Squeezing water from a stone: High-throughput sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards

Jimmy A McGuire [Corresp., 1, 2], **Darko D Cotoras** [3, 4], **Brendan O'Connell** [5], **Shobi Z S Lawalata** [1], **Cynthia Y Wang-Claypool** [1, 2], **Alexander Stubbs** [1, 2], **Xiaoting Huang** [6], **Guinevere O U Wogan** [7], **Sarah M Hykin** [1, 2], **Sean B Reilly** [1, 2], **Ke Bi** [8], **Awal Riyanto** [9], **Evy Arida** [9], **Lydia L Smith** [1], **Heather Milne** [3], **Jeffrey W Streicher** [10], **Djoko T Iskandar** [11]

[1] Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California, United States
[2] Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States
[3] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, California, United States
[4] Entomology Department, California Academy of Sciences, San Francisco, California, United States
[5] Department of Biomolecular Engineering and Bioinformatics, Baskin School of Engineering, University of California, Santa Cruz, Santa Cruz, California, United States
[6] Key Laboratory of Marine Genetics and Breeding, Ocean University of China, Qingdao, China
[7] Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, California, United States
[8] Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, California, United States
[9] Museum Zoologicum Bogoriense (MZB), Indonesia Institute of Sciences (LIPI), Cibinong, West Java, Indonesia
[10] Department of Life Sciences, The Natural History Museum, London, London, United Kingdom
[11] School of Life Sciences and Technology, Institute of Technology Bandung, Bandung, West Java, Indonesia

Corresponding Author: Jimmy A McGuire
Email address: mcguirej@berkeley.edu

We used Massively Parallel High-Throughput Sequencing to obtain genetic data from a 145-year old holotype specimen of the flying lizard, *Draco cristatellus*. Obtaining genetic data from this holotype was necessary to resolve an otherwise intractable taxonomic problem involving the status of this species relative to closely related sympatric *Draco* species that cannot otherwise be distinguished from one another on the basis of museum specimens. Initial analyses suggested that the DNA present in the holotype sample was so degraded as to be unusable for sequencing. However, we used a specialized extraction procedure developed for highly degraded ancient DNA samples and MiSeq shotgun sequencing to obtain just enough low-coverage mitochondrial DNA (547 base pairs) to conclusively resolve the species status of the holotype as well as a second known specimen of this species. The holotype was prepared before the advent of formalin-fixation and therefore was most likely originally fixed with ethanol and never exposed to formalin. Whereas conventional wisdom suggests that formalin-fixed samples should be the most challenging for DNA sequencing, we propose that evaporation during long-term alcohol storage and consequent water-exposure may subject older ethanol-fixed museum specimens to hydrolytic damage. If so, this may pose an even greater challenge for

sequencing efforts involving historical samples.

Squeezing water from a stone: High-Throughput Sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards

Jimmy A. McGuire[1,2], Darko D. Cotoras[3,4], Brendan O'Connell[5], Shobi Z. S. Lawalata[1,2,6], Cynthia Y. Wang-Claypool[1,2], Alexander Stubbs[1,2], Xiaoting Huang[7], Guinevere O. U. Wogan[8], Sarah M. Hykin[1,2], Sean B. Reilly[1,2], Ke Bi[9], Awal Riyanto[10], Evy Arida[10], Lydia L. Smith[1], Heather Milne[3], Jeffrey W. Streicher[11], and Djoko T. Iskandar[12].

[1] Museum of Vertebrate Zoology, 3101 VLSB, University of California, Berkeley, CA 94720, USA
[2] Department of Integrative Biology. University of California, Berkeley, CA 94720, USA
[3] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[4] Entomology Department, California Academy of Sciences, 55 Music Concourse Dr., Golden Gate Park, San Francisco, CA 94118, USA
[5] Department of Biomolecular Engineering and Bioinformatics, Baskin School of Engineering, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[6] United in Diversity Foundation, Jalan Hayam Wuruk No. 28, Jakarta 10220, Indonesia
[7] Key Laboratory of Marine Genetics and Breeding, Ocean University of China, Ministry of Education, 5 Yushan Road, Qingdao 266003, China
[8] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA
[9] Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, 238 Koshland Hall, University of California, Berkeley, CA 94720, USA
[10] Museum Zoologicum Bogoriense (MZB), Indonesia Institute of Sciences (LIPI), Jalan Raya Bogor-Jakarta KM 46, Cibinong 16911, Indonesia
[11] Department of Life Sciences, The Natural History Museum, London SW7 5BD, United Kingdom
[12] School of Life Sciences and Technology, Istitut Teknologi Bandung, Labtek XI Building, 10 Jalan Ganesa, Bandung 40132, Indonesia

Corresponding Author: Jimmy A. McGuire

Email address: mcguirej@berkeley.edu

ABSTRACT

We used Massively Parallel High-Throughput Sequencing to obtain genetic data from a 145-year old holotype specimen of the flying lizard, *Draco cristatellus*. Obtaining genetic data from this holotype was necessary to resolve an otherwise intractable taxonomic problem involving the status of this species relative to closely related sympatric *Draco* species that cannot otherwise be distinguished from one another on the basis of museum specimens. Initial analyses suggested that the DNA present in the holotype sample was so degraded as to be unusable for sequencing. However, we used a specialized extraction procedure developed for highly degraded ancient DNA samples and MiSeq shotgun sequencing to obtain just enough low-coverage mitochondrial DNA (547 base pairs) to conclusively resolve the species status of the holotype as well as a second known specimen of this species. The holotype was prepared before the advent of formalin-fixation and therefore was most likely originally fixed with ethanol and never exposed to formalin. Whereas conventional wisdom suggests that formalin-fixed samples should be the most challenging for DNA sequencing, we propose that evaporation during long-term alcohol storage and consequent water-exposure may subject older ethanol-fixed museum specimens to hydrolytic damage. If so, this may pose an even greater challenge for sequencing efforts involving historical samples.

INTRODUCTION

The advent of Massively Parallel High-Throughput Sequencing (HTS) has dramatically altered the manner in which geneticists conduct their research. This is certainly true for molecular phylogeneticists and population geneticists, who now routinely have access to large multilocus genetic datasets for non-model organisms. Because HTS using the Illumina platform involves sequencing of small fragments of DNA, this approach offers the potential to access previously unattainable genome-scale sequence data even for degraded historical samples (e.g., Prüfer et al., 2014; Palkopoulou et al., 2015). Millions of fluid-preserved specimens in museum collections predate the development of allozyme and DNA sequencing technologies, and thus lack specially preserved tissue samples for genetic analysis. Formalin-fixed fluid specimens usually having highly fragmented and cross-linked DNA, are often refractory to sequencing efforts using traditional Sanger sequencing. However, recent studies have shown that it is possible to obtain genomic DNA sequences from some of these fluid-preserved museum specimens. Hykin, Bi & McGuire (2015) demonstrated that low-coverage genomic sequences could be recovered from a 30-year old formalin-fixed museum specimen, though they were unsuccessful with a ~100-year old specimen. Ruane and Austin (2016) sequenced Ultra-Conserved Elements (UCEs) from both formalin-fixed (n=11) and ethanol-fixed (n=10) museum specimens, including one sample that was collected between 1878 and 1911. Both had mixed success, with the quantity of DNA recovered in the extraction stage likely playing the largest role in the performance of their sequencing efforts. Notably, the samples that failed in Ruane and Austin's (2016) experiment included subsets of both their formalin- (7 of 16) and alcohol-fixed (4 of 5) samples, indicating that old alcohol-preserved museum specimens are not necessarily less problematic than those initially fixed with formalin. This is surprising given that contemporary tissue samples earmarked for genetic analysis are routinely stored in 95% ethanol.

Our objective in this study was to address an otherwise intractable problem in flying lizard taxonomy using Illumina HTS and ancient DNA methods for a 145-year old fluid-preserved holotype specimen. The nettlesome taxonomic issue involves a small clade of poorly known flying lizards (genus *Draco*, Agamidae) that, for reasons outlined below, was unlikely to be resolved without obtaining genetic data from the holotype specimen of one of the constituent species, *Draco cristatellus*. Determining species limits within this small clade (the *Draco fimbriatus* group) has proven challenging for taxonomists, and we first describe the convoluted taxonomic history of the clade as a justification for our ultimate solution to this question involving HTS. The *D. fimbriatus* group currently includes four recognized species: *D. abbreviatus*, *D. cristatellus, D. fimbriatus*, and *D. maculatus*. This taxonomic framework is based on Manthey (2008) and was followed by the widely utilized Reptile Database (Uetz, 2006). For reasons that we will describe in a subsequent paper, we instead utilize an alternative taxonomy that includes *D. cristatellus*, as well as *D. fimbriatus* (= *D. abbreviatus* above), *D. hennigi*, *D. punctatus* (= *D. fimbriatus* above), and *D. maculatus.* We note that our taxonomy differs from that of Manthey (2008) primarily as a consequence of having information that indicates that the type locality of *D. fimbriatus* is the Malay Peninsula rather than Java. We further note that our recognition of both *D. cristatellus* and *D. punctatus* is tentative, as a primary objective of this paper is to resolve whether these are in fact distinct species.

The *Draco fimbriatus* group is taxonomically challenging. Although *Draco maculatus* is abundant, easily sampled in the field, and easily distinguished from other members of the group based on external phenotype, the remaining members of this clade are only rarely encountered, with relatively few specimens represented in museum collections. These species are canopy specialists (McGuire, 2003), making them more difficult to detect and more challenging to collect than other *Draco* taxa. Furthermore, the species comprising the *D. fimbriatus* group, as well as several other *Draco* clades are primarily distinguished on the basis of differences in coloration of their display structures (dewlap and patagia for most *Draco* taxa, just the dewlap among the relevant members of the *D. fimbriatus* group). For example, two major clades — the 'Philippines *volans* group' (McGuire and Alcala, 1999) and the '*Draco lineatus* group' (McGuire et al., 2007) — are each composed of multiple species that are primarily distinguished on the basis of coloration. Because coloration fades in preservative, recognizing species-specific coloration characteristics generally requires experience with the species in the field and/or access to color imagery of the specimen in life. Thus, as museum specimens, the members of these clades are functionally cryptic sympatric species. In summary, for the *D. fimbriatus* group, the paucity of museum specimens, and the rarity with which specimens are observed in the field from throughout their collective ranges by single observers, has greatly impeded taxonomic progress.

Within the *Draco fimbriatus* group, a particularly challenging issue relates to the taxonomic standings of *D. cristatellus* Günther 1872 and *D. punctatus* Boulenger 1900. *Draco cristatellus* was described based on a single specimen collected by Mr. Alfred Hart Everett in Matang, Sarawak between 1869 and March of 1872 (when Günther's manuscript describing the species was submitted for publication). Although Everett collected the type specimen, the Trustees of the British Museum purchased it from Mr. W. Cutter, thereby making it available to Günther for description (see Günther, 1872). Because Günther presumably did not see the living specimen, he did not evaluate the coloration of the dewlap in life, which is essential for species identification within this group. Günther described the dewlap as 'golden-yellow, with a brown anterior edge', presumably from its preserved state. Subsequently, Boulenger (1900) described *D.*

punctatus from Bukit Larut on the Malay Peninsula, noting that the dewlap was lemon yellow in coloration. Although he did not attempt to diagnose *D. punctatus* from *D. cristatellus*, Boulenger (1900) was clearly aware of the latter species and explicitly considered his *D. punctatus* holotype to be taxonomically distinct. Indeed, Boulenger (1900) noted that he had examined a second specimen of *D. punctatus* from Sarawak that was also collected by Everett, remarking that he had previously referred that second specimen to *D. cristatellus*. Boulenger (1900) might be the last author to have had a clear idea about the taxonomic distinctiveness of *D. cristatellus* and *D. punctatus,* and it is a pity that he did not identify the character differences that he used to render his taxonomic decision. Although de Rooij (1915) recognized *D. fimbriatus*, *D. cristatellus* and *D. punctatus* as distinct species, subsequent authors synonymized one or more members of the group. Hennig (1936) synonymized *D. cristatellus* with *D. fimbriatus*, while continuing to recognize *D. punctatus.* In his monographic *Draco* taxonomic study, Musters (1983) opted to synonymize both *D. cristatellus* and *D. punctatus* with *D. fimbriatus*. In his competing taxonomic treatment, Inger (1983) recognized two species, *D. cristatellus* and *D. fimbriatus,* as valid species, but placed *D. punctatus* in the synonymy of *D. cristatellus*. Inger's (1983) recognition of two species was based in part on Grandison's (1972) report on two sympatric *D. fimbriatus* group species with distinct dewlap colorations on Gunung (Mt.) Benom on the Malay Peninsula. Whereas Grandison (1972) identified the two sympatric species as *D. fimbriatus* and *D. punctatus* (without commenting on the status of the Bornean *D. cristatellus*), Inger (1983) instead opted to treat *D. punctatus* as a synonym of *D. cristatellus*. This sensible decision was presumably made on the basis of the similar dewlap colorations of the *D. cristatellus* and *D. punctatus* holotypes ('golden-yellow, with a brown anterior edge' vs. 'lemon yellow'). Inger (1983) furthermore attempted to differentiate his conceptions of *D. fimbriatus* and *D. cristatellus* using a statistical analysis of eight linear measurements and scale counts. Although he successfully sorted his sample into two groups on the basis of overlapping but significantly distinct character state differences, his *a priori* placement of *D. punctatus* in the synonymy of *D. cristatellus* effectively precluded the possibility that three species — *D. cristatellus, D. fimbriatus*, and *D. punctatus* — might all co-occur on the Greater Sunda Shelf (and particularly on Borneo). Here we address this open question taking advantage of two critical developments: (1) the acquisition and analysis of a key specimen (TNHC 56763) obtained by JAM from Santubong, Sarawak, Malaysian Borneo in 1996, and (2) an analysis of the 145-year old *D. cristatellus* holotype using ancient DNA extraction methods and HTS on the Illumina platform.

MATERIALS AND METHODS

(a) DNA extraction and sequencing from the *Draco cristatellus* holotype

We obtained from the Natural History Museum in London liver tissue from the holotype of *Draco cristatellus* (specimen BMNH 1872.2.19.4). This specimen was originally collected and prepared prior to March 1872, well before the advent of formalin-fixation.

At that time, the standard practice for fluid preservation of reptiles, amphibians, and fishes was direct preservation in "pure spirits of wine" (Günther, 1880). Thus, the holotype was most likely initially fixed in 90-100% ethanol (= Günther's "pure spirits of wine") and never exposed to formalin. Nevertheless, we opted to perform our initial DNA extraction using the methodology described in Hykin, Bi & McGuire (2015) for formalin-fixed tissues, with the goal being to perform an exome-capture experiment with this sample. The Hykin, Bi & McGuire (2015) procedure involves a series of initial ethanol washes followed by treatment in a heated alkali

buffer solution to break cross-linkages before standard phenol-chloroform extraction. When this extraction returned a very low (potentially zero) yield, we performed a second pair of phenol-chloroform extractions involving phase-lock gel tubes followed by SPRI bead clean-up. This second round of extractions was performed with and without exposure to heated alkali solution. These extractions also failed to return sufficient DNA to move forward with library preparation. Despite minimal DNA yield, we made an attempt to PCR-amplify and sequence a short fragment of the mitochondrial ND2 gene from both DNA extracts. These experiments resulted in the amplification and sequencing of human ND2 in two separate experiments. Both of our low-yield DNA extractions were then sent to MYcroarray Inc. in Ann Arbor, MI where they were subjected to an extra silica purification designed for low-concentration fragment retention, and prepared as libraries. However, the library preparation retrieved only artifact and we did not proceed to targeted enrichment of selected exons or sequencing.

At this stage, we engaged with a lab specializing in genetic analysis of ancient DNA samples, with extraction and sequencing performed in a facility specifically designed for work with ancient samples (the Shapiro Lab at UC Santa Cruz). No reptile work had previously been done in this facility and all work followed lab standards for working with historical samples (Fulton, 2011). The DNA extraction protocol was based on Dabney et al. (2013), Tin, Economo & Mikheyev (2014), and D. D. Cotoras et al. (2017). An initial subsample of 40 mg of tissue was subdivided into ~1mm pieces and suspended in lysis buffer. The composition of the 100 mL lysis buffer aliquot was: 5.3 mL 1 M Tris-HCl (pH 8.0), 5.3 mL 0.2 M EDTA, 10.6 mL 20% Sarkosyl, 1 mL 2-mercaptoethanol, and 77.8 mL distilled water. The tissue was digested with a total of 1 mL of lysis buffer with 1 mg/mL proteinase K, initially incubated overnight at 56°C, and then raised to 72°C for 1 hr. The 72°C incubation step was undertaken in case the sample had been exposed to formalin in order to reverse any potential crosslinks. Silica-based purification followed the centrifugation-based protocol described by Dabney et al. (2013). Briefly, 0.5 mL of 3M sodium acetate was added to the lysate and them transferred to a tube with 13 mL of binding buffer. The binding buffer is prepared in a 50 mL tube by first adding 23.88 g of guanidine hydrochloride and then adding water to bring the volume to 30 mL. A key element of this purification protocol is the high salt concentration of this binding buffer, which enhances recovery of short DNA fragments. After complete dissolution of the guanidine hydrochloride, 25 µl of Tween-20 and sufficient isopropanol to bring the total volume to 50 mL were added. The mixture of sample, binding buffer, and sodium acetate was transferred into a Zymo extension reservoir attached to a MiniElute spin column. The spin column was then centrifuged for 10 minutes at 1,000 rpm, after which the spin column was transferred to a 1.5 mL Eppendorf tube. We performed a dry spin for 1 minute at 13,000 rpm, followed by 2 washes with 750 µl of PE buffer (1 minute spin at 6,000 rpm). To ensure the entire PE buffer was removed, we did a dry spin for 1 minute at maximum speed. We eluted the purified extract in two volumes of 25 µl of TET. Each sample was centrifuged for 30 seconds at 13,200 RPM after 3-5 minutes of incubation. Because the elution displayed pigmentation, 25 µl of the extract was purified on a column filled with cross-linked polyvinylpyrrolidone (PVPP) (Arbeli and Fuentes, 2007). We also produced an extraction control consisting of lysis buffer that was subjected to the same set of procedures.

For genomic sequencing, we prepared two barcoded Illumina sequencing libraries (one for the holotype sample and one for the control) using the Meyer and Kircher (2010) protocol, starting with 5 µl of the PVPP purified DNA extraction. The same volume was used for the extraction control. The libraries were sequenced on an Illumina MiSeq machine using 150-cycle

v3 chemistry (2x75). Following sequencing, adaptors were removed from reads and sequences were merged using SeqPrep2 (https://github.com/jeizenga/SeqPrep2). Default parameters were used with the exception of the following: -q 20 -L 30 -A AGATCGGAAGAGCACACGTC -B AGATCGGAAGAGCGTCGTGT. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) confirmed that the sequence quality was good, with the normal base quality drop in the final five bases.

(b) Sanger sequence data for *Draco fimbriatus* group specimens

Our team has generated a large number of complete ND2 sequences for *Draco* specimens, including for 65 exemplars representing the *D. fimbriatus* group. These sequences were available for comparison with ND2 sequence fragments obtained from the *D. cristatellus* holotype. PCR-amplification was undertaken using the primers Metf1 and ALAr2, with cycle sequencing involving these external primers plus the internal primers Metf5 and ND2r6 (see McGuire and Kiew 2001 for details).

(c) Exome-capture and screening of mitochondrial DNA

For another project (McGuire et al. in prep), we generated an exome-capture data set using the MyBaits in-solution capture system for a set of 350 samples spanning all of *Draco*. This sample set included 14 *D. fimbriatus* group samples. The target loci for the exome capture include 1400 exons and flanking sequences derived from transcriptome sequences (jointly representing 709 loci), which were supplemented with an additional 540 lizard-specific UCE loci. Libraries enriched for our target loci were barcoded and sequenced on an Illumina Hi Seq 4000. Although our experiment was specifically designed to avoid capturing mitochondrial genes, mitochondrial sequences are so abundant in genomic DNA extractions that some mitochondrial molecules inevitably find their way into the off-target by-catch (non-target DNA sequences that are obtained during an exome-capture experiment). We took advantage of this imperfect filter to obtain mitochondrial sequences for comparison with the *D. cristatellus* holotype. For TNHC 56763, we used Geneious version 8.1.7 (Kearse et al. 2012) to obtain a mostly complete representation of mitochondrial coding genes by mapping our raw exome capture data (including off-target sequences) to the complete mitochondrial genome of *Acanthasaura armata* available on GenBank (AB266452.1). A preliminary assessment of the identity of the sequences was performed with a BLAST search after collapsing duplicate sequences with fastx_collapser (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html#fastx_collapser_usage). The result of the BLAST search was visualized with the program MEGAN (Huson et al., 2007). Processed reads were mapped with BWA mem (Li and Durbin, 2009) against the reference partial mitochondrial genome of TNHC 56763. Duplicates were removed with samtools rmdup (Li et al., 2009). A total of 17 unique reads mapped against the reference after duplicate removal. The represent a total of 777 bp of the 8114 bp reference. Most of the mapped regions had 1x coverage and portions of four contigs had 2x coverage. The average length of the mapped reads was 53 bp. Finally, for each of the 14 *D. fimbriatus* group samples included in our exome-capture experiment, we used Geneious to map our raw sequencing reads to 10 mitochondrial contigs obtained for the *D. cristatellus* holotype. The raw sequence data is available on the SRA database.

(d) Analysis of DNA sequence variation

Our analysis of DNA sequence variation included alignment of homologous DNA sequences and a simple count of nucleotide base substitutions between the *Draco cristatellus* holotype, sample TNHC 56763 from Santubong, Sarawak, and our selection of *D. fimbriatus* and *D. punctatus* samples from the Malay Peninsula, Sumatra, the Mentawai Islands, Java, and Borneo. Specimens examined are listed in Table 1. We also performed a heuristic parsimony analysis to obtain a phylogram for the *D. fimbriatus* group and performed a non-parametric bootstrap analysis with 1000 replicates to assess branch support. We did not perform a more rigorous maximum likelihood or Bayesian analysis because our primary objective was to assess uncorrected relative branch lengths. Phylogenetic analyses were performed in PAUP version 4 (Swofford 2002).

(e) Data availability and Permits

Sanger sequence data are available on GenBank (will be submitted if accepted) and a matrix for the mitochondrial ND2 gene for the *D. fimbriatus* group is included as supplemental materials. This research was undertaken in accordance with UC Berkeley Animal Use Protocol Number AUP-2014-12-6954. Fieldwork was undertaken with research permits issued by the Economic Planning Unit of Malaysia (UPE:40/200/19 SJ.363) and the Indonesia Institute of Sciences (LIPI: No. 2411/FRP/SM/X?2008 and No. 0115/FRP/SM/VI/2009).

RESULTS

The MiSeq run generated a total of 1,086,926 paired-end reads after combining the data from 2 different libraries (prepared identically) from the same extract. Raw sequences were merged if possible and duplicates collapsed, producing a total of 538,995 reads, which were BLAST searched (Altschul et al. 1990) against the NCBI database. 115,266 reads were successfully assigned, and of these, 47,400 hits corresponded to bacteria, 40,751 were assigned to mammals (of which 29,720 were specifically assigned to human), and 48 were assigned to *Anolis carolinensis*. Only three reads were assigned to *Draco*, each of which involved the mitochondrial ND2 sequence posted on GenBank for *Draco cristatellus* sample TNHC 56763 (see below). The paucity of *Draco* hits is not surprising given that there are no *Draco* reference genomes to map to. The extraction control was sequenced producing a total of 104,417 PE reads. After processing (adaptor removal, merged if possible, and duplicate removal) a total of 15,480 reads were assigned by the BLAST search. Bacteria were represented by 6,805 reads, mammals by 4,711 reads (of those, 3103 were assigned to human, and reptiles were assigned no reads (one read was assigned to chicken (*Gallus gallus*).

Of the three holotype ND2 reads, two were broadly overlapping, and the joint ND2 data obtained from the holotype totaled only 125 bp. To search for additional mitochondrial contigs in the holotype MiSeq data, we first generated a partial mitochondrial genome for TNHC 56763 from exome-capture off-target by-catch, which returned 8114 bp of protein-coding gene sequence data. Mapping the holotype data to the TNHC 56763 mitochondrial assembly resulted in the recovery of an additional 10 reads totaling 596 bp of mitochondrial sequence data representing six genes (COI, COXIII, ATPase8, ND4, ND4L, ND5). Thus, a total of 13 contigs were assembled with an average length of 61 bp. Most of the mapped regions had 1x coverage and parts of four contigs had 2x coverage. No reads were recovered when mapping the extraction blank against the same reference. In comparing the holotype sequence data with TNHC 56763 across the 721 bp of homologous sequence data, we found that the two samples were only weakly divergent from one another, sharing the same base calls at 710 of 721 positions for a raw

sequence divergence of 1.5%. We then mapped the raw reads from the exome captures for the remaining 13 *D. fimbriatus* group samples to the 10 *D. cristatellus* holotype contigs, which returned as few as two and as many as six homologous sequences per sample.

Comparison of the mitochondrial data obtained from the *Draco cristatellus* holotype with homologous data obtained for *D. fimbriatus* group samples found that the sample TNHC 56763 from Santubong, Sarawak, Malaysian Borneo was much more similar to the holotype than were any other *D. fimbriatus* group samples. When limiting our comparison to the six gene fragments for which we had between six and 13 corresponding sequences for comparison to the holotype, we found that TNHC 56763 was 0.9% divergent from the holotype, whereas all other samples ranged between 11.9% and 15.1% divergent. TNHC 56763 differed from the holotype at just five of 547 base positions, whereas the other samples differed from the holotype at from 36 of 303 bp to 70 of 465 base positions. The ND2 comparisons were most comprehensive because we had access to complete ND2 sequences for 65 *D. fimbriatus* group samples. Whereas TNHC 56763 differed from the holotype at three of 183 base positions, all other samples differed by at least 24 base positions. Notably, the two *D. punctatus* samples from Sarawak (the type locality for *D. cristatellus*) differed from the holotype at 28 and 30 of 183 base positions (15.3% and 16.4%, respectively).

A parsimony phylogenetic analysis of the ND2 gene including the *D. cristatellus* holotype strongly supports the monophyly of the holotype together with TNHC 56763 to the exclusion of all other *D. fimbriatus* group samples with 100% bootstrap support (Figure 1). Further, the 0.9% mitochondrial sequence divergence between TNHC 56763 and the holotype is within the scope of ND2 sequence variation that we observe between *D. punctatus* samples from the same locality. This does not consider the possibility that one or more of the five documented base substitutions could be sequencing errors resulting either from damage to the holotype DNA or random errors in our low coverage data, as we did not apply informatics pipelines developed to identify post-mortem damage of ancient DNA to our data (e.g. Mateiu and Rannala 2008, Molak and Ho 2011). Notably, one of the three inferred ND2 substitutions is a first position C->T change that would result in a proline to serine amino acid replacement, suggesting that this might be the result of post-mortem deamination of the template molecular or sequencing error.

DISCUSSION

In the present study, we applied HTS to a 145-year old fluid-preserved holotype specimen in an effort to disentangle an otherwise intractable taxonomic question. The problem stems from the fact that one of the species in this group, *Draco cristatellus*, was described using limited color information, and because fluid-preserved specimens representing multiple sympatric *D. fimbriatus* group species are often indistinguishable from one another without color information. Indeed, sympatric species in this complex are effectively cryptic once they have been prepared as museum specimens. This combination of circumstances rendered it virtually impossible to resolve the species status of *D. cristatellus* relative to *D. punctatus* and *D. fimbriatus,* two widespread species on the Greater Sunda Shelf. Importantly, a sample (TNHC 56763) collected in 1996 by JAM provides phylogenetic evidence for a third *D. fimbriatus* group species on Borneo, with the natural question being whether this sample is conspecific with the name-bearing holotype specimen of *D. cristatellus* housed in the British Museum of Natural History. Several species composition outcomes were possible, all of which were considered in a hypothesis-testing framework. In Hypothesis 1, *D. cristatellus* and *D. punctatus* are synonyms,

together representing a single species distinct from *D. fimbriatus* and TNHC 56763. In Hypothesis 2, *D. cristatellus* and *D. fimbriatus* are synonyms. In Hypothesis 3, *D. cristatellus* is a species distinct from *D. punctatus* and *D. fimbriatus* but conspecific with TNHC 56763. Finally, in Hypothesis 4, *D. cristatellus*, *D. fimbriatus*, *D. punctatus*, and TNHC 56763 all represent distinct species, with TNHC 56763 representing a fourth sympatric species on Borneo. The only way to conclusively test these alternative hypotheses was to obtain informative genetic data from the holotype specimen of *D. cristatellus* for comparison with TNHC 56763, and representative specimens of *D. fimbriatus* and *D. punctatus*.

We initially believed that genetic data would easily be retrieved from the *Draco cristatellus* holotype. The holotype was prepared before the advent of formalin-fixation, and we consequently had reason to believe that the specimen was originally fixed with ethanol and had never been exposed to formalin. Because tissue samples collected for genetic analysis are routinely stored in ethanol, we were confident that the holotype would still hold high molecular-weight DNA suitable for genomic sequencing. Our hope was to perform exome-capture with this sample and include it in a larger *Draco* phylogenomic data set. However, our initial attempts at extracting DNA from the sample using methods appropriate for historical and formalin-fixed tissues failed, forcing us to adjust both our approach and our expectations. Fortunately, our alternative hypotheses proved testable without comprehensive genomic data from the holotype. Indeed, analysis of the initial 125 bp of mitochondrial ND2 data identified when the holotype sequence data was subjected to GenBank BLAST allowed us to reject hypotheses 1, 2, and 4 in favor of hypothesis 3. The additional 422 bp of mitochondrial data obtained via mapping of holotype contigs to the reconstructed mitochondrial genome for TNHC 56763 simply provided additional confirmation that *D. cristatellus* and *D. punctatus* are each valid species, and that our specimen TNHC 56763 from Santubong, Sarawak is indeed a true *D. cristatellus* exemplar. This finding was a best-case scenario because all future specimens for which tissue samples are obtained can now be compared with known *D. cristatellus*, *D. punctatus*, and *D. fimbriatus* samples for genetic identification.

What lessons can be learned from our attempt to obtain genetic data from the *Draco cristatellus* holotype? First, even when initial attempts at extraction and quantification of DNA suggest that none is present, small numbers of DNA molecules may survive in the sample. For questions of simple species identification involving old and highly degraded samples, it may only be necessary to obtain limited data — even a few hundred base pairs of mitochondrial data may be sufficient to address the question. Our study shows that this is indeed possible even when initial assessments suggest that DNA in a tissue sample has been highly degraded. Obtaining data in these instances will likely require highly specialized extraction procedures such as the silica-column based extraction methodology utilized here, followed by short-fragment sequencing. Finally, we believe that the difficulty we confronted with this ethanol-fixed sample — which is consistent with the problems experienced by Ruane and Austin (2016) with their presumed alcohol-fixed samples (4 of 5 of which failed to sequence) — suggests that hydrolyzed ethanol-fixed tissues might be more problematic than formalin-fixed samples for genomic sequencing efforts (see Handt et al. 1994 for a description of DNA hydrolysis). This has important implications for curatorial practices. Not only is potential hydrolytic damage cause for concern with whole fluid specimens stored in ethanol, but it could be particularly problematic for tissue samples stored in ethanol that are not maintained in sub-zero degree conditions. Evaporation of ethanol from tissue vials might render even modern tissue samples virtually unusable for genetic analysis.

CONCLUSIONS

The development of HTS has revolutionized biological research by making genome-scale data readily available at a reasonable cost, even for non-model organisms. Systematists have fully embraced these advances in data acquisition for freshly sampled specimens, but are just beginning to harness HTS for the millions of fluid-preserved historical samples housed in natural history collections around the world. As we have shown here, acquiring genetic data from old museum specimens will sometimes present special challenges, but the information that can be gleaned from such specimens may be the only way to conclusively resolve previously intractable evolutionary and taxonomic questions.

ACKNOWLEDGEMENTS

The authors would like to thank Patrick Campbell and Donney Nicholson of The Natural History Museum, London for assistance processing the loan of liver tissue from the *D. cristatellus* holotype. Colin McCarthy and Oliver Crimmen kindly provided suggestions for locating information on A.H. Everett's Bornean collections and the history of fluid preservation at BMNH, respectively. We acknowledge Alison Devault and Jacob Enk at MYcroarray and MYreads both for advice and for generating the library preparation confirming the low yield of our initial DNA extraction. David Wake, Rauri Bowie, and members of the McGuire and Bowie labs provided valuable suggestions that improved the manuscript. JAM thanks the Economic Planning Unit of Malaysia and the Indonesian Institute of Sciences (LIPI) for providing scientific collecting permits.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, & D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.

Arbeli, Z. & C. L. Fuentes, 2007. Improved purification and PCR amplification of DNA from environmental samples. *FEMS Microbiology Letters* 272:269–275.

Boulenger, G. A. 1900. Descriptions of new batrachians and reptiles from the Larut Hills, Perak. *Annals and Magazine of Natural History* Series 7, Volume 6:186-193.

Cotoras, D. D., G. G. R. Murray, J. Kapp, R. G. Gillespie, C. Griswold, W. B. Simison, R. E. Green, & B. Shapiro. 2017. Ancient DNA resolves the history of *Tetragnatha* (Araneae, Tetragnathidae) spiders on Rapa Nui. *Genes* 8:403.

Dabney, J., M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Paabo, J.-L. Arsuaga, & M. Meyer. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110:15758–15763.

Fulton, T. L. 2011. Setting Up an Ancient DNA Laboratory. *Ancient DNA Methods and Protocols* (ed. by M. Shapiro, B.; Hofreiter), pp. 1–11. Humana Press, New York.

Grandison, A. G. C. 1972. The Gunong Benom Expedition 1967. 5. Reptiles and amphibians of Gunong Benom with a description of a new species of *Macrocalamus*. *Bulletin of the British Museum (Natural History)* 23:45-101.

462 Günther, A. 1872. On the reptiles and amphibians of Borneo. *Proceedings of the Zoological*
463 *Society of London* 1872:586-600.
464 Günther, A. 1880. An introduction to the study of fishes. Adam and Charles Black. Edinburgh.
465 743 pp.
466 Handt, O., M. Höss, M. Krings, & S. Pääbo. 1994. Ancient DNA: methodological challenges.
467 *Experientia* 50:524-529.
468 Huson, D. H., A. F. Auch, J. Qi, & S. C. Schuster. 2007. MEGAN analysis of metagenomic data.
469 *Genome Research* 17:377-386.
470 Hykin, S. M., K. Bi, & J. A. McGuire. 2015. Fixing Formalin: A Method to Recover Genomic-
471 Scale DNA Sequence Data from Formalin-Fixed Museum Specimens Using High-
472 Throughput Sequencing. *PLoS ONE* 10(10): e0141579. doi:10.1371/journal.pone.0141579
473 Inger, R. F. 1983. Morphological and ecological variation in the flying lizards (genus *Draco*).
474 *Fieldiana Zoology, New Series* 18:vi+35 pp.
475 Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A.
476 Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. L. Meintjes, & A. J. Drummond.
477 2012. Geneious Basic: An integrated and extendable desktop software platform for the
478 organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
479 Li, H. & R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler
480 Transform. *Bioinformatics*, 25:1754-60. [PMID: 19451168]
481 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
482 Durbin, & 1000 Genome Project Data Processing Subgroup. 2009. The Sequence
483 alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078-9. [PMID: 19505943]
484 Manthey, U. 2008. Agamid Lizards of Southern Asia. Draconinae 1. Edition Chimaira, Frankfurt
485 am Main, Germany.
486 Mateiu, L., & B. Rannala. 2008. Bayesian inference of errors in ancient DNA caused by
487 postmortem degradation. *Molecular Biology and Evolution* 25:1503-1511.
488 McGuire, J. A. 2003. Allometric prediction of locomotor performance: An example from
489 Southeast Asian flying lizards. *American Naturalist* 161:347-369.
490 McGuire, J. A., & A. C. Alcala. 2000. A taxonomic revision of the flying lizards of the
491 Philippine Islands (Iguania: Agamidae: *Draco*), with a description of a new species.
492 *Herpetological Monographs* 14:81-138.
493 McGuire, J. A., & B. H. Kiew. 2001. Phylogenetic systematics of Southeast Asian flying lizards
494 (Iguania: Agamidae: *Draco*) as inferred from mitochondrial DNA sequence data. *Biological*
495 *Journal of the Linnean Society* 72:203-229.
496 McGuire, J. A., R. M. Brown, Mumpuni, A. Riyanto, & N. Andayani. 2007. The flying lizards of
497 the *Draco lineatus* group (Squamata: Iguania: Agamidae): A taxonomic revision with
498 descriptions of two new species. *Herpetological Monographs* 21:180-213.
499 Molak, M., & S. Y. W. Ho. 2011. Evaluating the impact of post-mortem damage in ancient
500 DNA: A theoretical approach. *Journal of Molecular Evolution* 73:244-255.
501 Meyer, M. & M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed
502 target capture and sequencing. *Cold Spring Harbor Protocols*, **5**, 1–11.
503 Musters, C. J. M. 1983. Taxonomy of the genus *Draco* L. (Agamidae, Lacertilia, Reptilia).
504 *Zoologische Verhandelingen* 199:1-120 + 4 plates.
505 de Rooij, N. 1915. The Reptiles of the Indo-Australian Archipelago. I. Lacertilia, Chelonia,
506 Emydosauria. E. J. Brill Ltd., Leiden.

Palkopoulou, E., S. Mallick, P. Skoglund, J. Enk, N. Rohland, H. Li, A. Omrak, S. Vartanyan, H. Poinar, A. Götherström, D. Reich, & L. Dalén. 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology* 25:1395-1400.

Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu1, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, Janet Kelso, & S. Pääbo. 2014. The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* 505(7481):43-49.

Ruane, S., & C. C. Austin. 2016. Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources* doi:10.1111/1755-0998.12655.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachussetts.

Tin M. M.-Y., E. P. Economo, & A. S. Mikheyev. 2014. Sequencing Degraded DNA from Non-Destructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun Phylogenetics. *PLoS ONE* 9(5): e96793. doi:10.1371/journal.pone.0096793.

Uetz, P. 2006. The EMBL Reptile Database CD-ROM edition, March 2006.

Figure 1(on next page)

Phylogenetic tree for the *Draco fimbriatus* group including the *D. cristatellus* holotype.

Figure 1. Phylogenetic tree for the *Draco fimbriatus* group based on a parsimony analysis of the complete mitochondrial ND2 gene (1032 bp). The *D. cristatellus* holotype includes 183 bp of sequence data. Only two of 28 available *D. maculatus* samples were included to simplify the image. Non-parametric bootstrap values (1000 replicates) are superimposed on the single most parsimonious phylogram for select clades. The photo in the bottom left is *Draco punctatus*.

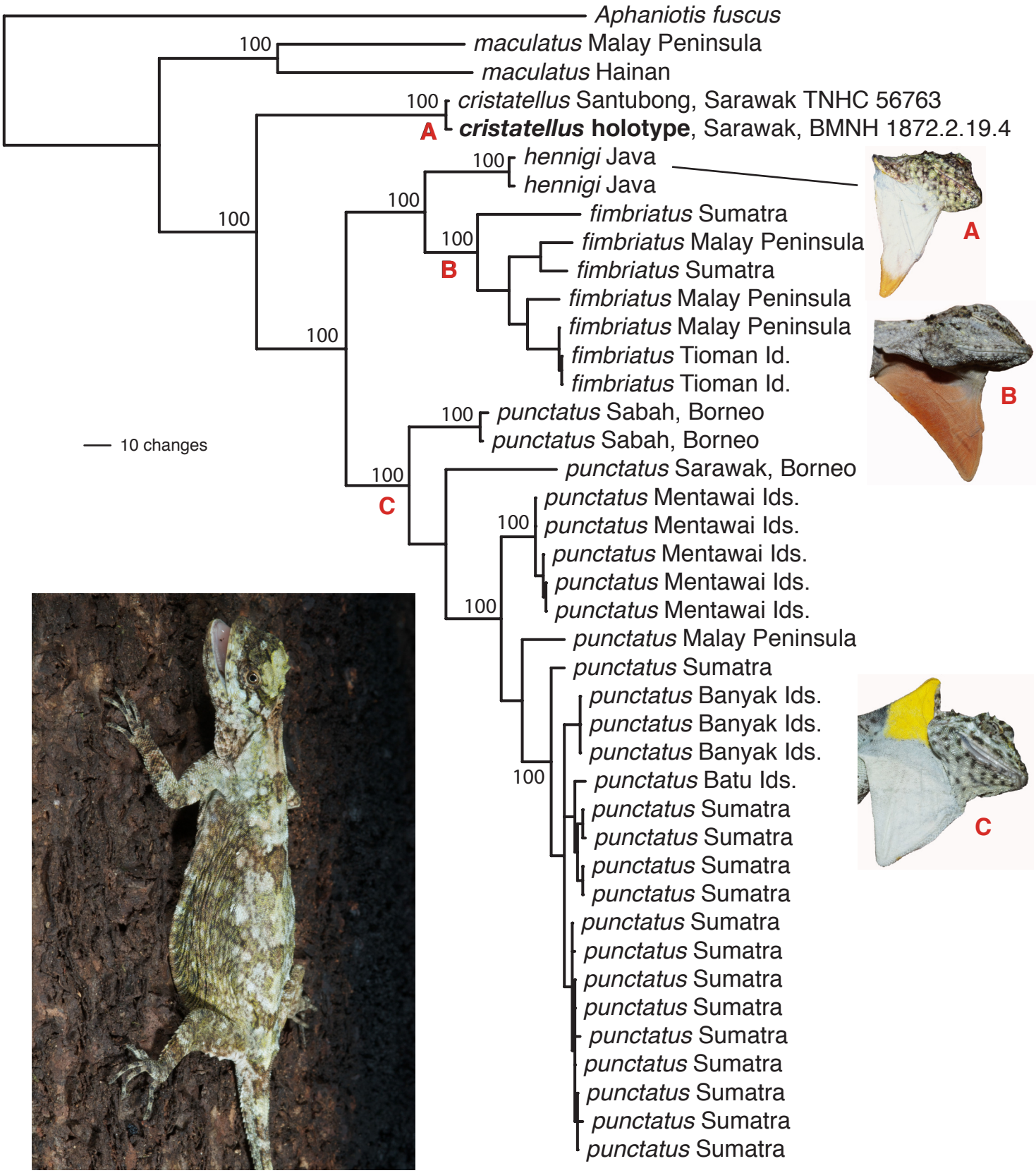

Table 1(on next page)

Numbers of base substitutions and sequence divergence values between the *Draco cristatellus* holotype and 14 exemplars representing the *D. fimbriatus* group.

Base pair differences between the *Draco cristatellus* holotype and each of 14 *D. fimbriatus* group samples for six mitochondrial genes. For ND2, the data used for comparisons were generated using standard Sanger sequencing. For all other genes, the data were derived from exome-capture off-target sequences. Mean sequence divergence values relative to the holotype are provided for each species.

1

	COX III contig3a	COI contig12a	COI contig12b	ND4L contig13	ND5 contig15	ND2 Sanger	Total	%
cristatellus Borneo TNHC 56763	0/43	1/79	1/81	0/79	0/64	3/183	5/547	0.9%
punctatus Borneo TNHC 56766		7/79	7/69			30/183	44/331	13.3%
punctatus Borneo TNHC 56764		6/79	8/81	10/69		28/183	52/412	12.6%
punctatus Malay Pen. LSUHC 5617	5/43			10/79		32/183	47/305	15.4%
punctatus Mentawai MVZ 270632	4/43	10/79	7/81	12/79		37/183	70/465	15.1%
punctatus Batu Ids MVZ 270636	4/43	10/79	7/81	12/79		35/183	68/465	14.6%
punctatus Sumatra MVZ 270835	4/43	9/79	7/81	11/79		35/183	66/465	14.2%
punctatus Banyak Ids MVZ 270829	4/43	10/79	7/81	9/59		37/183	67/465	14.4%
fimbriatus Mal Pen TNHC 57954	4/43	7/71	7/81	9/79	12/64	29/183	68/521	13.1%
fimbriatus Mal Pen TNHC 58565	4/43	10/79	7/80	9/79	12/64		42/345	12.2%
fimbriatus Sumatra MZB Lace.14276		9/79	5/56	8/79		32/183	54/397	13.6%
fimbriatus Sumatra MVZ 239473		9/79	8/81	8/79	11/64		36/303	11.9%
hennigi LSUMZ 81446		11/79		8/79	12/64	23/183	54/405	13.3%
hennigi LSUMZ 81447		11/79		8/79	13/64	23/183	55/405	13.6%

cristatellus	5/547 414/290	0.9%
punctatus	8 200/156	14.2%
fimbriatus	6	12.8%
hennigi	111/810	13.7%

2
3
4
5
6
7
8