Received: 14 May 2019 | Revised: 22 July 2019 | Accepted: 31 July 2019

DOI: 10.1111/1755-0998.13072

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault

Shahan Derkarabetian | Ligia R. Benavides | Gonzalo Giribet

Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

Correspondence
Shahan Derkarabetian, Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA.
Email: sderkarabetian@gmail.com

Funding information
Harvard Dean's Fund; Division of Environmental Biology, Grant/Award Number: DEB-1601208 and DEB-1754278

Abstract

Natural history collections play a crucial role in biodiversity research, and museum specimens are increasingly being incorporated into modern genetics-based studies. Sequence capture methods have proven incredibly useful for phylogenomics, providing the additional ability to sequence historical museum specimens with highly degraded DNA, which until recently have been deemed less valuable for genetic work. The successful sequencing of ultraconserved elements (UCEs) from historical museum specimens has been demonstrated on multiple tissue types including dried bird skins, formalin-fixed squamates and pinned insects. However, no study has thoroughly demonstrated this approach for historical ethanol-preserved museum specimens. Alongside sequencing of "fresh" specimens preserved in >95% ethanol and stored at −80°C, we used extraction techniques specifically designed for degraded DNA coupled with sequence capture protocols to sequence UCEs from historical museum specimens preserved in 70%–80% ethanol and stored at room temperature, the standard for such ethanol-preserved museum collections. Across 35 fresh and 15 historical museum samples of the arachnid order Opiliones, an average of 345 UCE loci were included in phylogenomic matrices, with museum samples ranging from six to 495 loci. We successfully demonstrate the inclusion of historical ethanol-preserved museum specimens in modern sequence capture phylogenomic studies, show a high frequency of variant bases at the species and population levels, and from off-target reads successfully recover multiple loci traditionally sequenced in multilocus studies including mitochondrial loci and nuclear rRNA loci. The methods detailed in this study will allow researchers to potentially acquire genetic data from millions of ethanol-preserved museum specimens held in collections worldwide.

KEYWORDS
degraded DNA, ethanol preservation, museum collections, Opiliones, phylogenomics, ultraconserved elements

1 | INTRODUCTION

Natural history museums hold a wealth of biological information in their collections, and museum specimens are increasingly being used in modern research, enhancing their value, visibility and impact (Beaman & Cellinese, 2012; Nelson & Ellis, 2019; Short, Dikow, & Moreau, 2018; Soltis & Soltis, 2016; Webster, 2017). This research has been facilitated in many ways including the use of cutting-edge technology for morphological work (e.g., Fernández, Kvist, Lenihan, Giribet, & Ziegler, 2014; van de Kamp et al., 2018), the development of high-throughput imaging techniques and digitization of specimens (e.g., Blagoderov, Kitching, Livermore, Simonsen, & Smith, 2012), or

the creation of searchable databases of museum holdings, occurrence records and natural history information (e.g., iDigBio, www.idigbio.org). Recently, there has been a burst of research using historical museum specimens (i.e., specimens in natural history collections that were not collected with an intention to conduct molecular work, such as pinned insects, dry specimens or specimens preserved in low-grade ethanol) in phylogenomics, further increasing the importance of natural history collections in modern studies of biodiversity and evolution. The increase in the use of historical specimens in studies relying on genetic data is due largely in part to the development of laboratory methods, such as DNA extraction protocols specifically designed for degraded DNA samples (Tin, Economo, & Mikheyev, 2014) and reduced-genome data collection methods like sequence capture approaches, combined with many modern DNA sequencing methods (Burrell, Disotell, & Bergey, 2015; Faircloth et al., 2012; Suchan et al., 2016). In this regard, museum specimens can be considered genomic resources (McCormack, Rodríguez-Gómez, Tsai, & Faircloth, 2017) and can be used for diverse types of studies. Examples of such studies are fine-scale population genomics and genotyping (Bi et al., 2013; Lim & Braun, 2016), genome sequencing (Staats et al., 2013), metagenomics (Der Sarkissian et al., 2017), epigenomics (Rubi, Knowles, & Dantzer, 2019), barcoding (Miller, Beentjes, van Helsdingen, & IJland, 2013); Prosser, deWaard, Miller, & Hebert, 2016), species delimitation (Hedin, Derkarabetian, Blair, & Paquin, 2018; Kehlmaier et al., 2019), and phylogenomics (Blaimer, Lloyd, Guillory, & Brady, 2016; Hedin, Derkarabetian, Ramírez, Vink, & Bond, 2018; Ruane & Austin, 2017; Sproul & Maddison, 2017; Starrett et al., 2017; Wood, González, Lloyd, Coddington, & Scharff, 2018), including phylogenomic studies that incorporate genetic data from rare, endangered and/or extinct taxa held in historical collections (Hedin, Derkarabetian, Blair, et al., 2018; Oliveros et al., 2019; Tsai et al., 2019). This was not the case just a few years ago, when historical museum samples were not routinely used for molecular work.

For phylogenomics, sequence capture approaches have been highly successful, providing the ability to include data from museum specimens where DNA is typically highly degraded. The utility of sequence capture of ultraconserved elements (UCEs) with historical museum specimens has been thoroughly demonstrated for multiple preserved tissue types including bird skins (McCormack, Tsai, & Faircloth, 2016), formalin-fixed squamates (Hykin, Bi, & McGuire, 2015; Ruane & Austin, 2017), pinned insects (Blaimer, Lloyd, et al., 2016; Faircloth, Branstetter, White, & Brady, 2015), dried mollusc tissue (Abdelkrim et al., 2018), and mammal skin clips (Swanson, Oliveros, & Esselstyn, 2019), including successful sequencing of specimens over 100 years old (Blaimer, Lloyd, et al., 2016; McCormack et al., 2016; Ruane & Austin, 2017). Despite the obvious success of sequence capture methods across preservation types, these approaches have not yet been used to their full potential with standard historical museum specimens preserved in ethanol (typically 70%–80% ethanol, but often of lower grade due to evaporation, and stored at room temperature) spanning comparable time frames as those of the previously mentioned studies, although Ruane & Austin (2017) succesfully sequenced a single ethanol-preserved snake specimen over 100 years old. Ethanol-preserved collections include some vertebrate taxa, the great majority of non-insect arthropods and most soft-bodied invertebrates, including cnidarians, annelids and most non-dried molluscs, among many other groups, representing a large proportion of specimens held in museum collections worldwide. Recent studies utilizing sequence capture of UCEs in arachnids have successfully included some historical ethanol-preserved museum specimens in phylogenomic analyses (Hedin, Derkarabetian, Alfaro, Ramírez, & Bond, 2019; Hedin, Derkarabetian, Blair, et al., 2018; Hedin, Derkarabetian, Ramírez, et al., 2018; Wood et al., 2018) using standard extraction and sequence capture protocols, but success was limited to specimens collected within the last ~30 years, despite some attempts targeting older specimens.

In this study, we used DNA extraction techniques specifically designed for degraded DNA coupled with modified sequence capture methods to sequence UCEs from standard ethanol-preserved museum specimens of the arachnid order Opiliones (commonly called harvestmen or daddy long-legs) collected up to ~150 years ago. We successfully demonstrate the inclusion of these museum specimens in phylogenomic analyses, and explore the effect the percentage of museum specimens in the dataset has on locus inclusion in final matrices. Similarly, our results further suggest the potential for these UCE data in species delimitation analyses and population-level studies incorporating museum specimens. An additional benefit of the UCE sequence capture approach is the sequencing of mitochondrial genes as "bycatch" (do Amaral et al., 2015; Zarza et al., 2018). As such, for fresh and museum specimens we also explore the utility of sequencing multiple bycatch loci that are typical targets of past multilocus phylogenetic studies using traditional Sanger sequencing, such as nuclear ribosomal RNAs and mitochondrial genes.

2 | MATERIALS AND METHODS

2.1 | Taxon sample

For this study, we refer to samples preserved in >95% ethanol and kept at −20 to −80°C as "fresh" samples and those specimens preserved in 70%–80% ethanol and kept at room temperature as "museum" samples, although most modern collections today have both types of specimens. A total of 35 fresh specimens were included, 18 available from previously published UCE studies (Derkarabetian et al., 2018; Starrett et al., 2017), and 17 newly sequenced for this study (Table S1). Newly sequenced fresh samples were taken from the Museum of Comparative Zoology (MCZ), the San Diego State University Terrestrial Arthropod Collection (SDSUTAC), the California Academy of Sciences (CAS), and the Natural History Museum of Denmark–University of Copenhagen (NHMUC). The museum specimens were taken from the MCZ and NHMUC (Table 1; Table S1). Included samples targeted all four suborders of Opiliones with an emphasis on the laniatorean family Triaenonychidae. Museum samples were chosen based on age, spanning more recent collections from 2008 back to some of the oldest nontype Opiliones specimens

TABLE 1 Sequencing and matrix statistics for all museum samples included in this study

Species	Voucher	Year	Age (years)	Initial conc. (ng/µl)	Input DNA (ng)	Reads	Contigs	Mean length (bp)	Raw UCEs	Final matrix	Total length (bp)
Phalangium opilio	MCZ 36220	1967	52	25	500	484,239	156,596	316.5	822	425	74,164
Leiobunum formosum	MCZ 36740	1865	154	0.68	17	15,759	13,959	262.3	93	6	702
Leiobunum vittatum	MCZ 36455	1950	69	2.14	96	620,638	66,318	620.4	441	268	35,313
Leiobunum vittatum	MCZ 36816	1938	81	2.66	120	476,117	10,668	1,072.4	52	14	1,599
Vonones ornata	MCZ 37491	1919	100	1.22	58	228,322	30,168	285.8	60	35	3,688
Vonones ornata	MCZ 37493	1899	120	1.59	76	515,460	55,988	298.7	135	105	12,644
Vonones ornata	MCZ 37499	1976	43	2.82	135	1,951,115	119,220	300.5	245	184	23,452
Vonones sayi	MCZ 37537	1955	64	0.69	33	751,203	117,072	294.1	172	95	11,165
Callihamina sp.	NHMUC18A	2008	11	43.3	500	837,546	391,935	357.7	831	495	93,627
Heteronuncia robusta	MCZ 35919	1914	105	2.73	123	141,542	17,215	350.1	143	92	10,980
new genus B	NHMUC32A	2008	11	37.5	500	883,638	467,808	345.5	945	504	98,440
new genus nr *Triaenobunus*	NHMUC21B	2008	11	16.3	456	404,246	142,389	330	701	461	75,287
Paranuncia ingens	MCZ 48421	1958	61	3.64	175	864,243	54,976	315.3	66	46	5,160
Triaenonychidae sp.	MCZ 52828	1992	27	2.76	77	83,842	28,541	308.2	391	270	36,077
Triaenonychoides breviops	MCZ 31332	1965	54	3.86	185	994,227	94,932	320.2	271	214	27,676

Note: Reads refers to the total reads after quality control. The number of raw UCEs, UCEs in the final matrix, and total length refer to the 65/65 dataset. Full statistics for all samples are given Table S1.

held in the MCZ collected in 1865. The oldest specimens chosen were commonly collected larger bodied taxa represented in the collections by many samples (e.g., *Leiobunum*, *Vonones*), specifically chosen to maximize the amount of DNA available for library preparation. A total of five specimens from the two North American species of the cosmetid genus *Vonones* were included (one fresh and four museum) to assess UCE locus recovery across samples of the same size/tissue input using museum specimens of varying age spanning the years 1976–1899, and to assess the potential for species-level variation and utility of UCEs derived from museum samples.

2.2 | Sequence capture

DNA from fresh specimens was extracted using the Qiagen DNeasy Blood and Tissue kit following the manufacturer's protocol, with a final elution volume of 150 µl. Museum specimens were extracted following the protocol of Tin et al. (2014), an approach using silica-based magnetic beads and specifically designed for nondestructive DNA extraction from degraded specimens. Details of the extraction protocol used, derived from Tin et al. (2014), are provided in the Supporting Information. Preliminary testing of this approach resulted in increased DNA yield relative to standard phenol–chloroform extractions with an average five-fold increase in the amount of recovered DNA across eight replicated samples (Table S1). For museum specimens one set of appendages was used for extractions, except for the smaller-bodied *Callihamina* NHMUC18A, which consisted of a whole-body extraction. Before DNA extractions, tissue was placed in distilled H_2O overnight at 4°C to facilitate the removal of ethanol. Museum extractions were conducted in a laminar flow hood, tissue was destructively extracted by grinding with sterile disposable pestles, and DNA was eluted in a final volume of 50 µl. Although the Tin et al. (2014) method was originally demonstrated with nondestructive sampling, we used destructive sampling to maximize the amount of DNA extracted. All extractions were quantified using a Qubit fluorometer and visualized on an Agilent Tape Station to assess DNA degradation. Up to 500 ng of DNA for all fresh and recent museum specimens (e.g., <20 years) was sonicated in 130 µl with a Covaris S220 Focused-ultrasonicator using the default settings for a target peak of 500 bp. Museum specimens >20–30 years old did not require sonication as DNA was already naturally degraded to the appropriate size for sequence capture library preparation.

Sequence capture protocols for fresh specimens followed previously published methods used in arachnid studies (Derkarabetian et al., 2018; Hedin, Derkarabetian, Blair, et al., 2018; Starrett et al., 2017). The hybridization followed a standard protocol using the Arbor Biosciences myBaits kit version 4 (arborbiosci.com) and the arachnid-specific probe set (Faircloth, 2017). Specific adjustments to the standard sequence capture protocol used for the museum samples are highlighted here. We do note that many of these adjustments have previously been used in various UCE studies cited throughout the text. For museum specimens, 3× bead clean-ups were conducted throughout library preparation using an Ampure substitute (Rohland & Reich, 2012), 5 µm stubs (Glenn et al., 2019) were used in adapter ligations with a ligation time of 60 min, and 18

cycles were used during post-ligation amplification. Hybridization was conducted at 60°C for 24 hr with 18 cycles for the post-hybridization amplification. For the older museum specimens (>30 years) only four to seven individual samples were included in a hybridization pool depending on the amount of extracted DNA. Adjusting the number of samples in a hybridization pool is effectively similar to adjusting the concentration of probes in the hybridization reaction (Quattrini et al., 2018). Samples included in this study were sequenced over multiple plates and sequencing experiments. Sequencing was conducted on an Illumina HiSeq2500, using either 150-bp paired-end reads and sequenced at the Bauer Core Facility at Harvard University or 125-bp paired-end reads and sequenced at the Brigham Young University DNA Sequencing Center.

2.3 | Bioinformatics and phylogenetic analyses

Sequencing reads were processed and final datasets were created using the PHYLUCE version 1.6 package (Faircloth, 2015). An Illumiprocessor wrapper (Faircloth, 2013) was used for quality control and adapter removal. Contigs were assembled using both VELVET (Zerbino & Birney, 2008) and TRINITY (Grabherr et al., 2011) and combined into a single contig file for matching to probes because UCE locus recovery increases when different assembly types are combined (Derkarabetian et al., 2018; Hedin, Derkarabetian, Ramírez, et al., 2018). Coverage statistics were calculated in PHYLUCE using only the TRINITY assemblies. For all datasets, sequences were aligned with MAFFT (Katoh & Standley, 2013) and trimmed with GBLOCKS (Castresana, 2000; Talavera & Castresana, 2007) using the settings -b1 0.5, -b2 0.5, -b3 6, -b4 6. We refer to this dataset with 50 samples (35 fresh and 15 museum) as the "CORE" data set. From this CORE data set, multiple subsets were created to test the effect the percentage of museum samples in the dataset has on locus counts in the final matrix. The percentage of museum samples included in the subsets are 30% ("CORE" dataset), 50%, 75%, 90%, and 100% (i.e., only museum samples). Additionally, we created datasets removing museum samples based on lowest final UCE count in the CORE dataset, equating to 24% (four removed) and 17% matrices (eight removed). For all datasets, we also tested different values when matching contigs to probes to account for potential contamination of nontarget sequences. Two sets of minimum coverage and minimum identity percentage values were used: first, we use 65/65 values as in previous arachnid sequence capture studies (Derkarabetian et al., 2018; Hedin et al., 2019; Hedin, Derkarabetian, Ramírez, et al., 2018; Starrett et al., 2017); and second, 82/80, which were recommended values to reduce the amount of nontarget UCEs recovered when using the arthropod probe sets (Bossert & Danforth, 2018). These values refer to the minimum percent coverage and minimum percent identity required to consider a contig successfully matched to a UCE probe sequence.

Phylogenies were estimated only for the CORE 65/65 and 82/80 matrices including loci with at least 50% taxon coverage. For both datasets, all individual loci were manually inspected in GENEIOUS (Kearse et al., 2012) to adjust obvious alignment errors and remove any nonorthologous sequences. To assess nonorthology, gene trees were reconstructed for all individual locus alignments with RAxML version 8 (Stamatakis, 2014) using the GTRGAMMA model and the rapid bootstrap algorithm with 200 bootstrap replicates. Potential nonorthologs were evidenced by highly divergent congeneric or confamilial sequences and/or by inspecting preliminary gene trees (Hedin et al., 2019); sequences that resulted in paraphyletic Opiliones suborders or families (i.e., universally supported lineages) and/or extremely long branch lengths were considered nonorthologs or contaminants and removed from subsequent analyses. We were very conservative when considering the oldest museum specimens. For example, in the case of MCZ 36740, to consider a sequence as orthologous we only included it in a UCE matrix if that locus also included at least one other *Leiobunum* or the relatively closely related *Phalangium opilio*. After the above manual inspections, phylogenies were reconstructed using concatenated and unpartitioned matrices in RAXML version 8 using the GTRGAMMA model and the rapid bootstrap algorithm with 200 bootstrap replicates run on the CIPRES Portal (Miller, Pfeiffer, & Schwartz, 2010).

For all museum samples and a subset of fresh samples (i.e., the closest relative for each museum sample) we assessed the recovery and potential utility of several "standard Sanger" sequencing loci (mitochondrial cytochrome *c* oxidase subunit I [COI], nuclear 18S and 28S rRNAs, and histone H3 [H3]) historically used in multilocus phylogenetic analyses (e.g., Giribet, Vogt, González, Sharma, & Kury, 2010; Wheeler et al., 2017). In GENEIOUS, local nucleotide BLAST searches were conducted for all assemblies compared to sequences of each gene derived from confamilial/congeneric samples available on GenBank from previously published studies. As the utility of UCE-derived mitochondrial genomes and COI datasets have been demonstrated in previous studies (Derkarabetian, Castillo, Koo, Ovchinnikov, & Hedin, 2019; do Amaral et al., 2015; Hedin, Derkarabetian, Blair, et al., 2018; Zarza et al., 2018), we explored the phylogenetic utility of 28S + 18S rRNA derived from samples included in this study with published data derived from multiple studies (Boyer & Giribet, 2007; Boyer, Karaman, & Giribet, 2005; Burns, Hedin, & Shultz, 2012; Giribet et al., 2010; Hedin, Tsurusaki, Macías-Ordóñez, & Shultz, 2012; Vélez, Fernández, & Giribet, 2014). A phylogeny was reconstructed with a matrix partitioned by locus using RAxML version 8 as described above.

Finally, to explore the potential of museum specimens in species delimitation and population-level studies, we created a dataset consisting only of the five *Vonones* samples. Variation was assessed via a "smilogram" created using the PHYLUCE script phyluce_align_get_smilogram_from_alignments. "Smilograms" indicate the frequency of variable base positions across UCE loci in a given dataset as a function of the distance from the center of the UCE alignment, where flanking regions tend to have more variability relative to the core UCE region.

3 | RESULTS

3.1 | Taxon sampling and UCE statistics

A total of 15 museum samples were successfully included in analyses, with collection dates ranging in age from 2008 to 1865 (11–154 years

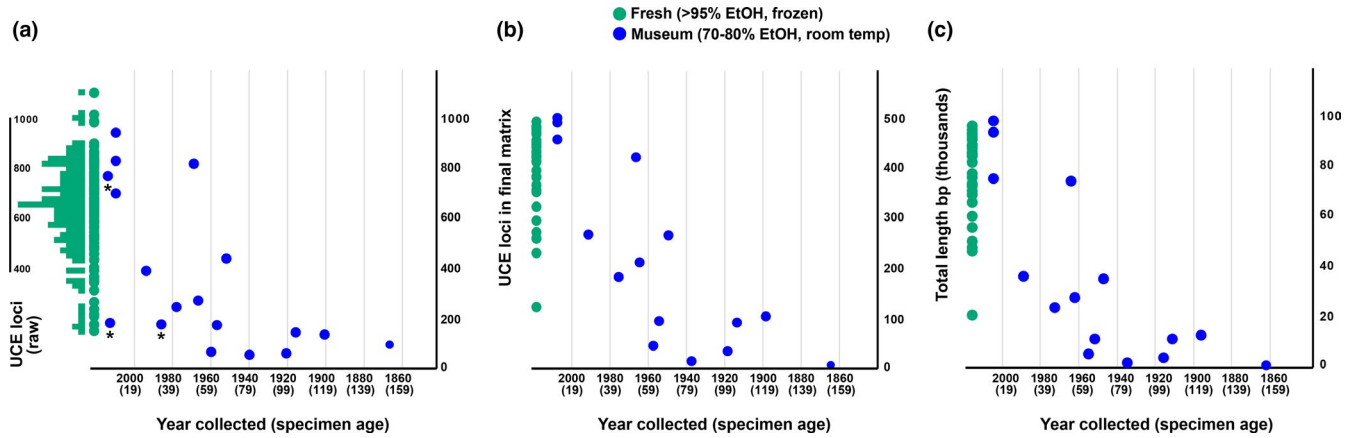

FIGURE 1 Number of UCE loci recovered in relation to specimen age. All fresh specimens are treated as a single time point regardless of their collection date. (a) The number of loci corresponds to raw UCE loci recovered from PHYLUCE. This plot includes fresh samples from previous studies with identical bioinformatic processing (Derkarabetian et al., 2019, 2018; Hedin, Derkarabetian, Ramírez, et al., 2018). Asterisks below museum samples indicate those samples from Hedin, Derkarabetian, Ramírez, et al. (2018). (b) The number of UCE loci corresponds to the number of loci in the CORE 65/65 matrix used in phylogenetic analyses. (c) Total length corresponds to the total number of base pairs (nongap) included in the final CORE 65/65 matrix after trimming with GBLOCKS and manual inspection. Note scale change [Colour figure can be viewed at wileyonlinelibrary.com]

old). Voucher, sequencing, and matrix statistics for all samples are presented in Table 1 and Table S1. Input DNA for museum specimens varied from 17 to 500 ng (average of 203 ng). We considered six additional museum specimens attempted as failed (Table S1); most of the failed samples were represented by very few reads, or an extremely low number of raw UCE loci (0–20) that were deemed to be contamination in preliminary explorations. These failed samples (six out of 21) were not included in any further analyses.

The mean coverage of all contigs for all samples was 20×, with a mean of 12× for fresh specimens and 37× for museum specimens. For the purposes of this study we consider a UCE locus sequenced if a contig is successfully matched to a UCE, regardless of contig length. When matching contigs to probes, using the 82/80 values resulted in an average decrease of 181 loci for the CORE dataset relative to the 65/65 dataset. Statistics for only the 65/65 analyses are mentioned below, but those for all datasets are included in Table S1. The average number of raw UCE loci, as determined by the "phyluce_assembly_get_fastas_from_match_counts" script, was 551 for all samples, with an average of 646 and 358 for fresh and museum specimens, respectively (Table S1). The number of loci recovered from recent museum specimens <20 years old spans the variation seen in fresh specimens, with a clear overall negative correlation with increasing specimen age across all museum specimens (Figure 1a,b). Not surprisingly, the oldest specimen (MCZ 36740, 154 years old; https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:36740) had the lowest number of loci in the final matrices (Figure 1a,b). Similarly, the total number of base pairs per sample included in the final matrix (following GBLOCKS and manual inspection) decreases with specimen age (Figure 1c). The mean locus length across all samples was 176 bp, with a mean of 193 bp for fresh specimens and 137 bp for museum specimens (Figure S1). The mean coverage of UCE loci for the 65/65 dataset was 29 × across all samples, with a mean of 14 × for fresh samples and a mean of 67 × for museum samples. Across multiple UCE studies using different preserved tissue types and different probe sets (this study; Blaimer, Lloyd, et al., 2016; McCormack et al., 2016), there are highly similar negative trends in the relationship between specimen age and the proportion of targeted UCEs successfully sequenced for each probe set (Figure 2). The average percentage of targeted UCE loci that are successfully sequenced decreases by 0.32%, 0.33% and 0.37% per year for ethanol-preserved, pinned and dried tissue types, respectively.

Increasing the percentage of museum samples included in the final matrix from 30% to 100% increased the amount of missing data for any given locus, resulting in more loci being dropped as bioinformatic filters removed loci with high levels of missing data, with a 25% average decrease (mean = 90 loci, range: 2–269) in the number of loci for museum specimens (Figure 3; Table S1). Decreasing the number of museum samples in the matrix from 30% to 17% increased the number of loci by an average of 58 (range: 13–103) across all fresh samples, and for those museum samples retained in the 17% matrix, increased the number of loci by an average of 60 loci (n = 6, range: 17–106). In the final CORE matrix with 50% taxon coverage, the average number of UCE loci across all samples was 345, with averages of 400 (range: 124–504) and 214 (range: 6–425) loci across fresh and museum samples, respectively (Table S1).

3.2 | Phylogenetic results

The resulting phylogeny of the CORE 65/65 dataset (Figure 4) shows all museum samples placed in the correct suborder and family, and resolved as expected based on current taxonomy. While there is some discordance in backbone topology between the 65/65 and 82/80 datasets (Figure 4; Figure S2), this discordance is largely *not* due to uncertainty in the placement of museum specimens, as all specimens (except "new genus nr *Triaenobunus*" NHMUC21B) are recovered as fully supported with their sister taxa in both trees.

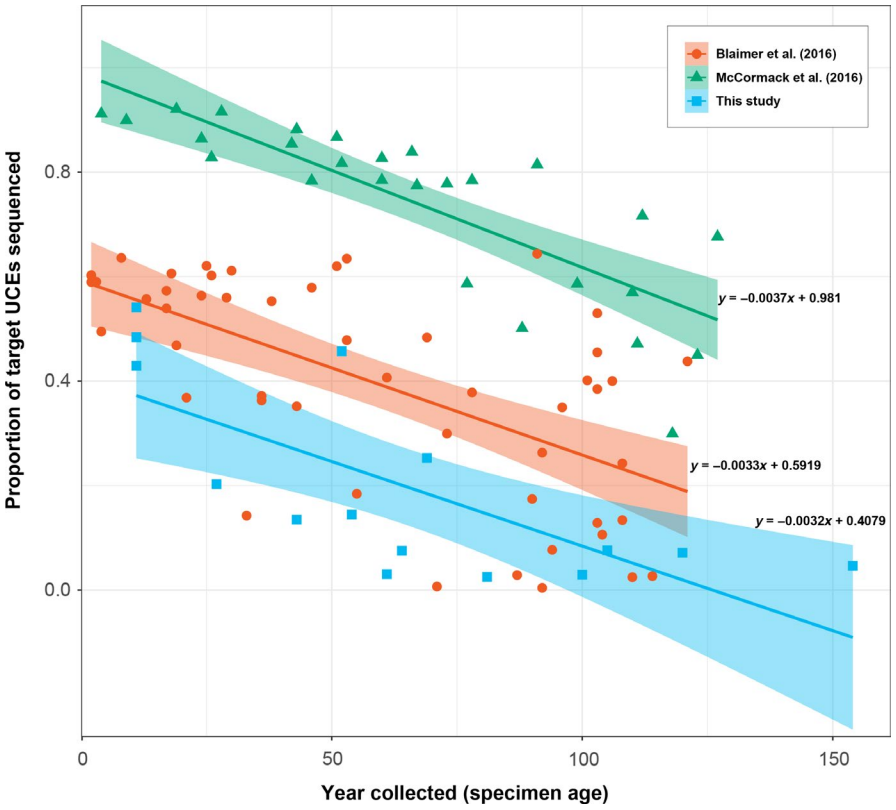

FIGURE 2 UCE sequencing capture success for historical specimens across multiple studies. Comparisons include this study using the Arachnida 1.1K probe set targeting 1,120 UCEs and two previously published studies using different UCE probe sets and preserved tissue types: McCormack et al. (2016) using dried bird tissue and the Tetrapod 5K probe set (5,060 UCEs); and Blaimer, Lloyd, et al. (2016) using pinned insects and the Hymenoptera 2.5K probe set (1,510 UCEs). For comparison across studies, numbers of loci recovered from each sample are scaled based on the total number of UCEs targeted by the specific probe set used [Colour figure can be viewed at wileyonlinelibrary.com]

Regarding the oldest specimen MCZ 36740 collected 154 years ago, while correctly placed in a clade with all other *Leiobunum* in both analyses, the relatively long branches and discordance in placement seen between analyses (Figure 4; Figure S2) indicate high levels of DNA degradation (see Discussion) and/or some additional inclusion of potential contamination/nonorthologous sequences despite our rigorous inspection, which removed over 90% of raw UCE loci as potential contaminants for this sample.

All *Vonones* samples form a fully supported monophyletic lineage and each species is reciprocally monophyletic. A 50% taxon-coverage matrix of all *Vonones* includes 119 loci, with a total length of 14,194 bp, including 1,188 variable sites, 80 of which are parsimony-informative. This species-level dataset shows increasing sequence variation in UCE loci with increasing distance from the UCE center, further indicating the utility of sequence capture in species and population-level analyses using museum specimens (Figure S3). All Triaenonychidae samples constitute a clade with internal relationships as expected based on ongoing work; here we provide brief justifications to support the correct placement of all triaenonychid museum samples. The *Triaenonychoides breviops* (MCZ 31332; https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:31332) sample from Chile collected in 1965 is recovered in a clade with the two other South American taxa included in this study, a relationship reflected in their shared and relatively unique genital morphology (Maury, 1987, 1988). The likely conspecific museum and fresh specimens of "new genus B" are fully supported as sister taxa. The unidentified triaenonychid from Madagascar (MCZ 52828; https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:52828) is recovered with all other Malagasy taxa included in this study. *Callihamina* sp. (NHMUC18A) is recovered as sister species to *Calliuncus odoratus* as expected based on morphology: these two genera are largely identical, separated solely by tarsal segmentation of the second leg, a highly variable character even within species. *Paranuncia ingens* (MCZ 48421; https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:48421) is recovered in a clade with *P. gigantea* and *Hickmanoxyomma*; these genera share highly similar genital morphology and are hypothesized to form a clade (Hunt, 1990). Both *Heteronuncia robusta* (MCZ 35919; https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:35919) and "new genus nr *Triaenobunus*" (NHMUC21B) are recovered with the fresh specimen of *Triaenobunus armstrongi*. This is a poorly studied and morphologically diverse Australian clade for which ongoing research indicates a high potential for undescribed taxa.

For the four traditional Sanger-sequenced loci targeted (COI, 18S rRNA, 28S rRNA, H3), UCE sequencing resulted in all loci being recovered from all samples in the subset of fresh specimens, except COI from "new genus B" (Table S1). For museum specimens, all four loci were recovered from 8/15 samples, while 18S and 28S rRNA were not recovered from two samples each, and COI and H3 were not recovered from four each. None of these loci were recovered from the oldest specimen attempted, namely *Leiobunum formosum* MCZ 36740. For COI, fresh samples were represented by full-length sequences, with the length of the top-hitting contigs over 1 kbp, while museum specimen COI sequences were not full length, and in some specimens collected prior to the 1960–1970s, COI was missing entirely. The 28S + 18S rRNA dataset had a total length of 3,006 bp and included UCE-derived data from 14/15 museum specimens and all nine fresh samples included in the subset (Figure 5; Table

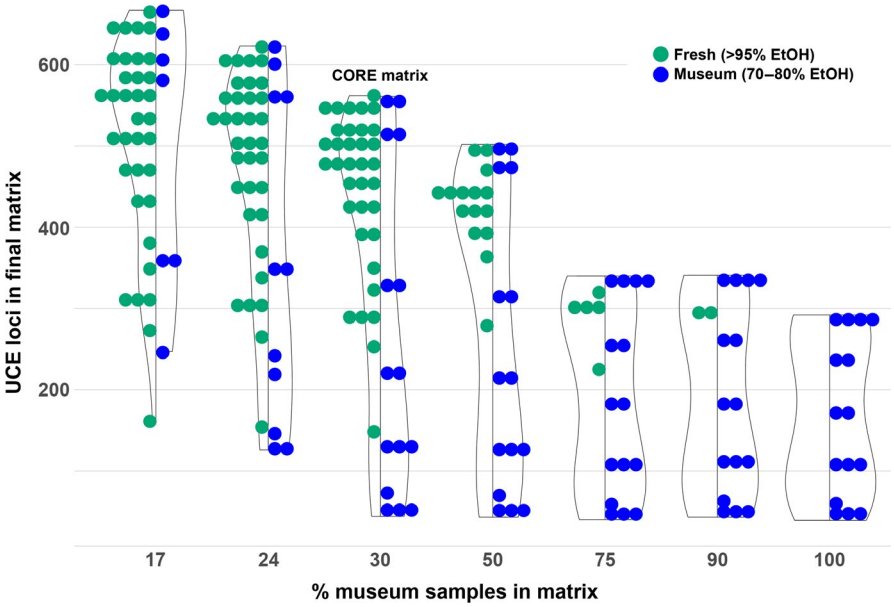

FIGURE 3 Number of loci in the final 50% taxon-coverage matrix per sample across data matrices with varying percentage of museum samples. The 30% matrix is the CORE matrix indicated in the text. Fresh samples are on the left side of each midline, museum samples are on the right. The left half of the violin plots is based on all samples included in the dataset (fresh and museum), while the right half of the violin plot is based on only museum samples [Colour figure can be viewed at wileyonlinelibrary.com]

S1). All fresh samples in each matrix were represented by full-length sequences. Sequence lengths of ribosomal genes for museum specimens were variable; all 2008 specimens were full length, with recovered sequence length decreasing beginning with specimens collected around 1960–1970. The resulting 28S + 18S rRNA phylogeny places all UCE samples as expected, including the museum sample of *Phalangium opilio* with a near identical sequence to two conspecific samples available from previous studies, and the two samples of *Leiobunum* each correctly recovered as the sister group to their respective conspecifics with very high support. Lack of resolution within *Vonones* could be attributed to the low level of variation found in these markers, which are often not useful for species-level phylogenetics in Opiliones.

4 | DISCUSSION

4.1 | Practical considerations

In this study we demonstrate the utility of sequence capture of UCEs in incorporating historical ethanol-preserved museum specimens in modern molecular systematics studies, allowing access to millions of historical specimens held in collections worldwide. It is clear that UCE locus recovery decreases with specimen age, and multiple UCE studies using different preserved tissue types and probe sets show similar decreasing rates of sequencing success in the range of 0.32%–0.37% loss of targeted UCE loci sequenced per year of preservation. Natural DNA degradation can have obvious impacts when using museum specimens in genomics-based research (McCormack et al., 2017). DNA degrades through both enzymatic and biochemical means, oxidation and hydrolysis being two major processes (Lindahl, 1993; Willerslev & Cooper, 2004). These processes can alter bases through deamination, mainly C to T transitions (Sawyer, Krause, Guschanski, Savolainen, & Pääbo, 2012), and depurination, which largely accounts for the fragmentation of DNA strands (Briggs et al., 2007). With increasing age of museum specimens, these processes lead to a decrease in sequencing success and decrease in sequence length. Here we show a clear decrease in average UCE locus length with increasing age, as reported in previous UCE studies focusing on museum specimens (Blaimer, Lloyd, et al., 2016; McCormack et al., 2016). Branch lengths may also increase with specimen age either through the above biochemical processes or from contamination (McCormack et al., 2016) and may possibly explain the relatively long branches of the oldest specimen, *Leiobunum formosum* MCZ 36740, as compared to its congeners.

Several previous studies have used ethanol-preserved museum specimens in sequencing experiments, using both traditional and modern sequence capture approaches. In a time-series study using spider specimens, Miller et al. (2013) showed high rates of COI sequencing success from specimens ~20 years old, and were able to acquire usable sequence data from specimens up to 60 years old. Targeting mitochondrial genomes via Illumina sequencing, Cotoras et al. (2017) successfully recovered partial mitochondrial genomes from spider samples collected from the 1940s, with a 242-bp fragment recovered from the oldest specimen collected in 1925 (92 years old). Most recently several UCE sequence capture studies have included museum specimens in phylogenomic analyses (Hedin et al., 2019; Hedin, Derkarabetian, Blair, et al., 2018; Hedin, Derkarabetian, Ramírez, et al., 2018), including one study specifically focusing on the utility of museum specimens (Wood et al., 2018). In these studies, successful sequencing of UCEs from the museum specimens was limited to those that were up to ~30 years old. Our study extends this success for sequence capture considerably to specimens 150+ years old, and the successful sequencing of mitochondrial loci from specimens up to 120 years old. In this study we were able to sequence historical ethanol-preserved specimens the same age as or older than preserved specimens reported in previous UCE sequence capture studies using other common preservation types, such as formalin fixation and dried tissue (Blaimer, Lloyd, et al., 2016; McCormack et al., 2016; Ruane & Austin, 2017).

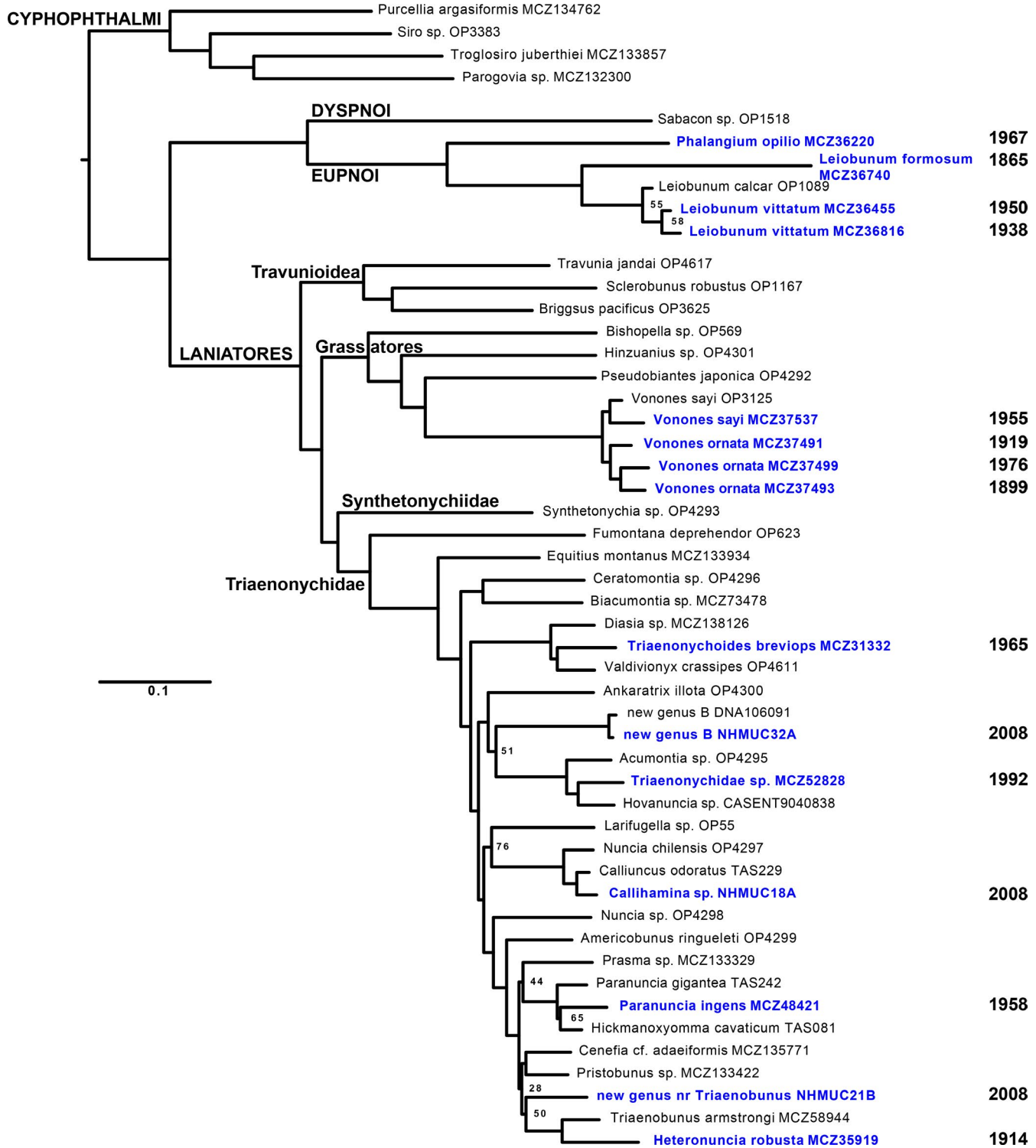

FIGURE 4 Phylogenetic results of the CORE 65/65 matrix. Museum samples are indicated in blue, with collection year for each museum specimen indicated at the right. Nodal support is only indicated for nodes with bootstrap values less than 80 [Colour figure can be viewed at wileyonlinelibrary.com]

The sample with the lowest number of UCEs in final matrices was the oldest specimen collected 154 years ago (1865). However, more loci were recovered from a specimen collected in 1899 than from specimens collected as recently as 1958. Similarly, across all samples, the specimen with the highest raw locus recovery was a museum specimen collected in 2008. The idiosyncrasies of preservation and storage history across collections, museums, and specimens might be reasons that a more predictable trend in the correlation of locus recovery and specimen age is not seen, although these factors were not tested here. Although we only included two specimens collected prior to the 1900s, it is possible that we approached the upper limit of specimen age for which we can

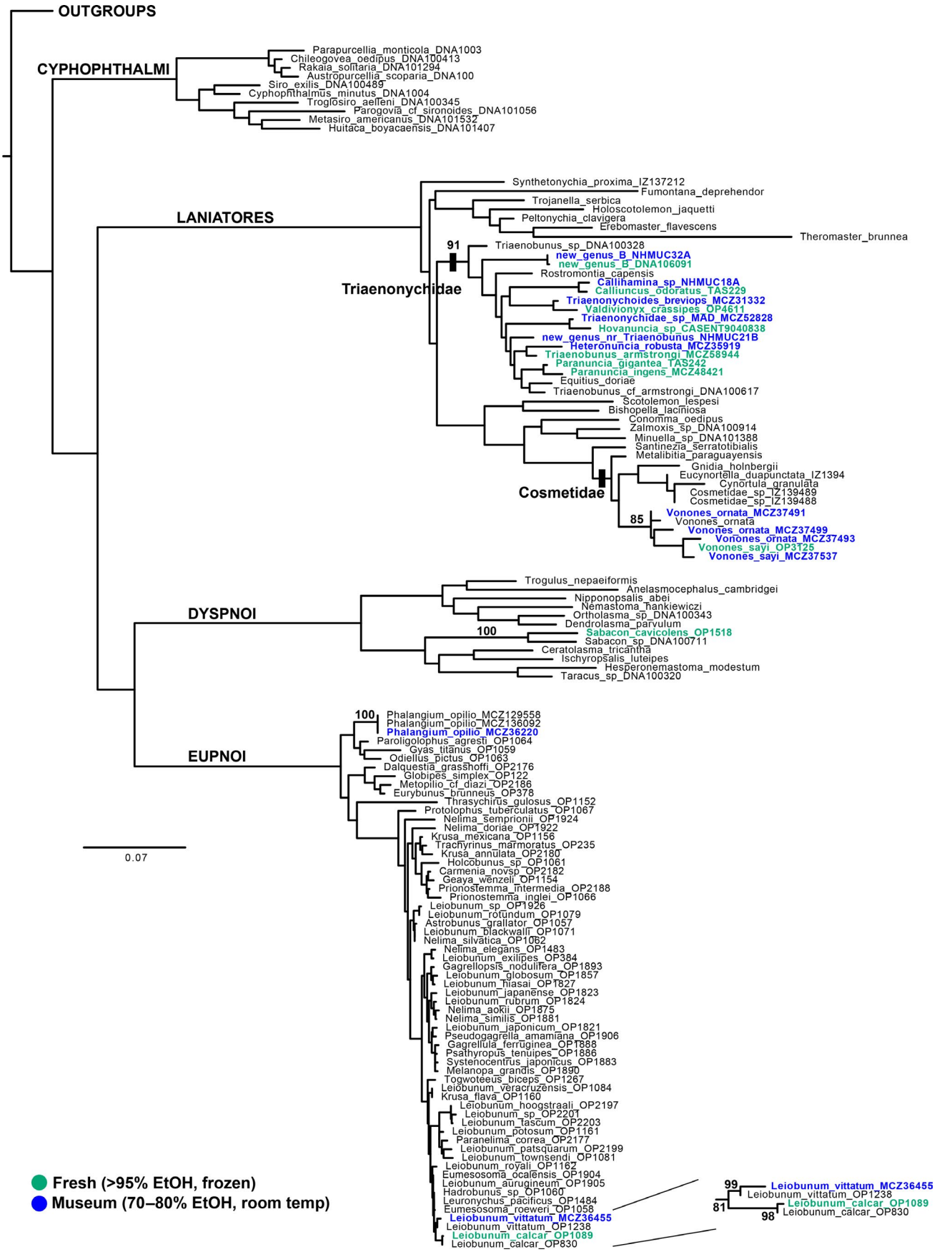

FIGURE 5 Phylogenetic results of the 18S + 28S rRNA matrix. UCE samples are indicated by color; fresh samples are green and museum samples are blue. Bootstrap support values are only shown for relevant nodes [Colour figure can be viewed at wileyonlinelibrary.com]

successfully sequence ethanol-preserved museum specimens using the current arachnid-specific probe set and methods used in this paper. However, the number of specimens in natural history collections older than 1865 is relatively small compared to more recent specimens, as only a handful of specimens exist mostly in European museums. The oldest specimen included in this study collected in 1865 had very few loci in the final matrices, relatively long branches in resulting phylogenies, and recovery of all traditional Sanger-sequenced loci failed. For this specimen, only 8% of raw UCE loci were retained after removing potential contamination, while an average of ~89% of raw loci were retained for fresh specimens. Despite the significant amount of contamination and "noise" due to natural DNA degradation, the successful placement of this sample with congeners suggests a measure of UCE sequencing success as enough phylogenetic signal is present to correctly infer placement of this specimen at the genus level.

Increasing locus recovery for 150+ year-old specimens and extending success even further back in time is likely given some protocol adjustments. A major challenge in incorporating museum specimens will be increasing the probability of success (e.g., numbers of loci) for samples with both low input DNA and high degradation. In this study, we chose large-bodied specimens to maximize output DNA, although some museum specimens yielded low input relative to comparably sized fresh specimens. More specialized and less commonly used extraction protocols have been applied with increasing success in small-bodied taxa and/or extractions using minimal tissue (Lienhard & Schäffer, 2019; Sproul & Maddison, 2017; Tin et al., 2014). Other extraction techniques tested on small-bodied low-input samples, for example using Chelex (Lienhard & Schäffer, 2019), may yield more DNA than the method used here, but this comparison remains to be tested. In extremely small-bodied specimens with low input DNA, whole genome amplification following DNA extraction has been successfully used to increase DNA yield for library preparation (Branstetter, Ješovnik, et al., 2017; Satler et al., 2018). It would be beneficial to test multiple whole genome amplification protocols specifically targeting small-bodied and highly degraded museum specimens. Of course, one obvious way to increase locus recovery for degraded specimens is to use a more specific probe set that targets more loci. A higher percentage (and number) of UCE loci were sequenced from pinned insects (39.7%; Blaimer, Lloyd, et al., 2016), formalin-fixed snakes (48.5%; Ruane & Austin, 2017) and dried birds (74.1%; McCormack et al., 2016) relative to the ethanol-preserved specimens in this study (20%). This is largely a result of two factors: (a) the higher number of loci targeted by the probe sets used in those other studies (5,060 loci in tetrapods and 1,510 loci in Hymenoptera) relative to the arachnid probe set targeting 1,120 loci; and (b) the more recent divergences those probe sets encompass. In the case of the arachnid-specific probe set, we reiterate the statements of Hedin et al. (2019) that the probe set encompasses taxa that diverged over 500 million years ago. Using probe sets that are more taxon-specific will increase locus recovery and decrease nontarget loci. For arachnids, progress has been made in this regard for the most diverse arachnid lineage, Acari (Van Dam, Trautwein, Spicer, & Esposito, 2019). The arachnid-specific probe set we used targets 1,120 loci, while the more specific Acari probe set targets 1,832 loci, with increased UCE recovery potentially even in nonacarine arachnids (Van Dam et al., 2019). Other considerations may certainly help locus recovery, including adjustments to the hybridization protocol (Li, Hofreiter, Straube, Corrigan, & Naylor, 2013; Paijmans, Fickel, Courtiol, Hofreiter, & Förster, 2016) and deeper sequencing, particularly for extremely important specimens (e.g., type or rare specimens).

Bossert and Danforth (2018) demonstrated that some UCE loci are shared across all arthropod UCE probe sets designed to date and correspond to genes with known function. Similarly, the vast majority (>98%) of arachnid UCEs were shown to be exonic in origin and annotated to genes with known functions (Hedin et al., 2019), making sequence capture of UCEs in arachnids essentially exon capture with the potential sequencing of adjacent intronic regions. The additional context the annotations and exon/intron boundaries provide for arachnid (or other arthropod) UCE origins is beneficial for downstream analyses, for example as an additional way to ensure sequence orthology for questionable sequences from museum specimens. Specific methods of data processing for museum specimens should be incorporated in sequence capture studies, for example bioinformatic or tree-based contamination removal (Kocot, Citarella, Moroz, & Halanych, 2013). Using higher thresholds when matching contigs to probes can reduce potential contamination, but the number of UCEs in the final matrices is significantly reduced, and some of the excluded loci are perhaps merely more variable on-target UCE loci. In our dataset, using the more conservative values of 82/80 to account for contamination reduced the number of UCE loci by an average of 34% across all samples. In museum samples, differentiating more variable UCE orthologs from orthologs where variation is due to natural DNA degradation is an additional complication. While contamination is a real concern when doing sequence capture with museum specimens, the amount and types of off-target contamination in empirical UCE datasets remains to be formally assessed.

Previous UCE studies demonstrated recovery of complete mitochondrial genomes (do Amaral et al., 2015; Zarza et al., 2018), and successful combination of UCE-derived COI sequences with traditional Sanger-sequenced COI data (Derkarabetian et al., 2019; Hedin, Derkarabetian, Blair, et al., 2018). Although not sequence capture studies, the retrieval of many traditional loci from museum samples has been demonstrated previously using modern sequencing technologies (Prosser et al., 2016), including the nuclear ribosomal complex (Sproul & Maddison, 2017). As in Branstetter, Longino, Ward, & Faircloth (2017), our study extends the utility of sequence capture of UCEs to include the ribosomal complex, here 18S and 28S rRNA, from both fresh and museum samples, furthering the ability to combine sequence capture data with many traditional Sanger-sequenced datasets. This success is particularly evident with the highly supported placement of two *Leiobunum* museum samples in a lineage with a relatively shallow diversification (Hedin et al., 2012).

4.2 | Sequence capture phylogenomics with historical specimens

The ability to incorporate museum specimens into modern molecular systematics offers plenty of research possibilities, as evidenced by the plethora of studies published in recent years. One obvious benefit in systematics is the inclusion of rare, difficult to collect, endangered or extinct taxa in phylogenomic analyses (Hedin et al., 2019; Hedin, Derkarabetian, Blair, et al., 2018; Hedin, Derkarabetian, Ramírez, et al., 2018; Oliveros et al., 2019; Tsai et al., 2019; Wood et al., 2018), or even datasets consisting entirely of museum specimens (Tsai et al., 2019). One important consideration will be the proportion of specimens in the final dataset that are historical museum specimens. Given that there is a decrease in UCE loci sequenced with increasing age, a higher proportion of museum specimens in the final matrix decreases the total number of loci retained for phylogenetic analyses. Given more specific probe sets and better genomic resources, obtaining thousands of loci in a final matrix is certainly possible for datasets composed entirely of museum specimens (Tsai et al., 2019). In this study we were able to include museum specimens of described and undescribed taxa for which recently collected specimens intended for genetic work are extremely limited or nonexistent. For example, in our ongoing examinations of Triaenonychidae taxa held in museums worldwide, *Paranuncia ingens* and *Callihamina* sp. were only available to us as historical museum specimens. Similarly, the unidentified and potentially undescribed taxon from Madagascar (Triaenonychidae sp. MCZ 52828) is only represented in collections by the single museum specimen sequenced in this study, for which we only used one set of appendages for DNA extractions.

Sequence capture methods have great utility at shallow taxonomic levels (Blaimer, LaPolla, Branstetter, Lloyd, & Brady, 2016; Derkarabetian et al., 2019; Hedin, Derkarabetian, Blair, et al., 2018; Smith, Harvey, Faircloth, Glenn, & Brumfield, 2013; Starrett et al., 2017; Tsai et al., 2019; Zarza et al., 2018) including population genomics (Bi et al., 2013), and given the high success in sequencing historical museum specimens, there is great potential for species delimitation. Clear phylogenetic structure and species-level variation are seen across all *Vonones* samples included in this study, further indicating the potential of incorporating museum specimens into both species- and population-level analyses (Hedin, Derkarabetian, Blair, et al., 2018; Tsai et al., 2019). Including historical type specimens in species delimitation studies is an obvious application, for example in assigning populations and specimens to species listed as federally endangered (Hedin, Derkarabetian, Blair, et al., 2018), or in helping to resolve long-standing taxonomic issues (Kehlmaier et al., 2019; McGuire et al., 2018). The use of nondestructive DNA extraction techniques (Tin et al., 2014) makes the inclusion of type specimens even more feasible.

For modern studies in taxa for which newly collected specimens destined for molecular and morphological work are both preserved in ethanol, preference may be given to molecular work, preserving specimens in >95% ethanol. This is perhaps to the detriment of potential morphological work, which for most taxa is better performed on specimens preserved in 70%–80% EtOH, as high-concentration ethanol makes specimens brittle and easily damaged during physical manipulation. We recover essentially the full complement of UCE loci from museum specimens collected in the last 10 years, with potential for hundreds of UCEs from specimens collected up to 60 years ago. This result makes it reasonable for more specimens in taxa that are targets of current/ongoing research to be preserved for morphological purposes, as recently collected ethanol-preserved "morphological specimens" are also viable for genetic work with essentially full recovery (Hedin, Derkarabetian, Blair, et al., 2018; Miller et al., 2013; Wood et al., 2018), as also recently suggested for dried bird tissue (Tsai et al., 2019). What remains to be tested is the relative success of sequencing DNA from specimens preserved in 70% ethanol and stored in freezers, something that is not common practice, as only "DNA-grade" specimens preserved in >95% EtOH are typically stored in freezers.

Given the success of including historical museum specimens in phylogenomic analyses, access to historical type specimen tissue should be more readily considered, for example in cases dealing with federally protected species where type specimens are critical in informing conservation decisions (Hedin, Derkarabetian, Blair, et al., 2018). Similarly, we argue that effort should be made to set aside tissue of historical type specimens for molecular work, whether the taxa are under active investigation or not, to halt DNA degradation and promote potential inclusion of type specimens in future molecular systematic studies.

ACKNOWLEDGEMENTS

This research was supported by funding from the Harvard Dean's Fund and NSF DEB-1754278 to G.G., and an NSF DDIG to S.D. (DEB-1601208). Specimens used in this study were provided by the Museum of Comparative Zoology, the San Diego State University Terrestrial Arthropod Collection, the Natural History Museum of Denmark–University of Copenhagen, and the California Academy of Sciences. Marshal Hedin graciously provided fresh specimens of key taxa. Some fresh specimens were collected in Tasmania under permit number FA16289. We thank Claire Bailey Reardon and Patrick Dennett of the Bauer Core Facility at Harvard University for assistance troubleshooting sequencing efforts, and Brant Faircloth and Carl Oliveros for initial discussion regarding sequence capture from museum specimens. This manuscript was significantly improved by constructive comments from Michael Hansen, Brant Faircloth, Marshal Hedin and two anonymous reviewers.

AUTHOR CONTRIBUTIONS

All authors contributed to research design, S.D. and L.R.B performed research, S.D. contributed new reagents, S.D. analysed data, S.D. wrote the paper, and all authors edited and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Raw reads for all newly sequenced samples have been deposited in the SRA (BioProject ID PRJNA553236). The final concatenated matrices, all individual locus alignments, and resulting phylogenies for the CORE 65/65 and 82/80 analyses are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.sh5b962.

ORCID

Shahan Derkarabetian https://orcid.org/0000-0002-9163-9277
Ligia R. Benavides https://orcid.org/0000-0003-4387-1220
Gonzalo Giribet https://orcid.org/0000-0002-5467-8429

REFERENCES

Abdelkrim, J., Aznar-Cormano, L., Fedosov, A. E., Kantor, Y. I., Lozouet, P., Phuong, P., … Puillandre, N. (2018). Exon-capture-based phylogeny and diversification of the venomous gastropods (Neogastropoda, Conoidea). *Molecular Biology and Evolution*, *35*, 2355–2374. https://doi.org/10.1093/molbev/msy144

Beaman, R. S., & Cellinese, N. (2012). Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, *209*, 7–17. https://doi.org/10.3897/zookeys.209.3313

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, *22*, 6018–6032. https://doi.org/10.1111/mec.12516

Blagoderov, V., Kitching, I. J., Livermore, L., Simonsen, T. J., & Smith, V. S. (2012). No specimen left behind: Industrial scale digitization of natural history collections. *ZooKeys*, *209*, 133–146. https://doi.org/10.3897/zookeys.209.3178

Blaimer, B. B., LaPolla, J. S., Branstetter, M. G., Lloyd, M. W., & Brady, S. G. (2016). Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Molecular Phylogenetics and Evolution*, *102*, 20–29. https://doi.org/10.1016/j.ympev.2016.05.030

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS ONE*, *11*, e0161531. https://doi.org/10.1371/journal.pone.0161531

Bossert, S., & Danforth, B. N. (2018). On the universality of target enrichment baits for phylogenomic research. *Methods in Ecology and Evolution*, *9*, 1453–1460. https://doi.org/10.1111/2041-210X.12988

Boyer, S. L., & Giribet, G. (2007). A new model Gondwanan taxon: Systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. *Cladistics*, *23*, 337–361. https://doi.org/10.1111/j.1096-0031.2007.00149.x

Boyer, S. L., Karaman, I., & Giribet, G. (2005). The genus *Cyphophthalmus* (Arachnida, Opiliones, Cyphophthalmi) in Europe: A phylogenetic approach to Balkan peninsula biogeography. *Molecular Phylogenetics and Evolution*, *36*, 554–567. https://doi.org/10.1016/j.ympev.2005.04.004

Branstetter, M. G., Ješovnik, A., Sosa-Calvo, J., Lloyd, M. W., Faircloth, B. C., Brady, S. G., & Schultz, T. R. (2017). Dry habitats were crucibles of domestication in the evolution of agriculture in ants. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170095. https://doi.org/10.1098/rspb.2017.0095

Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (2017). Enriching the ant tree of life: Enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, *8*, 768–776. https://doi.org/10.1111/2041-210X.12742

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prufer, K., … Paabo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, *104*, 14616–14621. https://doi.org/10.1073/pnas.0706065104

Burns, M., Hedin, M., & Shultz, J. W. (2012). Molecular phylogeny of the leiobunine harvestmen of eastern North America (Opiliones: Sclerosomatidae: Leiobuninae). *Molecular Phylogenetics and Evolution*, *63*, 291–298. https://doi.org/10.1016/j.ympev.2011.12.025

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44. https://doi.org/10.1016/j.jhevol.2014.10.015

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Cotoras, D., Murray, G., Kapp, J., Gillespie, R., Griswold, C., Simison, W., … Shapiro, B. (2017). Ancient DNA Resolves the History of Tetragnatha (Araneae, Tetragnathidae) Spiders on Rapa Nui. *Genes.*, *8*(12), 403.

Der Sarkissian, C., Pichereau, V., Dupont, C., Ilsøe, P. C., Perrigault, M., Butler, P., … Orlando, L. (2017). Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. *Molecular Ecology Resources*, *17*, 835–853. https://doi.org/10.1111/1755-0998.12679

Derkarabetian, S., Castillo, S., Koo, P. K., Ovchinnikov, S., & Hedin, M. (2019). An empirical demonstration of unsupervised machine learning in species delimitation. *Molecular Phylogenetics and Evolution*, *139*, https://doi.org/10.1016/j.ympev.2019.106562

Derkarabetian, S., Starrett, J., Tsurusaki, N., Ubick, D., Castillo, S., & Hedin, M. (2018). A stable phylogenomic classification of Travunioidea (Arachnida, Opiliones, Laniatores) based on sequence capture of ultraconserved elements. *ZooKeys*, *760*, 1–36. https://doi.org/10.3897/zookeys.760.24937

do Amaral, F. R., Neves, L. G., Resende, M. F. R., Mobili, F., Miyaki, C. Y., Pellegrino, K. C. M., & Biondo, C. (2015). Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE*, *10*, e0138446. https://doi.org/10.1371/journal.pone.0138446

Faircloth, B. C. (2013). Illumiprocessor: A trimmomatic wrapper for parallel adapter and quality trimming. https://doi.org/10.6079/J9ILL

Faircloth, B. C. (2015). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*, 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, *8*, 1103–1112. https://doi.org/10.1111/2041-210X.12754

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, *15*, 489–501. https://doi.org/10.1111/1755-0998.12328

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*, 717–726. https://doi.org/10.1093/sysbio/sys004

Fernández, R., Kvist, S., Lenihan, J., Giribet, G., & Ziegler, A. (2014). *Sine systemate chaos*? A versatile tool for earthworm taxonomy: Non-destructive imaging of freshly fixed and museum specimens using micro-computed tomography. *PLoS ONE*, *9*, e96617. https://doi.org/10.1371/journal.pone.0096617

Giribet, G., Vogt, L., González, A. P., Sharma, P., & Kury, A. B. (2010). A multilocus approach to harvestman (Arachnida: Opiliones) phylogeny with emphasis on biogeography and the systematics of Laniatores. *Cladistics*, *26*, 408–437. https://doi.org/10.1111/j.1096-0031.2009.00296.x

Glenn, T. C., Nilsen, R., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger Jr, J. W. & … B. C. (2019). Adapterama I: universal stubs and primers for thousands of dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *bioRxiv*. https://doi.org/10.1101/049114

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644. https://doi.org/10.1038/nbt.1883

Hedin, M., Derkarabetian, S., Alfaro, A., Ramírez, M. J., & Bond, J. E. (2019). Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. *PeerJ*, *7*, e6864. https://doi.org/10.7717/peerj.6864

Hedin, M., Derkarabetian, S., Blair, J., & Paquin, P. (2018). Sequence capture phylogenomics of eyeless *Cicurina* spiders from Texas caves, with emphasis on US federally-endangered species from Bexar County (Araneae, Hahniidae). *ZooKeys*, *769*, 49. https://doi.org/10.3897/zookeys.769.25814

Hedin, M., Derkarabetian, S., Ramírez, M. J., Vink, C., & Bond, J. E. (2018). Phylogenomic reclassification of the world's most venomous spiders (Mygalomorphae, Atracinae), with implications for venom evolution. *Scientific Reports*, *8*, 1636. https://doi.org/10.1038/s41598-018-19946-2

Hedin, M., Tsurusaki, N., Macías-Ordóñez, R., & Shultz, J. W. (2012). Molecular systematics of sclerosomatid harvestmen (Opiliones, Phalangioidea, Sclerosomatidae): Geography is better than taxonomy in predicting phylogeny. *Molecular Phylogenetics and Evolution*, *62*, 224–236. https://doi.org/10.1016/j.ympev.2011.09.017

Hunt, G. S. (1990). *Hickmanoxyomma*, a new genus of cavernicolous harvestmen from Tasmania (Opiliones: Triaenonychidae). *Records of the Australian Museum*, *42*, 45–68. https://doi.org/10.3853/j.0067-1975.42.1990.106

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

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. https://doi.org/10.1093/molbev/mst010

Kehlmaier, C., Zhang, X., Georges, A., Campbell, P. D., Thomson, S., & Fritz, U. (2019). Mitogenomics of historical type specimens of Australasian turtles: Clarification of taxonomic confusion and old mitochondrial introgression. *Scientific Reports*, *9*, 5841. https://doi.org/10.1038/s41598-019-42310-x

Kocot, K. M., Citarella, M. R., Moroz, L. L., & Halanych, K. M. (2013). PHYLOTREEPRUNER: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics*, *9*, EBO-S12813. https://doi.org/10.4137/EBO.S12813

Li, C., Hofreiter, M., Straube, N., Corrigan, S., & Naylor, G. J. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques*, *54*, 321–326. https://doi.org/10.2144/000114039

Lienhard, A., & Schäffer, S. (2019). Extracting the invisible: Obtaining high quality DNA is a challenging task in small arthropods. *PeerJ*, *7*, e6753. https://doi.org/10.7717/peerj.6753

Lim, H. C., & Braun, M. J. (2016). High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources*, *16*, 1204–1223. https://doi.org/10.1111/1755-0998.12568

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*, 709–715. https://doi.org/10.1038/362709a0

Maury, E. A. (1987). Triaenonychidae sudamericanos. IV. El género *Triaenonychoides* H. Soares 1968 (Opiliones, Laniatores). *Boletín De La Sociedad De Biología De Concepción*, *58*, 95–106.

Maury, E. A. (1988). Triaenonychidae Sudamericanos. III. Descripción de los nuevos géneros *Nahuelonyx* y *Valdivionyx* (Opiliones, Laniatores). *Journal of Arachnology*, *16*, 71–83.

McCormack, J. E., Rodríguez-Gómez, F., Tsai, W. L., & Faircloth, B. C. (2017). Transforming museum specimens into genomic resources 1. In: M. S. Webster (Eds.), *The Extended Specimen* (pp. 143–156). Boca Raton, FL: CRC Press, Boca.

McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, *16*, 1189–1203. https://doi.org/10.1111/1755-0998.12466

McGuire, J. A., Cotoras, D. D., O'Connell, B., Lawalata, S. Z. S., Wang-Claypool, C. Y., Stubbs, A., … Iskandar, D. T. (2018). Squeezing water from a stone: High-throughput sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards. *PeerJ*, *6*, e4470. https://doi.org/10.7717/peerj.4470

Miller, J. A., Beentjes, K. K., van Helsdingen, P., & IJland, S. (2013). Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. *ZooKeys*, *365*, 245–261. https://doi.org/10.3897/zookeys.365.5787

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010*, pp. 1–8, New Orleans, LA. Retrieved from http://www.phylo.org/sub_sections/portal/sc2010_paper.pdf

Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 20170391. https://doi.org/10.1098/rstb.2017.0391

Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., … Faircloth, B. C. (2019). Earth history and the passerine superradiation. *Proceedings of the National Academy of Sciences*, *116*, 7916–7925. https://doi.org/10.1073/pnas.1813206116

Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*, 42–55. https://doi.org/10.1111/1755-0998.12420

Prosser, S. W. J., deWaard, J. R., Miller, S. E., & Hebert, P. D. N. (2016). DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources*, *16*, 487–497. https://doi.org/10.1111/1755-0998.12474

Quattrini, A. M., Faircloth, B. C., Dueñas, L. F., Bridge, T. C. L., Brugler, M. R., Calixto-Botía, I. F., … McFadden, C. S. (2018). Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. *Molecular Ecology Resources*, *18*, 281–295. https://doi.org/10.1111/1755-0998.12736

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*, 939–946. https://doi.org/10.1101/gr.128124.111

Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, *17*, 1003–1008. https://doi.org/10.1111/1755-0998.12655

Rubi, T. L., Knowles, L. L., & Dantzer, B. (2019). Museum epigenomics: Characterizing cytosine methylation in historic museum specimens. *bioRxiv*, 620583.

Satler, J. D., Herre, E. A., Jandér, K. C., Eaton, D. A., Machado, C. A., Heath, T. A., & Nason, J. D. (2018). Inferring processes of coevolutionary diversification in a community of Panamanian strangler figs and associated pollinating wasps. *bioRxiv*. https://doi.org/10.1101/490862

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE*, *7*(3), e34131. https://doi.org/10.1371/journal.pone.0034131

Short, A. E. Z., Dikow, T., & Moreau, C. S. (2018). Entomological collections in the age of big data. *Annual Review of Entomology*, *63*, 513–530. https://doi.org/10.1146/annurev-ento-031616-035536

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2013). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, *63*, 83–95. https://doi.org/10.1093/sysbio/syt061

Soltis, D. E., & Soltis, P. S. (2016). Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Diversity*, *38*, 264–270. https://doi.org/10.1016/j.pld.2016.12.001

Sproul, J. S., & Maddison, D. R. (2017). Sequencing historical specimens: Successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*, *17*, 1183–1201. https://doi.org/10.1111/1755-0998.12660

Staats, M., Erkens, R. H. J., van de Vossenberg, B., Wieringa, J. J., Kraaijeveld, K., Stielow, B., … Bakker, F. T. (2013). Genomic treasure troves: Complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE*, *8*, e69189. https://doi.org/10.1371/journal.pone.0069189

Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W. Jr, McCormack, J. E., & Faircloth, B. C. (2017). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Molecular Ecology Resources*, *17*, 812–823. https://doi.org/10.1111/1755-0998.12621

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., … Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*, *11*, e0151651. https://doi.org/10.1371/journal.pone.0151651

Swanson, M. T., Oliveros, C. H., & Esselstyn, J. A. (2019). A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1902), 20190672. https://doi.org/10.1098/rspb.2019.0672

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*, 564–577. https://doi.org/10.1080/10635150701472164

Tin, M.-M.-Y., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for rad-tagging and low-coverage shotgun phylogenetics. *PLoS ONE*, *9*, e96793. https://doi.org/10.1371/journal.pone.0096793

Tsai, W. L. E., Mota-Vargas, C., Rojas-Soto, O., Bhowmik, R., Liang, E. Y., Maley, J. M., … McCormack, J. E. (2019). Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. *Molecular Phylogenetics and Evolution*, *136*, 29–34. https://doi.org/10.1016/j.ympev.2019.03.017

Van Dam, M. H., Trautwein, M., Spicer, G. S., & Esposito, L. (2019). Advancing mite phylogenomics: Designing ultraconserved elements for Acari phylogeny. *Molecular Ecology Resources*, *19*, 465–475. https://doi.org/10.1111/1755-0998.12962

van de Kamp, T., Schwermann, A. H., dos Santos Rolo, T., Lösel, P. D., Engler, T., Etter, W., … Krogmann, L. (2018). Parasitoid biology preserved in mineralized fossils. *Nature Communications*, *9*, 3325. https://doi.org/10.1038/s41467-018-05654-y

Vélez, S., Fernández, R., & Giribet, G. (2014). A molecular phylogenetic approach to the New Zealand species of Enantiobuninae (Opiliones: Eupnoi: Neopilionidae). *Invertebrate Systematics*, *28*, 565–589. https://doi.org/10.1071/IS14030

Webster, M. S. (2017). *The extended specimen: Emerging frontiers in collections-based ornithological research*. Boca Raton, FL: CRC Press.

Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A., Griswold, C. E., … Zhang, J. (2017). The spider tree of life: Phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics*, *33*, 574–616. https://doi.org/10.1111/cla.12182

Willerslev, E., & Cooper, A. (2004). Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 3–16. https://doi.org/10.1098/rspb.2004.2813

Wood, H. M., González, V. L., Lloyd, M., Coddington, J., & Scharff, N. (2018). Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution*, *127*, 907–918. https://doi.org/10.1016/j.ympev.2018.06.038

Zarza, E., Connors, E. M., Maley, J. M., Tsai, W. L., Heimes, P., Kaplan, M., & McCormack, J. E. (2018). Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (Sarcohyla; Hylidae). *PeerJ*, *6*, e6045. https://doi.org/10.7717/peerj.6045

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829. https://doi.org/10.1101/gr.074492.107

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Derkarabetian S, Benavides LR, Giribet G. Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Mol Ecol Resour*. 2019;19:1531–1544. https://doi.org/10.1111/1755-0998.13072