x

Budenberg, W.J. & Powell, W. (1992). The role of honeydew as an ovipositional stimulant for two species of syrphids. *Entomologia Experimentalis et Applicata, 64*(1), 57-61. doi: 10.1111/j.1570-7458.1992.tb01594.x

Campbell, D.R., Bischoff, M., Lord, J.M. & Robertson, A.W. (2010). Flower color influences insect visitation in alpine New Zealand. *Ecology, 91*(9), 2638-2649. doi: 10.1890/09-0941.1

Campoy, A., Aracil, A., Pérez-Bañón, C. & Rojo, S. (2020). An in-depth study of the larval head skeleton and the external feeding structures related with the ingestion of food particles by the eristaline flower flies *Eristalis tenax* and *Eristalinus aeneus. Entomologia Experimentalis et Applicata, 168*(10), 783-798. doi: 10.1111/eea.12974

Dziock, F. (2005). Evolution of prey specialization in aphidophagous syrphids of the genera *Melanostoma* and *Platycheirus* (Diptera: Syrphidae) 1. Body size, development and prey traits. European Journal of Entomology, 102(3), 413-421. doi: 10.14411/eje.2005.059

Francuski, L., Matić, I., Ludoški, J. & Milankov, V. (2011). Temporal patterns of genetic and phenotypic variation in the epidemiologically important drone fly, *Eristalis tenax. Medical and Veterinary Entomology, 25*(2), 135-147. doi: 10.1111/j.1365-2915.2011.00956.x

Gojkovic, N., Francuski, L., Ludoski, J. & Milankov, V. (2020). DNA barcode assessment and population structure of aphidophagous hoverfly *Sphaerophoria scripta*: Implications for conservation biological control. *Ecology and Evolution, 10*(7), 9428-9443. doi: 10.1002/ece3.6631

Jamali, R.A., Memom, N., Shah, M.A., Khan, K. & Ansari, A. (2018). Prevalence of aphidophagous hoverflies (Syrphidae: Syrphinae) in relation to their prey, green aphids on Brassica in Dadu. *Journal of Animal and Plant Sciences, 28*(5), 1447.

Jiang, S., Li, H., He, L. & Wu, K. (2022). Population fitness of *Eupeodes corollae* Fabricius (Diptera: Syrphidae) feeding on different species of aphids. *Insects, 13*(6), 494. doi: 10.3390/insects13060494

Jovičić, S., Burgio, G., Diti, I., Krašić, D., Markov, Z., Radenković, S. & Vujić, A.(2017). Influence of landscape structure and land use on M*erodon* and *Cheilosia* (Diptera: Syrphidae): contrasting responses of two genera. *Journal of Insect Conservation, 21*(1), 53-64. doi: 10.1007/s10841-016-9951-1

Khaganinia, S. & Kazerani, F. (2014). A review of the genus *Cheilosia* (Diptera, Syrphidae) from Iran. *Vestnik Zoologii, 48*(5), 401-410. doi: 10.2478/vzoo-2014-0048

Koval, A.G., Guseva, O.G. & Shpanev, A.M. (2018). Hoverflies (Diptera, Syrphidae) in Agrolandscapes pf St. Petersburg and Leningrad Province. *Entomological Review, 98*(6), 702-708. doi: 10.1134/S0013873818060064

Krivosheina, N.P. (2021). Ecological relations of the hoverfly larvae (Diptera, Syrphidae, Eristalinae) bark inhabitants with xylobiont insects. *Biology Bulletin of the*

Russian Academy of Sciences, 47(6), 605-616. doi: 10.1134/S1062359020060096

Lillo, I., Perez-Bañón, C. & Rojo, S. (2021). Life cycle, population parameters, and predation rate of the hover fly *Eupeodes corollae* fed on the aphid *Myzus persicae. Entomologia Experimentalis et Applicata, 169*(11), 1027-1038. doi: 10.1111/eea.13090

Mengual, X., Stahls, G. & Rojo, S. (2008). First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial COI and nuclear 28S rRNA genes: conflict and congruence with the current tribal classification. *Cladistics, 24*(4), 543-562. doi: 10.1111/j.1096-0031.2008.00200.x

Mielczarek, L.E. (2013). *Melangyna soszynskii* sp. n. (Diptera: Syrphidae) from the Sayan Mountains. *Polskie Pismo Entomologiczne, 82*(4), 339-342. doi: 10.2478/v10200-012-0047-2

Milic, D., Radenkovic, S., Radisic, D., Andric, A., Nikolic, T., Vujic, A. & Silva, D.P. (2019). Stability and changes in the distribution of *Pipiza* hoverflies (Diptera, Syrphidae) in Europe under projected future climate conditions. *PLoS One, 14*(9), p.e0221934. doi: 10.1371/journal.pone.0221934

Nedeljković, Z., Ačanski, J., Đan, M., Obreht-Vidaković, D., Ricarte, A., Vujić, A. & Biesmeijer, J.C. (2015). An integrated approach to delimiting species borders in the genus *Chrysotoxum* Meigen, 1803 (Diptera: Syrphidae), with description of two new species. *Contributions to Zoology, 84*(4), 285-304. doi: 10.1163/18759866-08404002

Nedeljković, Z., Ačanski, J., Vujić, A., Obreht, D., Đan, M., Ståhls, G. & Radenković, S. (2013). Taxonomy of *Chrysotoxum festivum* Linnaeus, 1758 (Diptera: Syrphidae) - an integrative approach. *Zoological Journal of the Linnean Society, 169*(1), 84-102. doi: 10.1111/zoj.12052

Reemer, M. & Rotheray, G.E. (2009). Pollen feeding larvae in the presumed predatory syrphine genus *Toxomerus* Macquart (Diptera, Syrphidae). *Journal of Natural History, 43*(15-16), 939-949. doi: 10.1080/00222930802610576

Rizza, A., Campobasso, G., Dunn, P.H. & Stazi, M. (1988). *Cheilosia corydon* (Diptera: Syrphidae), a candidate for the biological control of musk thistle in North America. *Annals of the Entomological Society of America, 81*(2), 225-232. doi: 10.1093/aesa/81.2.225

Rotheray, E. L., Lepais, O., Nater, A., Krützen, M., Greminger, M., Goulson, D. & Bussière, L. F. (2012). Genetic variation and population decline of an

endangered hoverfly *Blera fallax* (Diptera: Syrphidae). *Conservation Genetics, 13*(5), 1283-1291. doi: 10.1007/s10592-012-0371-9

Tijana, N., D.R., Dubravka, M., V.M., Sonja, T., Snežana, J., Smiljka, S. & A.V. (2013). Models of the potential distribution and habitat preferences of the genus *Pipiza* (Syrphidae: Diptera) on the Balkan peninsula. *Archives of Biological Sciences, 65*(3), 1037-1052. doi: 10.2298/ABS1303037N

Vanhaelen, N., Francis, F. & Haubruge, E. (2004). Purification and characterization of glutathione S-transferases from two syrphid flies (*Syrphus ribesii* and *Myathropa florae*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 137*(1), 95-100. doi: 10.1016/j.cbpc.2003.10.006

Yan, Y., Hu, L. & National Library of Medicine. (2021). The complete mitochondrial genome of *Platycheirus albimanus* (Diptera: Syrphidae: Syrphinae) and phylogenetic analysis of the Syrphidae. *Mitochondrial DNA, 6*(2), 528-530. doi: 10.1080/23802359.2021.1872455.

Tabanidae

Kelly-Hope, L., Paulo, R., Thomas, B., Brito, M., Unnasch, T.R. & Molyneux, D. (2017). Loa loa vectors *Chrysops* spp.: perspectives on research, distribution, bionomics, and implications for elimination of lymphatic filariasis and onchocerciasis. *Parasites & Vectors, 10*(1), 172. doi: 10.1186/s13071-017-2103-y

Lall, S.B. (1969). Effect of starvation on the sensitivity of tarsal and labellar contact chemoreceptors in *Hybomitra lasiophthalma* (Diptera: Tabanidae). *Annals of the Entomological Society of America, 62*(5), 1217. doi: 10.1093/aesa/62.5.1217

Teskey, H.J., Shemanchuk, J.A. & Weintraub, J. (1987). *Hybomitra agora*, a new species of Tabanidae (Diptera) from Western North America. *Canadian Entomologist, 119*(12), 1117-1122. doi: 10.4039/Ent1191117-12

Tachinidae

Al-Dobai, S., Reitz, S. & Sivinski, J. (2012). Tachinidae (Diptera) associated with flowering plants: estimating floral attractiveness. Biological Control, 61(3), 230-239. doi: 10.1016/j.biocontrol.2012.02.008

Bora, D. & Deka, B. (2014). Role of visual cues in host searching behaviour of *Exorista sorbillans* Widemann, a parasitoid of Muga Silk Worm, *Antheraea assama*

Westwood. *Journal of Insect Behavior, 27*(1), 92-104. doi: 10.1007/s10905-013-9409-1

Cerretti, P., O'Hara, J.E., Wood, D.M., Shima, H., Inclan, D.J. & Stireman III, J.O. (2014). Signal through the noise? Phylogeny of the Tachinidae (Diptera) as inferred from morphological evidence. Systematic Entomology, 39(2), 335-353. doi: 10.1111/syen.12062

Dindo, M.L., Rezaei, M. & De Clercq, P. (2019). Improvements in the rearing of the Tachinid parasitoid *Exorista larvarum* (Diptera: Tachinidae): Influence of adult food on female longevity and reproduction capacity. *Journal of Insect Science, 19*(2). doi: 10.1093/jisesa/iey122

Foerster, L.A. & Doetzer, A.K. (2002). Host instar preference of *Peleteria robusta* (Wiedman) (Diptera: Tachinidae) and development in relation to temperature. *Neotropical Entomology, 31*(3), 405-409. doi: 10.1590/S1519-566X2002000300009

Kara, K., Tschosnig, H. & Atay, T. (2020). Checklist of Turkish Tachinidae (Insecta, Diptera) with new records. *Journal of the Entomological Research Society*, 22(2), 163-190.

Martel, V., Thireau, J. & Régnière, J. (2022). Manual inoculation of host larvae with first instar maggots as a rearing technique for the larval parasitoid *Actia interrupta* (Diptera: Tachinidae). *Biocontrol Science and Technology, 32*(1), 110-113. doi: 10.1080/09583157.2021.1967292

Regniere, J., Thireau, J., Saint-Amant, R. & Martel, V. (2021). Modeling climatic influences on three parasitoids of low-density spruce budworm populations. Part 3: *Actia interrupta* (Diptera: Tachinidae). *Forests, 12*(11), 1471. doi: 10.3390/f12111471

Sahebari, F.S., Khaghaninia, S. & Talebi, A.A. (2018). New data of the subfamily Tachininae (Diptera: Tachinidae) from north-western Iran. *Zoology and Ecology*, 28(3), 252-258. doi: 10.1080/21658005.2018.1490107

Scaccini, D. (2018). Remarks on the biology of *Dinera ferina* (Diptera Tachinidae) as parasitoid of the two Italian *Platycerus* species (Coleoptera Lucanidae). *Bulletin of Insectology, 71*(1), 39.

Tachi, T. & Shima, H. (2010). Molecular phylogeny of the subfamily Exoristinae (Diptera, Tachinidae), with discussions on the evolutionary history of female oviposition strategy. *Systematic Entomology, 35*(1), 148-163. doi: 10.1111/j.1365-3113.2009.00497.x

Tipulidae

Bearup, D., Petrovskii, S. Blackshaw, R. & Hastings, A. (2013). Synchronized dynamics of *Tipula paludosa* metapopulation in a Southwestern Scotland agroecosystem: Linking pattern to process. *The American Naturalist, 182*(3), 393-409. doi: 10.1086/671162

Blackshaw, R.P. & Moore, J.P. (2011). Within-generation dynamics of leatherjackets (*Tipula paludosa* Meig.) *Journal of Applied Entomology, 136*(8), 605-613. doi: 10.1111/j.1439-0418.2011.01696.x

Cheever, B.M. & Webster, J.R. (2014). Effects of consumers and nitrogen availability on heterotrophic microbial activity during leaf decomposition in headwater streams. *Freshwater Biology, 59*(8), 1768-1780. doi: 10.1111/fwb.12383

Cook, D.M., Henriksen, E.D., Upchurch, R. & Peterson, J.B.D. (2020). Isolation of polymer-degrading bacteria and characterization of the hindgut bacterial community from the detritus-feeding larvae of *Tipula abdominalis* (Diptera: Tipulidae). *Applied and Environmental Microbiology, 73*(17), 5683-5686. doi: 10.1128/AEM.00213-07

Jo, J. (2017). Description of larval and pupal stages of *Tipula (Nippotipula) sinica* (Diptera, Tipulidae) from South Korea with ecological notes. *Animal Systematics, Evolution and Diversity, 33*(1), 56-59. doi: 10.5635/ASED.2017.33.1.036

Kania, I. & Nel, A. (2013). New fossil Tipulidae from the volcano-sedimentary latest Oligocene of Bes-Konak (Turkey). *Polish Journal of Entomology, 82*(4), 327-338. doi: 10.2478/v10200-012-0046-3

Meats, A. (1975). The development dynamics of *Tipula paludosa* and the relation of climate to its growth pattern, flight season and geographical distribution. *Oecologia, 19*(2), 177-128. doi: 10.1007/BF00369096

Peck, D.C. & Olmstead, D. (2009). Invasive *Tipula* (Diptera: Tipulidae) in turfgrass of the Northeast United States: Geographic distribution and local incidence three years after detection. *Journal of Economic Entomology, 102*(2), 652-658. doi: 10.1603/029.102.0225

Peck, D.C., Olmstead, D. & Morales, A. (2008). Application timing and efficacy of alternatives for the insecticidal control of *Tipula paludosa* Meigen (Diptera: Tipulidae), a new invasive pest of turf in the northeastern United States. *Pest Management Science, 64*(10), 989-1000. doi: 10.1002/ps.1592

Petersen, M.J. (2013). Evidence of a climatic niche shift following North American introductions of two crane flies (Diptera; genus *Tipula*). *Biological Invasions, 15*(4), 885-897. doi: 10.1007/s10530-012-0337-3

Podeniene, V., Gelhaus, J.K. & Yadamsuren, O. (2006). The last instar larvae and pupae of *Tipula (Arctotipula)* (Diptera, Tipulidae) from Mongolia. *Proceedings of the Academy of Natural Sciences of Philadelphia, 155*(1), 79-105. doi: 10.1635/i0097-3157-155-1-79.1

Pritchard, G. (1983). Biology of Tipulidae. *Annual Review of Entomology, 28*(1), 1-22. doi: 10.1146/annurev.en.28.010183.000245

Rao, S., Liston, A., Crampton, L. & Takeyasu, J. (2006). Identification of larvae of exotic *Tipula paludosa* (Diptera: Tipulidae) and *T. oleracea* in North America using mitochondrial cytB sequences. *Annals of the Entomological Society of America, 99*(1), 33-40. doi: 10.1603/0013-8746(2006)099[0033:IOLOET]2.0.CO2

Rogers, T.E. & Doran-Peterson, J. (2010). Analysis of cellulolytic and hemicellulolytic enzyme activity within the *Tipula abdominalis* (Diptera: Tipulidae) larval gut and characterization of *Crocebacterium ilecola* gen. nov., sp. nov., isolated from the *Tipula abdominalis* larval hindgut. *Insect Science, 17*(3), 291-302. doi: 10.1111/j.1744-7917.2010.01346.x

Salmela, J. & Petrašiūnas, A. (2014). Checklist of the infraorder Tipulomorpha (Trichoceridae, Tipuloidea) (Diptera) of Finland. *ZooKeys, 441*(441), 21-36. doi: 10.3897/zookeys.441.7533

Smith, R.M., Young, M.R. & Marquiss, M. (2008). Bryophyte use by an insect herbivore: does the crane-fly *Tipula montana* select food to maximise growth? *Ecological Entomology, 26*(1), 83-90. doi: 10.1046/j.1365-2311.2001.00297.x

Smith-Herron, A.J., Cook, J.L. & Cook, T.J. (2014). *Brevispina tasii* (Brevispinidae) n. fam., n. gen., n. sp (Apicomplexa: Eugregarinidae: Stenophoroidea) parasitizing the crane fly *Tipula umbrosa* (Diptera: Tipulidae: Tipulinae) from East Texas leaf litter. *Comparative Parasitology, 81*(1), 1-9. doi: 10.1654/4656.1

<u>Trichoceridae</u>

Laurence, B.R. (1955). On the life history of *Trichocera saltator* (Harris) (Diptera, Trichoceridae). *Proceedings of the Zoological Society of London, 126*(2), 235-244. doi: 10.1111/j.1096-3642.1956.tb00434.x

Ulidiidae

Goyal, G., Nuessly, G.S., Seal, D.R., Steck, G.J., Capinera, J.L. & Boote, K.J. (2017). Developmental studies of maize-infesting picture-winged flies (Diptera: Ulidiidae). *Environmental Entomology, 46*(4), 946-953. doi: 10.1093/ee/nvx082

Kaplin, V. G., Nartshuk, E. P., Volodina, I. A. & Kuryanovich, A. A. (2022). A new alfalfa nectarophage from the genus *Timia* Wiedemann (Diptera, Ulidiidae) in the forest-steppe of the Middle Volga Region. *Entomological Review, 101*(9), 1304-1309. doi: 10.1134/S0013873821090086

Owens, D., Larsen, N. & Nuessly, G.S. (2017). Post-harvest crop destruction effects on picture-winged fly (Diptera: Ulidiidae) emergence. *The Florida Entomologist, 100*(2), 422-425. doi: 10.1653/024.100.0203

Xylophagidae

Lee, J., Kim, Y. & Suh, S.J. (2022). First record of the awl fly genus *Xylophagus* (Diptera: Xylophagidae) from Korea. *Animal Systematics, Evolution and Diversity, 38*(1), 34-38.

Nagatomi, A. & Liu, N. (1995). Spermatheca and female terminalia of Pantophthalmidae and xylophagidae s. lat. (Diptera). *Annals of the Entomological Society of America, 88*(5), 603-626. doi: 10.1093/aesa/88.5.603

Appendix B: Results of 4% Clustering Analysis

Population Genetic Structure Measures

The full analysis was completed using a 4% clustering threshold and the single-linkage method instead of BINs (Table B1a). Both the clustering threshold and method can be adjusted by the user. 18 clusters across two geographic regions were included in the analysis for Greenland. The results reflected those of the BIN analysis. All of the Nei F_{ST} values for the clusters were low, showing low levels of genetic differentiation. The results for Hedrick's G_{ST} and Jost's D were more variable, with five clusters having a higher value (above 0.5) and 13 having a lower value for both measures. The AMOVA analysis revealed low levels of variation between the regions for all clusters. From the figures produced by the mismatch distribution analysis, 11 clusters' populations appeared to be unimodal, and seven appeared to be bimodal. However, the dip test statistic indicated that there was no significant multimodality.

For the Canada dataset, 602 clusters across 32 regions/polygons were included (Table B1b). The results were similar to those observed in the original BIN-based analysis. Most Nei F_{ST} values were below 0.1, indicating low levels of genetic differentiation. Most Hedrick G_{ST} and Jost's D values were above 0.8, indicating higher levels of genetic differentiation. The AMOVA for most clusters revealed low levels of variation between regions, with most phi values being below 0.1. The mismatch analysis revealed few significant signs of multimodality, suggesting that the populations are still expanding.

Comparison Across Traits

For the Greenland clustered analysis, nine were terrestrial and nine were aquatic. For larval diet, nine were phytophagous, five were saprophagous, three were fungivores and one was predaceous. For adult diet, three were saprophagous, two were phytophagous, one was predaceous, and 12 were non-feeding. Similar to the BIN analysis, there was no significant relationship or noticeable trend in either the ANOVA or PGLS analysis, except for a slight trend toward higher measures in saprophagous clusters in the PGLS analysis. The traits and measures were also compared to the range size of each clusters. For both the BIN and cluster analysis, there were no significant relationships between range size and population genetic structure measures or traits for either the ANOVA or PGLS analysis.

For the Canada analysis, 141 clusters were aquatic, 52 were semi-aquatic, and 409 were terrestrial. For adult diet, two clusters were fungivores, 239 were non-feeding, 28 were both phytophagous and predaceous, 93 were saprophagous, seven were both fungivores and phytophagous, 147 were phytophagous, 83 were predaceous and three were both saprophagous and phytophagous. For larval diet, 33 were fungivores, 20 were omnivores, 178 were phytophagous, 118 were predaceous, three were both saprophagous and phytophagous, 17 were both fungivores and phytophagous, 32 were

parasitic, and 201 were saprophagous. The relationship to traits did not differ from the BIN analysis. There was a significant relationship between the indicators of population genetic structure and habitat. An aquatic lifestyle had significantly higher levels of genetic differentiation than other habitats from the multiple regression using Jost's D and $G_{ST,}$ and terrestrial lifestyles had significantly lower levels of genetic differentiation (Table B2). F_{ST} showed the opposite trend, suggesting that aquatic lifestyles have significantly lower genetic differentiation and terrestrial lifestyles have significantly higher. Larval diet was significantly related to indicators of population genetic structure. Having a parasitic larval stage appears to result in significantly lower levels of genetic differentiation than clusters with other diets for Jost's D and G_{ST}. The F_{ST} results differed, suggesting that higher levels of genetic differentiation were observed in clusters with a phytophagous larval stage, a saprophagous larval stage, and those whose larval stage was a mix of fungivore and phytophagous.

Unlike the BIN analysis, no relationship was found between indicators of population genetic structure and distance. However, F_{ST} was still positively associated with both latitude and longitude. There was also a relationship between indicators of population genetic structure and the number of cells in which the cluster was found and the number of records of each cluster. As the number of cells increased, so did the genetic differentiation. Oppositely, as the number of records increased, the genetic differentiation decreased.

Table B1: Table outlining the population genetic structure measures using clustering at a 4% distance threshold from a) Greenland and b) Canada.

a) Greenland

Cluster	Species Name	Region/Cell	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
11	Spilogona sanctipauli	7721, 7478	0.011	0.18	0.16	-0.0017	0.46	0.044	0.99
12	Protophormia terraenovae	7721, 7478	0.037	0.84	0.83	0.053	0.008	0.038	1
15	Phytomyza puccinelliae	7721, 7478	0.0098	0.28	0.27	0.006	0.2	0.033	1
21	Limnophyes ninae	7721, 7478	0.0062	0.36	0.35	0.0065	0.009	0.086	0.78
24	Limnophyes pumilio	7721, 7478	0.025	0.89	0.89	0.038	0.001	0.05	0.99
25	Limnophyes brachytomus	7721, 7478	0.0015	0.1	0.099	0.00062	0.21	0.057	0.97

Cluster	Species Name	Region/Cell	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
60	Lycoriella riparia	7721, 7478	0.0084	0.23	0.22	-0.0051	0.55	0.1	0.45
66	Lycoriella flavipeda	7721, 7478	0.0027	0.16	0.15	0.0047	0.001	0.016	1
70	Camptochaeta delicata	7721, 7478	0.0081	0.25	0.24	0.0087	0.027	0.025	1
84	Corynoneura arctica	7721, 7478	0.0072	0.3	0.29	0.011	0.002	0.077	0.71
92	Nephrotoma lundbecki	7721, 7478	0.029	0.38	0.34	0.024	0.13	0.078	0.9
94	Cricotopus patens	7721, 7478	0.05	0.75	0.72	0.054	0.048	0.12	0.24
104	Smittia extrema	7721, 7478	0.016	0.47	0.45	0.028	0.001	0.044	1
114	Megaselia cirriventris	7721, 7478	0.023	0.69	0.68	0.041	0.001	0.033	1

Cluster	Species Name	Region/Cell	F_{ST}	G_{ST}	Jost's D	AMOVA Phi Value	AMOVA p-value	Dip Test Value	Dip Test p-value
137	Paraphaenocladius impensus	7721, 7478	0.0061	0.2	0.19	0.01	0.003	0.025	1
148	Diamesa arctica	7721, 7478	0.011	0.78	0.78	0.0083	0.012	0.098	0.41
150	Diamesa chorea	7721, 7478	0.014	0.23	0.2	0.021	0.007	0.04	1
181	Drymeia groenlandica	7721, 7478	0.042	0.43	0.38	0.063	0.002	0.025	1

b) Canada

Measure	Max	Min
Global F_{ST}	0.4	0.002
Global G_{ST}	1	0.031
Jost's D	1	0.019
AMOVA Phi Value	0.58	-0.03
AMOVA p-value	1	0.001
Dip Test Value	0.5	0.031
Dip Test p-value	1	0.004

Table B2: Table showing the results from the multiple regression comparing population genetic structure measures to biological traits and other important factors for Canada using a 4% clustering threshold for a) Jost's D, b) G_{ST} and c) F_{ST}. Significant values are bolded.

a) Jost's D

Trait/Factors	t-value	p-value
Aquatic	**4.34**	**1.66e-05**
Semi-Aquatic	-0.95	0.34
Terrestrial	**-2.22**	**0.027**
Adult Diet: Fungivore & Phytophagous	0.006	0.99
Adult Diet: Non-Feeding	0.73	0.47
Adult Diet: Phytophagous	0.8	0.43
Adult Diet: Predaceous	0.97	0.33
Adult Diet: Phytophagous & Predaceous	1.092	0.28
Adult Diet: Saprophagous	0.45	0.65
Adult Diet: Saprophagous & Phytophagous	-0.31	0.76
Larval Diet: Fungivore & Phytophagous	-1.08	0.28

Trait/Factors	t-value	p-value
Larval Diet: Omnivores	-1.49	0.14
Larval Diet: Parasitic	**-2.07**	**0.039**
Larval Diet: Phytophagous	-1.13	0.26
Larval Diet: Predaceous	-1.43	0.15
Larval Diet: Saprophagous	-1.26	0.21
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	1.44	0.15
Number of Records	-0.44	0.66
Max Distance	1.57	0.12
Longitude Coordinate	-0.27	0.78
Latitude Coordinate	0.48	0.63

b) G_{ST}

Trait/Factors	t-value	p-value
Aquatic	**4.53**	**7.1e-06**
Semi-Aquatic	-0.802	0.42
Terrestrial	**-2.051**	**0.041**
Adult Diet: Fungivore & Phytophagous	0.009	0.99
Adult Diet: Non-Feeding	0.75	0.45
Adult Diet: Phytophagous	0.804	0.42
Adult Diet: Predaceous	0.99	0.32
Adult Diet: Phytophagous & Predaceous	1.12	0.26
Adult Diet: Saprophagous	0.47	0.64
Adult Diet: Saprophagous & Phytophagous	-0.22	0.83
Larval Diet: Fungivore & Phytophagous	-1	0.32
Larval Diet: Omnivores	-1.44	0.15
Larval Diet: Parasitic	**-1.98**	**0.049**

Trait/Factors	t-value	p-value
Larval Diet: Phytophagous	-1.02	0.31
Larval Diet: Predaceous	-1.31	0.19
Larval Diet: Saprophagous	-1.18	0.24
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	1.58	0.11
Number of Records	-0.57	0.57
Max Distance	1.57	0.12
Longitude Coordinate	-0.19	0.85
Latitude Coordinate	0.61	0.54

c) F_{ST}

Trait/Factors	t-value	p-value
Aquatic	**-2.98**	**0.00304**
Semi-Aquatic	0.63	0.53
Terrestrial	**3.14**	**0.0018**
Adult Diet: Fungivore & Phytophagous	-0.47	0.64
Adult Diet: Non-Feeding	0.73	0.46
Adult Diet: Phytophagous	0.76	0.45
Adult Diet: Predaceous	0.99	0.32
Adult Diet: Phytophagous & Predaceous	1.088	0.28
Adult Diet: Saprophagous	0.45	0.65
Adult Diet: Saprophagous & Phytophagous	0.99	0.32
Larval Diet: Fungivore & Phytophagous	**3.39**	**0.00075**
Larval Diet: Omnivores	0.202	0.84
Larval Diet: Parasitic	1.6	0.11

Trait/Factors	t-value	p-value
Larval Diet: Phytophagous	**2.601**	**0.0095**
Larval Diet: Predaceous	1.808	0.071
Larval Diet: Saprophagous	**3.23**	**0.0013**
Larval Diet: Saprophagous & Phytophagous	N/A	N/A
Number of Cells	**3.49**	**0.00052**
Number of Records	**-3.32**	**0.00097**
Max Distance	0.56	0.58
Longitude Coordinate	**3.36**	**0.00083**
Latitude Coordinate	**6.78**	**2.98e-11**