Assessment of targeted enrichment locus capture across time and museums using odonate specimens

Aaron Goodman, Ethan Tolman, Rhema Uche-Dike1, John Abbott, Jesse W. Breinholt, Seth Bybee, Paul B. Frandsen, J. Stephen Gosnell, Rob Guralnick, Vincent J. Kalkman, Manpreet Kohli1, Judicael Fomekong Lontchi, Pungki Lupiyaningdyah, Lacie Newton, Jessica L. Ware

DOI:
https://doi.org/10.1093/isd/ixad011

Insect Systematics and Diversity, 7(3), 5; 2023, 1–9
https://doi.org/10.1093/isd/ixad011
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular phylogenetics, phylogenomics, and phylogeography

Assessment of targeted enrichment locus capture across time and museums using odonate specimens

Aaron Goodman[1,2,*,†], Ethan Tolman[1,2,†], Rhema Uche-Dike[1,2,†], John Abbott[3], Jesse W. Breinholt[4,5], Seth Bybee[6], Paul B. Frandsen[7], J. Stephen Gosnell[2,8], Rob Guralnick[5], Vincent J. Kalkman[9], Manpreet Kohli[1,8], Judicael Fomekong Lontchi[6], Pungki Lupiyaningdyah[6,10], Lacie Newton[1], Jessica L. Ware[1]

[1]Division of Invertebrate Zoology, American Museum of Natural History, New York City, NY 10024, USA, [2]City University of New York, Graduate Center, New York City, NY 10016, USA, [3]Alabama Museum of Natural History and UA Museums Department of Research and Collections, The University of Alabama, Tuscaloosa, AL 35487, USA, [4]Rapid Genomics, Gainesville, FL 32601, USA, [5]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA, [6]Department of Biology and Monte L. Bean Museum, Brigham Young University, Provo, UT 84602, USA, [7]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84602, USA, [8]Department of Natural Sciences, Baruch College, City University of New York, 17 Lexington Avenue, New York, NY 10010, USA, [9]Naturalis Biodiversity Center, 2300 RA Leiden, The Netherlands, [10]Research Center for Biosystematics and Evolution, Museum Zoologicum Bogoriense, BRIN, Cibinong 16911, Indonesia, *Corresponding author, mail: agoodman@amnh.org

[†]Equal first authors.
Subject Editor: Kazunori Yoshizawa

Received on 7 December 2022; revised on 10 May 2023; accepted on 29 May 2023

The use of gDNAs isolated from museum specimens for high throughput sequencing, especially targeted sequencing in the context of phylogenetics, is a common practice. Yet, little understanding has been focused on comparing the quality of DNA and results of sequencing museum DNAs. Dragonflies and damselflies are ubiquitous in freshwater ecosystems and are commonly collected and preserved insects in museum collections hence their use in this study. However, the history of odonate preservation across time and museums has resulted in wide variability in the success of viable DNA extraction, necessitating an assessment of their usefulness in genetic studies. Using Anchored Hybrid Enrichment probes, we sequenced DNA from samples at 2 museums, 48 from the American Museum of Natural History (AMNH) in NYC, USA and 46 from the Naturalis Biodiversity Center (RMNH) in Leiden, Netherlands ranging from global collection localities and across a 120-year time span. We recovered at least 4 loci out of an >1,000 locus probe set for all samples, with the average capture being ~385 loci (539 loci on average when a clade of ambiguous taxa omitted). Neither specimen age nor size was a good predictor of locus capture, but recapture rates differed significantly between museums. Samples from the AMNH had lower overall locus capture than the RMNH, perhaps due to differences in specimen storage over time.

Key words: dragonflies, anchored hybrid enrichment, museum, phylogeny, quality

Introduction

The age of next generation sequencing (NGS) for phylogenomics has sparked a renaissance of museum science as, unlike for Sanger (dideoxy or capillary electrophoresis) sequencing methods, such specimens have been shown to amplify using NGS methods (e.g., Yeates et al. (2016)). Sanger amplification and sequencing methods perform better when DNA fragments are not short, but for older samples (>10 yr in age) and those that may have been degraded by preservation method (Hykin et al. 2015) or pest management (Espeland et al. 2010), long fragments are not common. By contrast, NGS methods best amplify short DNA fragments, which are common in older samples; these make insect museum collections potentially useful for such work. Collections based NGS research using insect specimens older than 10 yr are becoming more common (e.g., Coleoptera: Van Dam et al. 2017; Diptera: Buenaventura 2021; Hemiptera: Dietrich et al. 2017; Lepidoptera: Mayer et al. 2021).

However, the age of a sample can influence DNA yield and locus capture, as may preservation method and storage, as has been shown for terrestrial insects (e.g., Blaimer et al. 2016). It is also

unclear how successful sequencing would be for older specimens of aquatic insects, which are stored either in ethanol (seen often in collections of Ephemeroptera, Plecoptera and Hemiptera adults, Odonata, Ephemeroptera, and Plecoptera nymphs), dried after acetone submersion (most common for odonate adults), or pinned (seen commonly for Hemiptera adults but often also for older collections of Ephemeroptera adults, and Plecoptera adults). For non-holometabolous aquatic insects only one targeted enrichment study has been done (for Odonata by Bybee et al. (2021), but they used recently collected samples).

To address these issues, we sampled dragonflies and damselflies collected between 1909 and 2001 and sequenced them for a modified version of Bybee et al. (2021) AHE Odonata probe set being used by authors. We then evaluated whether DNA yield, specimen age, specimen suborder (a proxy for size), and/or museum collection were good predictors of locus capture success and considered the usefulness of captured loci by attempting to reconstruct an established phylogenetic tree of Odonata. Existing work on Odonate phylogenetics, the large number of global collections of the groups, and the various preservation methods that are commonly employed to preserve Odonate specimens make the group an optimal focal species to consider the usefulness of museum specimens.

Dragonflies and damselflies (Insecta:Odonata) comprise over ~6,400 species categorized in 3 extant suborders: Anisoptera (dragonflies), Zygoptera (damselflies), and Anisozygoptera. Although Odonata systematics traditionally focused heavily on morphological characters for systematics, or relied on Sanger sequencing, recent genomic datasets have been critical for improved reconstruction of the evolutionary history of dragonflies and damselflies (Bybee et al. 2021, Kohli et al. 2021, Suvorov et al. 2022). With the development of a 478 locus anchored hybrid enrichment probe set for Odonata (Bybee et al. 2021), vast amounts of data can now quickly be gathered, allowing tests of hypotheses about the tempo and mode of dragonfly evolution, with far more dense taxon sampling. Sampling densely across the Odonata tree of life will require the use of museum specimens as some species are rare in nature or considered endangered.

Córdoba-Aguilar et al. (2023) reviewed odonate collections and databases globally, which vary in the number of specimens. Storage methods of odonates vary when one looks across time and across museums. Traditionally, odonates were pinned in drawers, or stored in paper envelopes after air drying upon collection. Since the 1980s nearly all specimens are preserved with acetone, which fixes the color in the insect cuticle, and stored in envelopes. Acetone has been shown to preserve insect DNA and as such, acetoned specimens are routinely sequenced (e.g., samples sequenced by (Ware et al. (2007), Pilgrim and Von Dohlen (2008), Letsch et al. 2016), and Bybee et al. (2008) were from odonates treated with acetone). However, the ability to sequence specimens that were not preserved using acetone has been less often explored. Sampling historically collected museum samples of dragonflies and damselflies would allow an expansion of phylogenetic taxon sampling and allow the comparison of populations across space and time.

Materials and Methods

Taxon Sampling

Damselflies and dragonflies from the RMNH and AMNH were selected with an emphasis on having a breadth of sizes, families, and ages. We initially selected samples that ranged in age from 2001 (~20 yr old) to 1909 (~112 yr old). Five specimens appeared to be quite old but had no collection date on the label: *Neurocordulia obsoleta*, *Polythore gigantea*, *Chalcopteryx scintilans*, *Hadrothemis defecta*, *Sapho orichlaceadate*; these were included in the phylogenetic analyses but omitted from our bivariate analyses. Briefly, we chose 94 specimens in total; of these, 64 were Anisoptera and 30 were Zygoptera, from 48 AMNH and 46 RMNH (See Supplementary Table 1). All samples had been dried and were stored in glassine envelopes; it was unclear from the label data the method of preservation but based on their coloration it seemed that most samples had not been treated with acetone for preservation.

DNA Extraction and Sequencing

A single leg was removed from each museum specimen using sterilized forceps and DNA was extracted using Qiagen 2011 Micro-prep kit protocols (Hilden Germany). DNA yield was quantified using a Qubit 4 Fluorometer. DNA extractions were sent to RAPID Genomics (Gainesville Florida) for library preparation and sequencing using Anchored Hybrid Enrichment probes detailed in Bybee et al. (2021). Florida State Center for Anchored Phylogenomics generated the AHE data for Bybee et al. (2021); here we used the sample methods as in Bybee et al. (2021). Total probes consisted of 1,306 loci, capturing 405 AHE loci, and 209 functional loci (Bybee et al. 2021). The average loci size 207.1 bp, with an N50 of 208 bp.

AHE Assembly and Analysis

We trimmed adapters from raw reads for each sample with fastp (Tang and Wong 2001) and checked quality using multiQC. Following trimming, we followed the general methods outlined in Breinholt et al. (2018) to assemble and assign orthology to each targeted capture locus. In brief, we assembled each locus individually using iterative baited assembly with SPAdes (Prjibelski et al. 2020). Following assembly, we screened each locus for orthology by first ensuring that the locus did not have BLAST hits to multiple places in the genome and, secondly, by ensuring best reciprocal hits between the reference and the query sequence.

Evaluation of Factors Impacting Capture Success

To analyze potential drivers of loci capture, we evaluated relationships among various sample-related factors (sample age, sample size (for which we used suborder as a proxy by comparing Zygoptera, which tend to be smaller, and Anisoptera, which tend to be larger), and museum source (due to differences in preservation methods)), the amount of genetic material recovered (DNA qubit quantification), and the number of loci captured. We focused our statistical analysis on the subset of data (n = 57) for which the collection periods overlapped for both museums (1923–1959).

We first used bivariate analysis to examine the direct relationships among sample-related factors, DNA concentration, and the number of loci captured. Model assumptions were checked graphically, and relationships were considered significant at the $P < 0.05$ level (Zuur et al. 2009). Next, we considered the combined effects and interactions among all sample-related factors on DNA concentration and the number of loci captured using multiple regression. Models were evaluated using stepwise backwards selection, with terms dropped until all remaining terms were significant at the $P < 0.05$ level (Zuur et al. 2009). Finally, we used path analysis to determine if any sample-related factors had direct impacts on the number of loci captured or if the effects were fully mediated through impacts on DNA concentration. Partially mediated and fully mediated models containing all possible links between

sample-related factors, DNA concentration, and number of loci captured were developed and compared to models that contained only museum and sample age or size as explanatory factors. Models were compared using AIC values since several models under consideration were saturated. Nested models were also compared using likelihood-ratio tests.

We used the R package *ggplot2* (Wickham et al. 2016) to plot the relationships among sample-related factors, DNA concentration, and locus capture. We used the lavaan (Rosseel 2012) and semPlot (Epskamp et al. 2017) packages to conduct path analysis on samples from 1923 to 1959.

Phylogenetic Analysis

We evaluated the ability of captured loci to reconstruct established phylogenetic trees for Odonata. We generated a multiple sequence alignment using the de novo sequences and the sequences from Bybee et al. (2021) by first aligning the probe regions for each locus using the MAFFT-linsi algorithm in MAFFT v.7.475 (Katoh and Standley 2013). We then concatenated the alignment using FASconCAT v1.11 (Kück and Meusemann 2010) and generated an optimal partitioning scheme using relaxed clustering with the model fixed to GTR+G for each subset in IQtree v.2.1.3 (Minh et al. 2020). We selected a model for each subset in the partitioning scheme using ModelFinder and estimated a maximum likelihood tree with 1,000 ultrafast bootstrap replicates in IQtree v.2.1.3 (Kalyaanamoorthy et al. 2017). Raw data (fastq files), data matrices, partition, and treefiles were uploaded to the dryad digital repository (doi:10.5061/dryad.kprr4xh8z).

Contamination Check

To further check for potential bacterial or fungal contamination, we blasted each locus for each of the taxa which appeared in a clade of uncertainty. Briefly, we used Geneious Prime 2019.2.3 (https://www.geneious.com) with the program BlastN for each locus for each taxon and recorded the top Blast hits (see Supplementary Table S1). Because using BLAST to align sequences to the NCBI database for identification has misassigned well supported scaffolds in Odonata genomes (Tolman et al. 2023), we also utilized taxon-annotated GC-coverage plots to further check for contamination in the screened loci using BlobTools v1.1.1 (Laetsch and Blaxter 2017). We mapped all paired end reads reads against the final screened loci from each problem taxa sing bwa (Li and Durbin 2009), sorted the bam file with samtools v1.13 (Danecek et al. 2021) using the command samtools sort, and made a taxonomic assignment for each loci with megablast using the parameters: task megablast and -e-value 1e-25. We calculated coverage using the blobtools function map2cov, created the blobtools database using the command blobdb, and generated the blobplot with the command blobtools plot.

Results

Capture Results

We captured between 4 and 1,049 loci across taxa in the dataset of de novo sequences. The concatenated alignment of probe regions with these data and Bybee et al. (2021) results in an alignment length of 575,468 bps with a total of 177,297 parsimony informative characters. Targeted enrichment resulted in a minimum of 4 loci for *Sinolestes edita* and a maximum of 1,049 loci for *Agyrthacantha dirupta* (see Supplementary Table 1).

In total, of our de novo sequences, we had 11 Aeshnidae samples (61–107 yr old), 7 Gomphidae (58–110 yr old), 4 Chlorogomphidae (27–88 yr old), 1 Cordulegastridae (110 yr old), 6 Synthemistidae (35–90 yr old), 6 Corduliidae (26–85 yr old), 1 Macromiidae (90 yr old), 27 Libellulidae (57–112 yr old), 2 Lestidae (73–88 yr old), 1 Argiolestidae (24 yr old), 2 Perilestidae (92–101 yr old), 7 Platycnemidae (31–98 yr old), 1 Megapodagrionidae (110 yr old), 2 Calopterygidae (unknown age – 104 yr old), 2 Polythoridae of unknown age, 4 Chlorocyphidae (32–91 yr old), 6 Euphaeidae (20–84 yr old), and 2 Synlestidae (76–83 yr old). Of these, most families had on average ~200–500 loci captured, regardless of age (see Supplementary Table 1).

Evaluation of Factors Impacting Capture Success

Response variables (DNA concentration and number of loci recovered) were log-transformed to ensure models met assumptions. To account for zero measurements, the smallest respective value measured for each variable was added to each measurement prior to transformation. Analysis of bivariate relationships indicated no statistical impacts of sample age ($F_{1,55}$ = 0.069, P = 0.93) or size ($F_{1,55}$ = 0.393, P = 0.533) on Qubit quantification (Fig. 1A and B). However, sample origin did impact Qubit quantification ($F_{1,55}$ = 40.485, P < 0.01); more DNA was recovered in samples from the RMNH (Fig. 1A). Similarly, age ($F_{1,55}$ = 1.67, P = 0.20) and size ($F_{1,55}$ = 1.008, P = 0.319) were not related to the number of loci recovered (Fig. 1C and D), but significantly more loci were recovered in samples from the RMNH ($F_{1,55}$ = 25.592, P < 0.01) (Fig. 1C). There was also a significant positive relationship between qubit quantification and loci recovery ($F_{1,55}$ = 20.915, P < 0.01) (Fig. 1E and F), with concentration explaining approximately 19% in the number of loci recovered (adjusted R^2 value from model on untransformed variables; from model on transformed variables, adjusted R^2 = 26.2%). Multivariate analysis indicated interactions among sample-related factors did not impact the number of loci recovered, with the final selected model only retaining museum as an exploratory factor. However, regression revealed both museum and size impacted the amount of DNA recovered.

Path analysis indicated that the model that included direct links between museum and both DNA concentration and number of loci captured, in addition to direct links between DNA concentration and number of loci captured, led to the lowest observed AIC score by over 2 points, suggesting the model had substantial support compared to other models (Burnham and Anderson 2002). Similarly, likelihood-ratio tests indicated the model containing direct impacts of museum on loci capture rates in addition to effects mediated though DNA concentration was a better fit for the data (χ^2_1=26.81, P < 0.01) compared to model without the direct effects (fully mediated model).

Resolution of Museum Samples in Phylogenetic Reconstruction

Using the AHE pipeline and sequences from these samples coupled with already existing sequences from Bybee et al. (2021), we were able to reconstruct the Odonata phylogeny using likelihood inference. The tree generated (Figure 2) largely supports already existing literature, most recently, Bybee et al. (2021). Among the anisopterans, Libellulidae is recovered as sister to Corduliidae, and together with Synthemistidae and Macromiidae, this grouping forms a monophyletic group (the superfamily Libelluloidea). We recover Cordulegastroidea, a grouping composed of Chlorogromphidae and Cordulegastridae, which as a superfamily is sister to Libelluloidea, recovering a monophyletic Cavilabiata. In the suborder Zygoptera, we see a bifurcation with Coenagrionidae and Platycnemididae as sister groups. We recover Coenagrionoidea, including families Philosinidae, Rhipidolestidae,

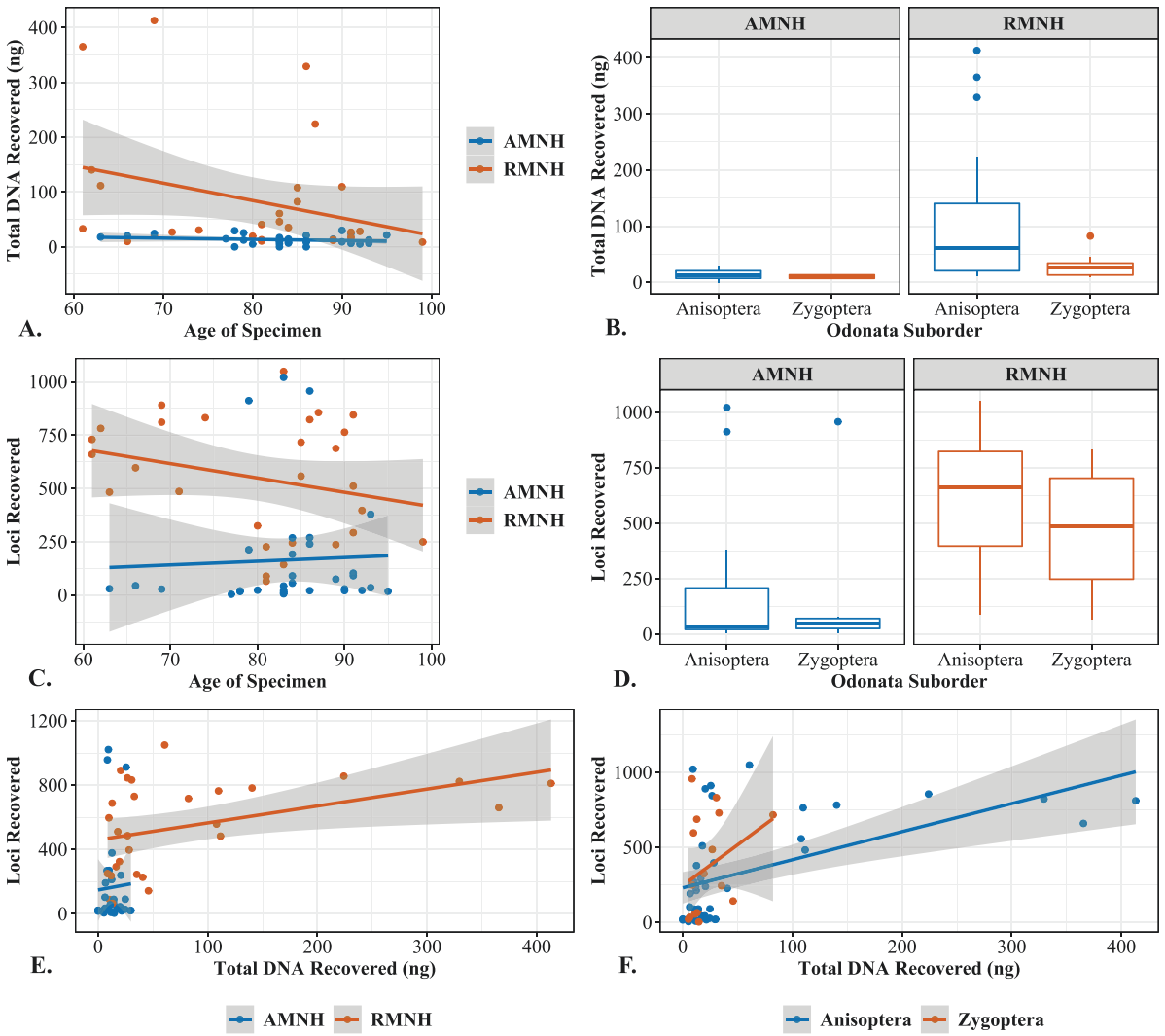

Fig. 1. Qubit DNA recovered, and age of specimen compared between museums (AMNH and RMNH) (A), and Odonata suborder (Anisozygopter and Zygoptera) (B). Number of loci recovered from AHE sequencing, and age of specimen compared between museums (C), and Odonata suborder (D). Number of loci recovered from AHE sequencing, and age of specimen compared between museums (E) and Odonata suborder (F). Shaded lines are linear regression trend lines, and the grey shading is the 95% confidence interval.

Devadattidae, Pentaphlebiidae, Polythoridae, Lestoideidae, Euphaeidae, Argiolestidae, Chlorocyphidae, and Calopterygidae.

We particularly draw attention to 25 samples which form a clade that is apparently united by a lack of data; these include all the taxa for which there were fewer than 70 loci sequenced, except for *Anatya* (6 loci) and *Sinolestes edita* (4 loci). These additional taxa are not received in this motley crew of a clade but are similarly presumably misplaced in the topology. Additionally, *Watanabeopetalia* from Bybee et al. (2021) was recovered in Zygoptera, likely due to low locus overlap. A heat map over locus overlap (see Supplementary Figure S1) suggests that all of the 25 taxa in this problematic clade shared only 0–10% overlap with the loci for which it was sequenced, and the loci recovered for other taxa; in addition, *Anatya* and *Sinolestes* similarly had only 0–10% overlap of loci. The remaining taxon which was apparently misplaced in the topology is *Idionyx carinata*, which has 50% overlap among the 912 loci recovered, suggesting this sequence is from a contaminated sample.

We reviewed the loci recovered for these ambiguously recovered samples. There are 58 loci for which no insect hits were recovered across these taxa; however, blasting to NCBI may not reveal much new information as the database does not have many Odonata genomes within it. In general, the most common Blast results were for Odonata (Ischnura, Sympetrum), and Formicidae. Given the incredibly small number of other AHE data for these probe regions on NCBI, and given the low number of Odonata genomes, we considered a Blast result of any species in Insecta to be unlikely contamination; any BLAST result not to Insecta was considered a possible contaminant. Indeed, the misplaced taxa not in the ambiguous clade (i.e., *Watanabeopetalia* and *Idionyx carinata*) had 64 and 56 non-Insecta BLAST results, respectively, which potentially affected their phylogenetic placement. However, the 25 taxa in the ambiguous clade taxa had only 100 non-insecta BLAST results across all loci. Genome level datasets (including Odonata) have been misclassified outside of arthropoda using BLAST (Tolman et al. 2023), so BLAST results for Odonata should be treated with caution. The blobplot did confirm that a number of the loci were not assigned as arthropoda, but there were few outliers when considering coverage and GC content (see Supplementary Figure S2). It is notable that the misplaced taxa had a lower average loci length, and loci N50 than the dataset as a whole (see Supplementary Table S2).

Fig. 2. Maximum Likelihood best tree with bootstrap values listed at each node; <95% indicated with a yellow circle. Triangles represent problematic taxa including the 25 taxa which possessed fewer than 70 loci which formed a clade, the genera *Anatya* (6 loci) and *Sinolestes edita* (4 loci), as well as our 2 misplaced taxa *Watanabeopetalia* and *Idionxy carinata*. Shaded text indicates all taxa from museums included as part of the test of our museumomics evaluation.

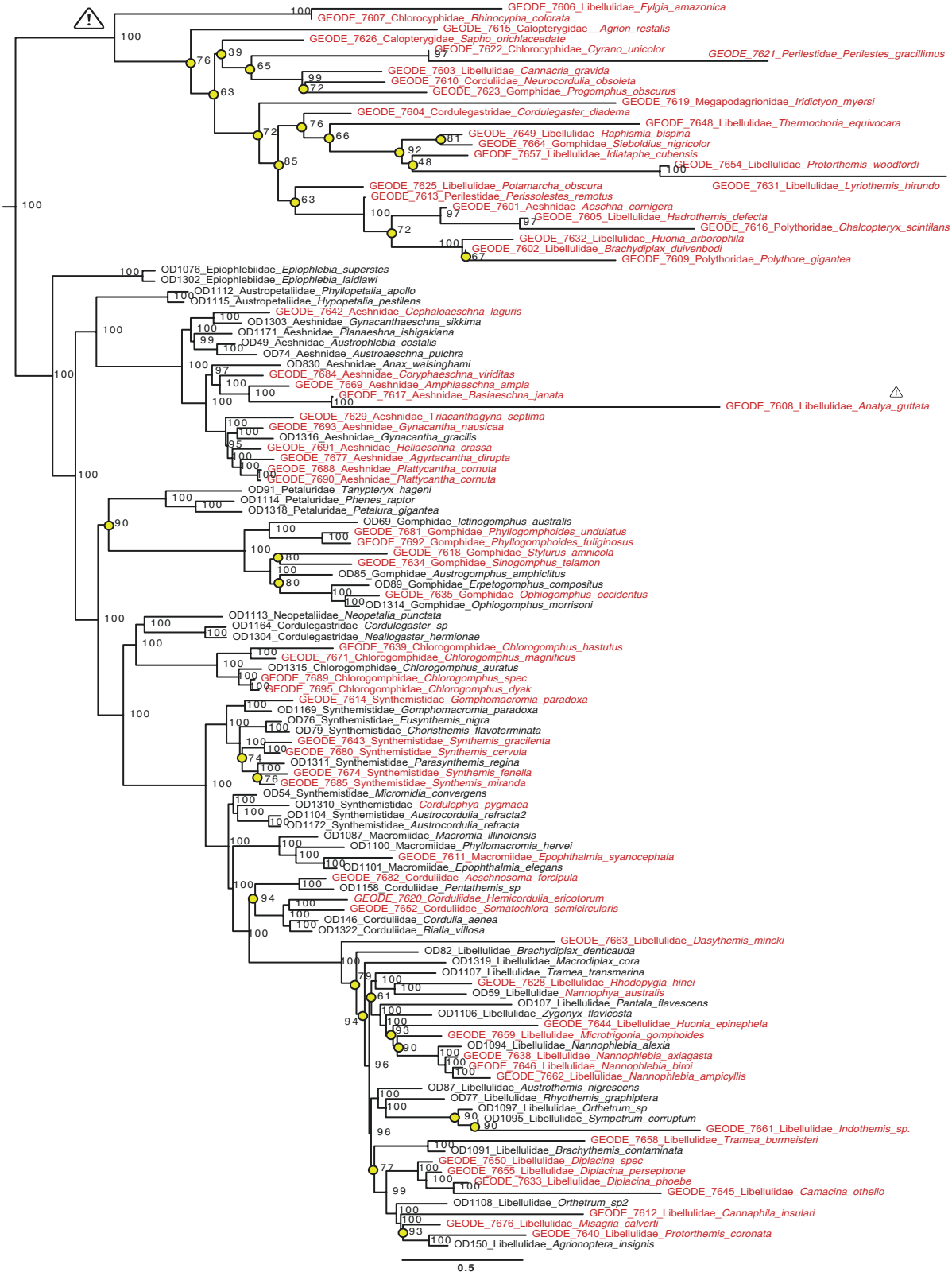

Fig. 2. Continued

Discussion

There are well over 1.5 million specimens of dragonflies and damselflies held in museums across the United States, and when all natural history collections worldwide are considered the number of preserved Odonata is likely numbering in several million. Odonata collectors continue to add to collections in museums, but many species remain rare and undersampled; the rate of collecting odonates has precipitously decreased, heightening the urgency of extracting

as much information from housed specimens as possible (Córdoba-Aguilar et al. 2023). These samples are invaluable records of geographic ranges, phenotypic variation, and populations; here, we show that they are very useful molecular data collection as well.

However, most odonate specimens were not collected with the intention (or knowledge) of future genomic studies, resulting in varying methods of preservation. Most labels for older odonata specimens lack knowledge of post-mortem treatment, which could drastically alter DNA quality among or within taxa, and among specimens of similar or differing ages. Common treatments for odonate specimens include freezing, ethyl acetate, and arsenic for killing, and boiling water and phenol as relaxing agents before pinning. Common storage agents for insects include submersion in 90–95% ethanol or acetone, however dragonflies are stored within envelopes or pinned using such agents as facilitators for quick drying; the comparison of quick drying versus air-drying for extraction of viable genetic material is an area of little study (Nakahama et al. 2019). Furthermore, several factors including time, dehydration, environmental exposure, and the presence of bacterial or fungal contamination, markedly affect the quality of DNA within museums, which is observed both within vertebrate and invertebrate taxa (Evans 2007, Zimmermann et al. 2008, Francis et al. 2010, Cárdenas and Moore 2019). Finally, at certain museums, older odonate specimens have been exposed to naphthalene or methyl bromide, common pesticide agents; the effects of which on genetic material currently remain unknown. All such factors from the moment of capture to the DNA extraction provide a spectrum of efficacy in acquiring robust genetic material for diversity studies.

The results of our study suggest that high numbers of loci may commonly be recovered from specimens, but that differences in storage procedures among museums may impact recovery rates. In general, we noted a positive correlation between the amount of recovered DNA and the number of loci recovered per sample. Increases in DNA yield have the potential to increase the chances of recovering loci due to the higher abundance of presumed viable DNA for sequencing. However, DNA concentration only explained 20% of the noted variation in loci recovered, suggesting Qubit values are not necessarily a strong indicator of the success of AHE. Specimens with low qubit values still yielded a high number of informative loci.

Path analysis indicated specimen source had both direct effects on loci capture and indirect effects mediated through DNA yield. We observed stark differences between Qubit readings and recovered loci between our museums. Specimens housed within the RMNH possessed higher average readings than the AMNH, both in Qubit (RMNH: 124ng, AMNH: 11ng), and recovered loci (RMNH: 662bp, AMNH: 109bp), suggesting external factors pertaining to storage which differ between the RMNH and the AMNH. A litany of external factors pertaining to storage between both museums could produce differences in DNA quality and quantity including stability in climate conditions, or presence of pests and pathogens. Analysis of samples from more museums will be required to better understand these impacts.

We note that samples from AMNH seemed to be more susceptible to low recovery and low locus overlap. This included a number of samples (25) that formed their own clade, separated from both the main Zygoptera and the main Anisoptera clade. This grouping is most likely due to a lack of overlap among these taxa (all have only 0–10% overlap in loci), and generally low locus recovery (all had less than 70 loci recovered). In general, a lack of data might be the reason for the lack of their resolution considering our mostly Insecta Blast results. Our results indicate that as a rule of thumb, taxa with 10% or less loci overlap deserve extra scrutiny when evaluating their placement in the topology. *Idionyx carinatus* was recovered in a problematic position in the tree despite having amplified over 900 loci with 50% overlap. We posit that this could be due to the presence of non-Insecta contaminants (e.g., bacteria, fungi, human, dog, etc.) for 64 of 912 recovered loci. A greater abundance of microbial activity could explain why less DNA and fewer loci were recovered from AMNH samples, in general. Museums will have variable storage quality, and humidity levels may influence microbial contamination. Further research must be done to determine the storage factors that influence the success of AHE sequencing, so reasonable expectations can be made for specimens from museums around the world.

Surprisingly, we found no impact of specimen age or size on the number of loci recovered or DNA yield. Previous studies of vertebrate and invertebrate taxa have found that older specimens possess fewer loci recovered (McCormack et al. 2016, Brewer et al. 2019, McGaughran 2020, Mayer et al. 2021). However, we produced the same result when we included taxa outside the overlapping age range of both museums (1909–2001). Our success in recovering loci from old samples is likely due to the short length of the target loci (average length < 210 bp). This success is unsurprising, as average read lengths near 70 have been recovered from the genomic DNA of large vertebrates that were nearly 40,000 yr old (Palkopoulou et al. 2015). Overall, this suggests even older, smaller samples may prove useful in genetic studies.

Seeing the relative success of this method in DNA and loci recovery from museum specimens, another angle of interest worth highlighting is the aspect of accessibility and capacity building in regions with a dearth of resources. The 2 museums used in this study are in the global north—Europe and North America, which possess a wealth of resources to employ the best practices in the preservation of these specimens. Since the advent of more sustainable and capital-intensive methods, biological repositories in the global south have found themselves having some catching up to do. Further, many museums lack climate-controlled compactor style storage, using naphthalene to control pests, and storing insects in suboptimal temperatures (Gyanpriya Maharaj & Kehinde Kemabonta, personal communication), which prevents these specimens from being as useful for molecular work, only exacerbating current inequities from colonial legacies.

Dragonflies and damselflies are remarkable insects and their position in the insect tree of life, as sister to the Neoptera, make them vital to our understanding of the evolution of Pterygota. To better study their evolution, museum collections are an invaluable resource that should be sampled to expand taxon sampling across space and time.

Supplementary Material

Supplementary material is available at *Insect Systematics and Diversity* online.

Specimen Collection Statement

The authors attest that all legal and regulatory requirements, including export and import collection permits, have been followed for the collection of specimens from source populations at any international, national, regional, or other geographic level for all relevant field specimens collected as part of this study.

Acknowledgments

We thank Ruth Salas for her tireless efforts to curate the Odonata collection at the AMNH. We thank Max Caspers for his work to curate the Odonata collection at the RMNH.

Funding

This work was supported by the National Science Foundation (NSF) Awards (Abbott: 2002489; Ware: 2002473; Bybee: 2002432; Guralnick: 2002457): Genealogy of Odonata (GEODE): Dispersal and color as drivers of 300 million years of global dragonfly evolution.

Author Contributions

Aaron M Goodman (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Validation-Equal, Visualization-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Ethan Tolman (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Validation-Equal, Visualization-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Rhema Uche Dike (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Validation-Equal, Visualization-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), John Abbott (Conceptualization-Equal, Funding acquisition-Equal, Project administration-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Jesse Breinholt (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Methodology-Equal, Resources-Equal, Software-Equal, Validation-Equal), Seth Bybee (Conceptualization-Equal, Project administration-Equal, Resources-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Paul Frandsen (Conceptualization-Equal, Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Project administration-Equal, Resources-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Stephen Gosnell (Formal analysis-Equal, Investigation-Equal, Visualization-Equal, Writing – review & editing-Equal), Robert Guralnick (Conceptualization-Equal, Funding acquisition-Equal, Project administration-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Vincent Kalkman (Conceptualization-Equal, Project administration-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Manpreet Kohli (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Validation-Equal, Visualization-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Judicael Fomekong Lontchi (Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Pungki Lupiyaningdyah (Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal), Lacie Newton (Data curation-Equal, Formal analysis-Equal, Investigation-Equal, Validation-Equal, Visualization-Equal, Writing – review & editing-Equal), Jessica Ware (Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal, Project administration-Equal, Resources-Equal, Validation-Equal, Visualization-Equal, Writing – original draft-Equal, Writing – review & editing-Equal)

Data Availability

The data underlying this article is uploaded to the Dryad Digital Repository (doi:10.5061/dryad.kprr4xh8z).

References

Blaimer BB, Lloyd MW, Guillory WX, Brady SG. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One. **2016**:11(8):e0161531. https://doi.org/10.1371/journal.pone.0161531

Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. Syst Biol. **2018**:67(1):78–93. https://doi.org/10.1093/sysbio/syx048

Brewer GE, Clarkson JJ, Maurin O, Zuntini AR, Barber V, Bellot S, Biggs N, Cowan RS, Davies NM, Dodsworth S. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. Front Plant Sci. **2019**:10:1102.

Buenaventura E. Museomics and phylogenomics with protein-encoding ultraconserved elements illuminate the evolution of life history and phallic morphology of flesh flies (Diptera: Sarcophagidae). BMC Ecol Evolut. **2021**:21(1):1–28.

Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. New York, NY: Spnnger-Veflag; **2002**.

Bybee SM, Kalkman VJ, Erickson RJ, Frandsen PB, Breinholt JW, Suvorov A, Dijkstra K-DB, Cordero-Rivera A, Skevington JH, Abbott JC, *et al*. Phylogeny and classification of Odonata using targeted genomics. Mol Phylogenet Evol. **2021**:160:107115. https://doi.org/10.1016/j.ympev.2021.107115

Bybee SM, Ogden TH, Branham MA, Whiting MF. Molecules, morphology and fossils: a comprehensive approach to odonate phylogeny and the evolution of the odonate wing. Cladistics. **2008**:24(4):477–514. https://doi.org/10.1111/j.1096-0031.2007.00191.x

Cárdenas P, Moore JA. First records of Geodia demosponges from the New England seamounts, an opportunity to test the use of DNA mini-barcodes on museum specimens. Mar Biodivers. **2019**:49:163–174.

Córdoba-Aguilar A, Beatty C, Bried J. Dragonflies and damselflies: model organisms for ecological and evolutionary research. Oxford University Press; **2023**.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. Twelve years of SAMtools and BCFtools. GigaScience. **2021**:10(2):giab008. https://doi.org/10.1093/gigascience/giab008

Dietrich CH, Allen JM, Lemmon AR, Lemmon EM, Takiya DM, Evangelista O, Walden KK, Grady PG, Johnson KP. Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). Insect Syst Diversity. **2017**:1:57–72.

Epskamp S, Epskamp MS, MplusAutomation S. Package 'semPlot'. Recuperado de. **2017**. https://cran.r-project.org/web/packages/semPlot/semPlot.pdf.

Espeland M, Irestedt M, Johanson KA, Åkerlund M, Bergh J-E, Källersjö M. Dichlorvos exposure impedes extraction and amplification of DNA from insects in museum collections. Front Zool. **2010**:7:1–7.

Evans TC. DNA damage. NEB expressions. **2007**:2:1–3. https://www.neb.com/-/media/nebus/files/expressions/expressions_2007_spring.pdf.

Francis CM, Borisenko AV, Ivanova NV, Eger JL, Lim BK, Guillén-Servent A, Kruskop SV, Mackie I, Hebert PD. The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. PLoS One. **2010**:5(9):e12575. https://doi.org/10.1371/journal.pone.0012575

Hykin SM, Bi K, McGuire JA. Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. PLoS One. **2015**:10(10):e0141579. https://doi.org/10.1371/journal.pone.0141579

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. **2017**:14:587–589.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. **2013**:30(4):772–780. https://doi.org/10.1093/molbev/mst010

Kohli M, Letsch H, Greve C, Béthoux O, Deregnaucourt I, Liu S, Zhou X, Donath A, Mayer C, Podsiadlowski L, *et al*. Evolutionary history and divergence times of Odonata (dragonflies and damselflies) revealed through transcriptomics. Iscience. **2021**:24:103324.

Kück P, Meusemann K. FASconCAT: convenient handling of data matrices. Mol Phylogenet Evol. **2010**:56(3):1115–1118. https://doi.org/10.1016/j.ympev.2010.04.024

Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Research. **2017**:6:1287. https://doi.org/10.12688/f1000research.12232.1

Letsch H, Gottsberger B, Ware JL. Not going with the flow: a comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides evidence for the role of lentic habitats on diversification. Mol Ecol. **2016**:25(6):1340–1353. https://doi.org/10.1111/mec.13562

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. **2009**:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Mayer C, Dietz L, Call E, Kukowka S, Martin S, Espeland M. Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old

museum specimens. Syst Entomol. **2021**:46(3):649–671. https://doi.org/10.1111/syen.12481

McCormack JE, Tsai WL, Faircloth BC. Sequence capture of ultraconserved elements from bird museum specimens. Mol Ecol Resour. **2016**:16(5):1189–1203. https://doi.org/10.1111/1755-0998.12466

McGaughran A. Effects of sample age on data quality from targeted sequencing of museum specimens: what are we capturing in time? BMC Genomics. **2020**:21(1):1–10.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. **2020**:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Nakahama N, Isagi Y, Ito M. Methods for retaining well-preserved DNA with dried specimens of insects. Eur J Entomol. **2019**:116:486–491. https://doi.org/10.14411/eje.2019.050

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, *et al*. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr Biol. **2015**:25(10):1395–1400. https://doi.org/10.1016/j.cub.2015.04.007

Pilgrim EM, Von Dohlen CD. Phylogeny of the Sympetrinae (Odonata: Libellulidae): further evidence of the homoplasious nature of wing venation. Syst Entomol. **2008**:33:159–174.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc. Bioinform. **2020**:70(1):e102. https://doi.org/10.1002/cpbi.102

Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw. **2012**:48(2):1–36.

Suvorov A, Scornavacca C, Fujimoto MS, Bodily P, Clement M, Crandall KA, Whiting MF, Schrider DR, Bybee SM. Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies. Syst Biol. **2022**:71(3):526–546. https://doi.org/10.1093/sysbio/syab063

Tang X, Wong D. Published. FAST-SP: a fast algorithm for block placement based on sequence pair. In: Proceedings of the 2001 Asia and South Pacific Design Automation Conference. New York, NY, United States: Association for Computing Machinery; **2002**. p. 521–526.

Tolman ER, Beatty CD, Bush J, Kohli M, Moreno CM, Ware JL, Weber KS, Khan R, Maheshwari C, Weisz D, *et al*. A Chromosome-length assembly of the Black Petaltail (Tanypteryx hageni) Dragonfly. Genome Biol Evol. **2023**:15(3):evad024. https://doi.org/10.1093/gbe/evad024

Van Dam MH, Lam AW, Sagata K, Gewa B, Laufa R, Balke M, Faircloth BC, Riedel A. Ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurf-weevils. PLoS One. **2017**:12(11):e0188044. https://doi.org/10.1371/journal.pone.0188044

Ware J, May M, Kjer K. Phylogeny of the higher Libelluloidea (Anisoptera: Odonata): an exploration of the most speciose superfamily of dragonflies. Mol Phylogenet Evol. **2007**:45(1):289–310. https://doi.org/10.1016/j.ympev.2007.05.027

Wickham H, Chang W, Wickham MH. Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version 2; **2016**. p. 1–189.

Yeates DK, Zwick A, Mikheyev AS. Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. Curr Opin Insect Sci. **2016**:18:83–88. https://doi.org/10.1016/j.cois.2016.09.009

Zimmermann J, Hajibabaei M, Blackburn DC, Hanken J, Cantin E, Posfai J, Evans TC. DNA damage in preserved specimens and tissue samples: a molecular assessment. Front Zool. **2008**:5(1):1–13.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. vol. 574. New York, NY: Springer; **2009**.