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.

1 **Squeezing water from a stone: High-Throughput Sequencing from a 145-year old holotype**
2 **resolves (barely) a cryptic species problem in flying lizards**
3

4 Jimmy A. McGuire[1,2], Darko D. Cotoras[3,4], Brendan O'Connell[5], Shobi Z. S. Lawalata[1,2,6],
5 Cynthia Y. Wang-Claypool[1,2], Alexander Stubbs[1,2], Xiaoting Huang[7], Guinevere O. U. Wogan[8],
6 Sarah M. Hykin[1,2], Sean B. Reilly[1,2], Ke Bi[9], Awal Riyanto[10], Evy Arida[10], Lydia L. Smith[1],
7 Heather Milne[3], Jeffrey W. Streicher[11], and Djoko T. Iskandar[12].
8
9 [1] Museum of Vertebrate Zoology, 3101 VLSB, University of California, Berkeley, CA 94720, USA
10 [2] Department of Integrative Biology. University of California, Berkeley, CA 94720, USA
11 [3] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, 1156 High
12 Street, Santa Cruz, CA 95064, USA
13 [4] Entomology Department, California Academy of Sciences, 55 Music Concourse Dr., Golden Gate Park,
14 San Francisco, CA 94118, USA
15 [5] Department of Biomolecular Engineering and Bioinformatics, Baskin School of Engineering, University
16 of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
17 [6] United in Diversity Foundation, Jalan Hayam Wuruk No. 28, Jakarta 10220, Indonesia
18 [7] Key Laboratory of Marine Genetics and Breeding, Ocean University of China, Ministry of Education, 5
19 Yushan Road, Qingdao 266003, China
20 [8] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA
21 94720, USA
22 [9] Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, 238
23 Koshland Hall, University of California, Berkeley, CA 94720, USA
24 [10] Museum Zoologicum Bogoriense (MZB), Indonesia Institute of Sciences (LIPI), Jalan Raya Bogor-
25 Jakarta KM 46, Cibinong 16911, Indonesia
26 [11] Department of Life Sciences, The Natural History Museum, London SW7 5BD, United Kingdom
27 [12] School of Life Sciences and Technology, Istitut Teknologi Bandung, Labtek XI Building, 10 Jalan
28 Ganesa, Bandung 40132, Indonesia
29
30 Corresponding Author: Jimmy A. McGuire
31
32 Email address: mcguirej@berkeley.edu
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49
50 **ABSTRACT**
51
52   We used Massively Parallel High-Throughput Sequencing to obtain genetic data from a 145-
53 year old holotype specimen of the flying lizard, *Draco cristatellus*. Obtaining genetic data from
54 this holotype was necessary to resolve an otherwise intractable taxonomic problem involving the
55 status of this species relative to closely related sympatric *Draco* species that cannot otherwise be
56 distinguished from one another on the basis of museum specimens. Initial analyses suggested
57 that the DNA present in the holotype sample was so degraded as to be unusable for sequencing.
58 However, we used a specialized extraction procedure developed for highly degraded ancient
59 DNA samples and MiSeq shotgun sequencing to obtain just enough low-coverage mitochondrial
60 DNA (547 base pairs) to conclusively resolve the species status of the holotype as well as a
61 second known specimen of this species. The holotype was prepared before the advent of
62 formalin-fixation and therefore was most likely originally fixed with ethanol and never exposed
63 to formalin. Whereas conventional wisdom suggests that formalin-fixed samples should be the
64 most challenging for DNA sequencing, we propose that evaporation during long-term alcohol
65 storage and consequent water-exposure may subject older ethanol-fixed museum specimens to
66 hydrolytic damage. If so, this may pose an even greater challenge for sequencing efforts
67 involving historical samples.
68
69
70 **INTRODUCTION**
71
72 The advent of Massively Parallel High-Throughput Sequencing (HTS) has dramatically altered
73 the manner in which geneticists conduct their research. This is certainly true for molecular
74 phylogeneticists and population geneticists, who now routinely have access to large multilocus
75 genetic datasets for non-model organisms. Because HTS using the Illumina platform involves
76 sequencing of small fragments of DNA, this approach offers the potential to access previously
77 unattainable genome-scale sequence data even for degraded historical samples (e.g., Prüfer et al.,
78 2014; Palkopoulou et al., 2015). Millions of fluid-preserved specimens in museum collections
79 predate the development of allozyme and DNA sequencing technologies, and thus lack specially
80 preserved tissue samples for genetic analysis. Formalin-fixed fluid specimens usually having
81 highly fragmented and cross-linked DNA, are often refractory to sequencing efforts using
82 traditional Sanger sequencing. However, recent studies have shown that it is possible to obtain
83 genomic DNA sequences from some of these fluid-preserved museum specimens. Hykin, Bi &
84 McGuire (2015) demonstrated that low-coverage genomic sequences could be recovered from a
85 30-year old formalin-fixed museum specimen, though they were unsuccessful with a ~100-year
86 old specimen. Ruane and Austin (2016) sequenced Ultra-Conserved Elements (UCEs) from both
87 formalin-fixed (n=11) and ethanol-fixed (n=10) museum specimens, including one sample that
88 was collected between 1878 and 1911. Both had mixed success, with the quantity of DNA
89 recovered in the extraction stage likely playing the largest role in the performance of their
90 sequencing efforts. Notably, the samples that failed in Ruane and Austin's (2016) experiment
91 included subsets of both their formalin- (7 of 16) and alcohol-fixed (4 of 5) samples, indicating
92 that old alcohol-preserved museum specimens are not necessarily less problematic than those
93 initially fixed with formalin. This is surprising given that contemporary tissue samples
94 earmarked for genetic analysis are routinely stored in 95% ethanol.

Preprints

95   Our objective in this study was to address an otherwise intractable problem in flying
96   lizard taxonomy using Illumina HTS and ancient DNA methods for a 145-year old fluid-
97   preserved holotype specimen. The nettlesome taxonomic issue involves a small clade of poorly
98   known flying lizards (genus *Draco*, Agamidae) that, for reasons outlined below, was unlikely to
99   be resolved without obtaining genetic data from the holotype specimen of one of the constituent
100   species, *Draco cristatellus*. Determining species limits within this small clade (the *Draco*
101   *fimbriatus* group) has proven challenging for taxonomists, and we first describe the convoluted
102   taxonomic history of the clade as a justification for our ultimate solution to this question
103   involving HTS. The *D. fimbriatus* group currently includes four recognized species: *D.*
104   *abbreviatus*, *D. cristatellus, D. fimbriatus*, and *D. maculatus*. This taxonomic framework is
105   based on Manthey (2008) and was followed by the widely utilized Reptile Database (Uetz, 2006).
106   For reasons that we will describe in a subsequent paper, we instead utilize an alternative
107   taxonomy that includes *D. cristatellus*, as well as *D. fimbriatus* (= *D. abbreviatus* above), *D.*
108   *hennigi*, *D. punctatus* (= *D. fimbriatus* above), and *D. maculatus*. We note that our taxonomy
109   differs from that of Manthey (2008) primarily as a consequence of having information that
110   indicates that the type locality of *D. fimbriatus* is the Malay Peninsula rather than Java. We
111   further note that our recognition of both *D. cristatellus* and *D. punctatus* is tentative, as a primary
112   objective of this paper is to resolve whether these are in fact distinct species.
113   The *Draco fimbriatus* group is taxonomically challenging. Although *Draco maculatus* is
114   abundant, easily sampled in the field, and easily distinguished from other members of the group
115   based on external phenotype, the remaining members of this clade are only rarely encountered,
116   with relatively few specimens represented in museum collections. These species are canopy
117   specialists (McGuire, 2003), making them more difficult to detect and more challenging to
118   collect than other *Draco* taxa. Furthermore, the species comprising the *D. fimbriatus* group, as
119   well as several other *Draco* clades are primarily distinguished on the basis of differences in
120   coloration of their display structures (dewlap and patagia for most *Draco* taxa, just the dewlap
121   among the relevant members of the *D. fimbriatus* group). For example, two major clades — the
122   'Philippines *volans* group' (McGuire and Alcala, 1999) and the '*Draco lineatus* group'
123   (McGuire et al., 2007) — are each composed of multiple species that are primarily distinguished
124   on the basis of coloration. Because coloration fades in preservative, recognizing species-specific
125   coloration characteristics generally requires experience with the species in the field and/or access
126   to color imagery of the specimen in life. Thus, as museum specimens, the members of these
127   clades are functionally cryptic sympatric species. In summary, for the *D. fimbriatus* group, the
128   paucity of museum specimens, and the rarity with which specimens are observed in the field
129   from throughout their collective ranges by single observers, has greatly impeded taxonomic
130   progress.
131   Within the *Draco fimbriatus* group, a particularly challenging issue relates to the
132   taxonomic standings of *D. cristatellus* Günther 1872 and *D. punctatus* Boulenger 1900. *Draco*
133   *cristatellus* was described based on a single specimen collected by Mr. Alfred Hart Everett in
134   Matang, Sarawak between 1869 and March of 1872 (when Günther's manuscript describing the
135   species was submitted for publication). Although Everett collected the type specimen, the
136   Trustees of the British Museum purchased it from Mr. W. Cutter, thereby making it available to
137   Günther for description (see Günther, 1872). Because Günther presumably did not see the living
138   specimen, he did not evaluate the coloration of the dewlap in life, which is essential for species
139   identification within this group. Günther described the dewlap as 'golden-yellow, with a brown
140   anterior edge', presumably from its preserved state. Subsequently, Boulenger (1900) described *D.*

141   *punctatus* from Bukit Larut on the Malay Peninsula, noting that the dewlap was lemon yellow in
142   coloration. Although he did not attempt to diagnose *D. punctatus* from *D. cristatellus*, Boulenger
143   (1900) was clearly aware of the latter species and explicitly considered his *D. punctatus* holotype
144   to be taxonomically distinct. Indeed, Boulenger (1900) noted that he had examined a second
145   specimen of *D. punctatus* from Sarawak that was also collected by Everett, remarking that he
146   had previously referred that second specimen to *D. cristatellus*. Boulenger (1900) might be the
147   last author to have had a clear idea about the taxonomic distinctiveness of *D. cristatellus* and *D.*
148   *punctatus,* and it is a pity that he did not identify the character differences that he used to render
149   his taxonomic decision. Although de Rooij (1915) recognized *D. fimbriatus*, *D. cristatellus* and
150   *D. punctatus* as distinct species, subsequent authors synonymized one or more members of the
151   group. Hennig (1936) synonymized *D. cristatellus* with *D. fimbriatus*, while continuing to
152   recognize *D. punctatus.* In his monographic *Draco* taxonomic study, Musters (1983) opted to
153   synonymize both *D. cristatellus* and *D. punctatus* with *D. fimbriatus*. In his competing
154   taxonomic treatment, Inger (1983) recognized two species, *D. cristatellus* and *D. fimbriatus,* as
155   valid species, but placed *D. punctatus* in the synonymy of *D. cristatellus*. Inger's (1983)
156   recognition of two species was based in part on Grandison's (1972) report on two sympatric *D.*
157   *fimbriatus* group species with distinct dewlap colorations on Gunung (Mt.) Benom on the Malay
158   Peninsula. Whereas Grandison (1972) identified the two sympatric species as *D. fimbriatus* and
159   *D. punctatus* (without commenting on the status of the Bornean *D. cristatellus*), Inger (1983)
160   instead opted to treat *D. punctatus* as a synonym of *D. cristatellus*. This sensible decision was
161   presumably made on the basis of the similar dewlap colorations of the *D. cristatellus* and *D.*
162   *punctatus* holotypes ('golden-yellow, with a brown anterior edge' vs. 'lemon yellow'). Inger
163   (1983) furthermore attempted to differentiate his conceptions of *D. fimbriatus* and *D. cristatellus*
164   using a statistical analysis of eight linear measurements and scale counts. Although he
165   successfully sorted his sample into two groups on the basis of overlapping but significantly
166   distinct character state differences, his *a priori* placement of *D. punctatus* in the synonymy of *D.*
167   *cristatellus* effectively precluded the possibility that three species — *D. cristatellus, D.*
168   *fimbriatus*, and *D. punctatus* — might all co-occur on the Greater Sunda Shelf (and particularly
169   on Borneo). Here we address this open question taking advantage of two critical developments:
170   (1) the acquisition and analysis of a key specimen (TNHC 56763) obtained by JAM from
171   Santubong, Sarawak, Malaysian Borneo in 1996, and (2) an analysis of the 145-year old *D.*
172   *cristatellus* holotype using ancient DNA extraction methods and HTS on the Illumina platform.
173
174   **MATERIALS AND METHODS**
175
176   **(a) DNA extraction and sequencing from the *Draco cristatellus* holotype**
177   We obtained from the Natural History Museum in London liver tissue from the holotype of
178   *Draco cristatellus* (specimen BMNH 1872.2.19.4). This specimen was originally collected and
179   prepared prior to March 1872, well before the advent of formalin-fixation.
180   At that time, the standard practice for fluid preservation of reptiles, amphibians, and fishes was
181   direct preservation in "pure spirits of wine" (Günther, 1880). Thus, the holotype was most likely
182   initially fixed in 90-100% ethanol (= Günther's "pure spirits of wine") and never exposed to
183   formalin. Nevertheless, we opted to perform our initial DNA extraction using the methodology
184   described in Hykin, Bi & McGuire (2015) for formalin-fixed tissues, with the goal being to
185   perform an exome-capture experiment with this sample. The Hykin, Bi & McGuire (2015)
186   procedure involves a series of initial ethanol washes followed by treatment in a heated alkali

187 buffer solution to break cross-linkages before standard phenol-chloroform extraction. When this
188 extraction returned a very low (potentially zero) yield, we performed a second pair of phenol-
189 chloroform extractions involving phase-lock gel tubes followed by SPRI bead clean-up. This
190 second round of extractions was performed with and without exposure to heated alkali solution.
191 These extractions also failed to return sufficient DNA to move forward with library preparation.
192 Despite minimal DNA yield, we made an attempt to PCR-amplify and sequence a short fragment
193 of the mitochondrial ND2 gene from both DNA extracts. These experiments resulted in the
194 amplification and sequencing of human ND2 in two separate experiments. Both of our low-yield
195 DNA extractions were then sent to MYcroarray Inc. in Ann Arbor, MI where they were
196 subjected to an extra silica purification designed for low-concentration fragment retention, and
197 prepared as libraries. However, the library preparation retrieved only artifact and we did not
198 proceed to targeted enrichment of selected exons or sequencing.
199   At this stage, we engaged with a lab specializing in genetic analysis of ancient DNA
200 samples, with extraction and sequencing performed in a facility specifically designed for work
201 with ancient samples (the Shapiro Lab at UC Santa Cruz). No reptile work had previously been
202 done in this facility and all work followed lab standards for working with historical samples
203 (Fulton, 2011). The DNA extraction protocol was based on Dabney et al. (2013), Tin, Economo
204 & Mikheyev (2014), and D. D. Cotoras et al. (2017). An initial subsample of 40 mg of tissue was
205 subdivided into ~1mm pieces and suspended in lysis buffer. The composition of the 100 mL lysis
206 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,
207 1 mL 2-mercaptoethanol, and 77.8 mL distilled water. The tissue was digested with a total of 1
208 mL of lysis buffer with 1 mg/mL proteinase K, initially incubated overnight at 56°C, and then
209 raised to 72°C for 1 hr. The 72°C incubation step was undertaken in case the sample had been
210 exposed to formalin in order to reverse any potential crosslinks. Silica-based purification
211 followed the centrifugation-based protocol described by Dabney et al. (2013). Briefly, 0.5 mL of
212 3M sodium acetate was added to the lysate and them transferred to a tube with 13 mL of binding
213 buffer. The binding buffer is prepared in a 50 mL tube by first adding 23.88 g of guanidine
214 hydrochloride and then adding water to bring the volume to 30 mL. A key element of this
215 purification protocol is the high salt concentration of this binding buffer, which enhances
216 recovery of short DNA fragments. After complete dissolution of the guanidine hydrochloride, 25
217 μl of Tween-20 and sufficient isopropanol to bring the total volume to 50 mL were added. The
218 mixture of sample, binding buffer, and sodium acetate was transferred into a Zymo extension
219 reservoir attached to a MiniElute spin column. The spin column was then centrifuged for 10
220 minutes at 1,000 rpm, after which the spin column was transferred to a 1.5 mL Eppendorf tube.
221 We performed a dry spin for 1 minute at 13,000 rpm, followed by 2 washes with 750 μl of PE
222 buffer (1 minute spin at 6,000 rpm). To ensure the entire PE buffer was removed, we did a dry
223 spin for 1 minute at maximum speed. We eluted the purified extract in two volumes of 25 μl of
224 TET. Each sample was centrifuged for 30 seconds at 13,200 RPM after 3-5 minutes of
225 incubation. Because the elution displayed pigmentation, 25 μl of the extract was purified on a
226 column filled with cross-linked polyvinylpyrrolidone (PVPP) (Arbeli and Fuentes, 2007). We
227 also produced an extraction control consisting of lysis buffer that was subjected to the same set
228 of procedures.
229   For genomic sequencing, we prepared two barcoded Illumina sequencing libraries (one
230 for the holotype sample and one for the control) using the Meyer and Kircher (2010) protocol,
231 starting with 5 μl of the PVPP purified DNA extraction. The same volume was used for the
232 extraction control. The libraries were sequenced on an Illumina MiSeq machine using 150-cycle

233 v3 chemistry (2x75). Following sequencing, adaptors were removed from reads and sequences
234 were merged using SeqPrep2 (https://github.com/jeizenga/SeqPrep2). Default parameters were
235 used with the exception of the following: -q 20 -L 30 -A AGATCGGAAGAGCACACGTC -B
236 AGATCGGAAGAGCGTCGTGT. FastQC (https://www.bioinformatics.babraham.ac.uk/
237 projects/fastqc/) confirmed that the sequence quality was good, with the normal base quality drop
238 in the final five bases.
239
240 **(b) Sanger sequence data for *Draco fimbriatus* group specimens**
241 Our team has generated a large number of complete ND2 sequences for *Draco* specimens,
242 including for 65 exemplars representing the *D. fimbriatus* group. These sequences were available
243 for comparison with ND2 sequence fragments obtained from the *D. cristatellus* holotype. PCR-
244 amplification was undertaken using the primers Metf1 and ALAr2, with cycle sequencing
245 involving these external primers plus the internal primers Metf5 and ND2r6 (see McGuire and
246 Kiew 2001 for details).
247
248 **(c) Exome-capture and screening of mitochondrial DNA**
249 For another project (McGuire et al. in prep), we generated an exome-capture data set using the
250 MyBaits in-solution capture system for a set of 350 samples spanning all of *Draco*. This sample
251 set included 14 *D. fimbriatus* group samples. The target loci for the exome capture include 1400
252 exons and flanking sequences derived from transcriptome sequences (jointly representing 709
253 loci), which were supplemented with an additional 540 lizard-specific UCE loci. Libraries
254 enriched for our target loci were barcoded and sequenced on an Illumina Hi Seq 4000. Although
255 our experiment was specifically designed to avoid capturing mitochondrial genes, mitochondrial
256 sequences are so abundant in genomic DNA extractions that some mitochondrial molecules
257 inevitably find their way into the off-target by-catch (non-target DNA sequences that are
258 obtained during an exome-capture experiment). We took advantage of this imperfect filter to
259 obtain mitochondrial sequences for comparison with the *D. cristatellus* holotype. For TNHC
260 56763, we used Geneious version 8.1.7 (Kearse et al. 2012) to obtain a mostly complete
261 representation of mitochondrial coding genes by mapping our raw exome capture data (including
262 off-target sequences) to the complete mitochondrial genome of *Acanthasaura armata* available
263 on GenBank (AB266452.1). A preliminary assessment of the identity of the sequences was
264 performed with a BLAST search after collapsing duplicate sequences with fastx_collapser
265 (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html#fastx_collapser_usage). The result
266 of the BLAST search was visualized with the program MEGAN (Huson et al., 2007). Processed
267 reads were mapped with BWA mem (Li and Durbin, 2009) against the reference partial
268 mitochondrial genome of TNHC 56763. Duplicates were removed with samtools rmdup (Li et al.,
269 2009). A total of 17 unique reads mapped against the reference after duplicate removal. The
270 represent a total of 777 bp of the 8114 bp reference. Most of the mapped regions had 1x
271 coverage and portions of four contigs had 2x coverage. The average length of the mapped reads
272 was 53 bp. Finally, for each of the 14 *D. fimbriatus* group samples included in our exome-
273 capture experiment, we used Geneious to map our raw sequencing reads to 10 mitochondrial
274 contigs obtained for the *D. cristatellus* holotype. The raw sequence data is available on the SRA
275 database.
276
277 **(d) Analysis of DNA sequence variation**

278 Our analysis of DNA sequence variation included alignment of homologous DNA sequences and
279 a simple count of nucleotide base substitutions between the *Draco cristatellus* holotype, sample
280 TNHC 56763 from Santubong, Sarawak, and our selection of *D. fimbriatus* and *D. punctatus*
281 samples from the Malay Peninsula, Sumatra, the Mentawai Islands, Java, and Borneo. Specimens
282 examined are listed in Table 1. We also performed a heuristic parsimony analysis to obtain a
283 phylogram for the *D. fimbriatus* group and performed a non-parametric bootstrap analysis with
284 1000 replicates to assess branch support. We did not perform a more rigorous maximum
285 likelihood or Bayesian analysis because our primary objective was to assess uncorrected relative
286 branch lengths. Phylogenetic analyses were performed in PAUP version 4 (Swofford 2002).
287
288 **(e) Data availability and Permits**
289 Sanger sequence data are available on GenBank (will be submitted if accepted) and a matrix for
290 the mitochondrial ND2 gene for the *D. fimbriatus* group is included as supplemental materials.
291 This research was undertaken in accordance with UC Berkeley Animal Use Protocol Number
292 AUP-2014-12-6954. Fieldwork was undertaken with research permits issued by the Economic
293 Planning Unit of Malaysia (UPE:40/200/19 SJ.363) and the Indonesia Institute of Sciences
294 (LIPI: No. 2411/FRP/SM/X?2008 and No. 0115/FRP/SM/VI/2009).
295
296 **RESULTS**
297
298 The MiSeq run generated a total of 1,086,926 paired-end reads after combining the data from 2
299 different libraries (prepared identically) from the same extract. Raw sequences were merged if
300 possible and duplicates collapsed, producing a total of 538,995 reads, which were BLAST
301 searched (Altschul et al. 1990) against the NCBI database. 115,266 reads were successfully
302 assigned, and of these, 47,400 hits corresponded to bacteria, 40,751 were assigned to mammals
303 (of which 29,720 were specifically assigned to human), and 48 were assigned to *Anolis*
304 *carolinensis*. Only three reads were assigned to *Draco*, each of which involved the mitochondrial
305 ND2 sequence posted on GenBank for *Draco cristatellus* sample TNHC 56763 (see below). The
306 paucity of *Draco* hits is not surprising given that there are no *Draco* reference genomes to map
307 to. The extraction control was sequenced producing a total of 104,417 PE reads. After processing
308 (adaptor removal, merged if possible, and duplicate removal) a total of 15,480 reads were
309 assigned by the BLAST search. Bacteria were represented by 6,805 reads, mammals by 4,711
310 reads (of those, 3103 were assigned to human, and reptiles were assigned no reads (one read was
311 assigned to chicken (*Gallus gallus*).
312   Of the three holotype ND2 reads, two were broadly overlapping, and the joint ND2 data
313 obtained from the holotype totaled only 125 bp. To search for additional mitochondrial contigs in
314 the holotype MiSeq data, we first generated a partial mitochondrial genome for TNHC 56763
315 from exome-capture off-target by-catch, which returned 8114 bp of protein-coding gene
316 sequence data. Mapping the holotype data to the TNHC 56763 mitochondrial assembly resulted
317 in the recovery of an additional 10 reads totaling 596 bp of mitochondrial sequence data
318 representing six genes (COI, COXIII, ATPase8, ND4, ND4L, ND5). Thus, a total of 13 contigs
319 were assembled with an average length of 61 bp. Most of the mapped regions had 1x coverage
320 and parts of four contigs had 2x coverage. No reads were recovered when mapping the extraction
321 blank against the same reference. In comparing the holotype sequence data with TNHC 56763
322 across the 721 bp of homologous sequence data, we found that the two samples were only
323 weakly divergent from one another, sharing the same base calls at 710 of 721 positions for a raw

324   sequence divergence of 1.5%. We then mapped the raw reads from the exome captures for the
325   remaining 13 *D. fimbriatus* group samples to the 10 *D. cristatellus* holotype contigs, which
326   returned as few as two and as many as six homologous sequences per sample.
327   Comparison of the mitochondrial data obtained from the *Draco cristatellus* holotype with
328   homologous data obtained for *D. fimbriatus* group samples found that the sample TNHC 56763
329   from Santubong, Sarawak, Malaysian Borneo was much more similar to the holotype than were
330   any other *D. fimbriatus* group samples. When limiting our comparison to the six gene fragments
331   for which we had between six and 13 corresponding sequences for comparison to the holotype,
332   we found that TNHC 56763 was 0.9% divergent from the holotype, whereas all other samples
333   ranged between 11.9% and 15.1% divergent. TNHC 56763 differed from the holotype at just five
334   of 547 base positions, whereas the other samples differed from the holotype at from 36 of 303 bp
335   to 70 of 465 base positions. The ND2 comparisons were most comprehensive because we had
336   access to complete ND2 sequences for 65 *D. fimbriatus* group samples. Whereas TNHC 56763
337   differed from the holotype at three of 183 base positions, all other samples differed by at least 24
338   base positions. Notably, the two *D. punctatus* samples from Sarawak (the type locality for *D.*
339   *cristatellus*) differed from the holotype at 28 and 30 of 183 base positions (15.3% and 16.4%,
340   respectively).
341   A parsimony phylogenetic analysis of the ND2 gene including the *D. cristatellus*
342   holotype strongly supports the monophyly of the holotype together with TNHC 56763 to the
343   exclusion of all other *D. fimbriatus* group samples with 100% bootstrap support (Figure 1).
344   Further, the 0.9% mitochondrial sequence divergence between TNHC 56763 and the holotype is
345   within the scope of ND2 sequence variation that we observe between *D. punctatus* samples from
346   the same locality. This does not consider the possibility that one or more of the five documented
347   base substitutions could be sequencing errors resulting either from damage to the holotype DNA
348   or random errors in our low coverage data, as we did not apply informatics pipelines developed
349   to identify post-mortem damage of ancient DNA to our data (e.g. Mateiu and Rannala 2008,
350   Molak and Ho 2011). Notably, one of the three inferred ND2 substitutions is a first position C-
351   >T change that would result in a proline to serine amino acid replacement, suggesting that this
352   might be the result of post-mortem deamination of the template molecular or sequencing error.
353
354   **DISCUSSION**
355
356   In the present study, we applied HTS to a 145-year old fluid-preserved holotype specimen in an
357   effort to disentangle an otherwise intractable taxonomic question. The problem stems from the
358   fact that one of the species in this group, *Draco cristatellus*, was described using limited color
359   information, and because fluid-preserved specimens representing multiple sympatric *D.*
360   *fimbriatus* group species are often indistinguishable from one another without color information.
361   Indeed, sympatric species in this complex are effectively cryptic once they have been prepared as
362   museum specimens. This combination of circumstances rendered it virtually impossible to
363   resolve the species status of *D. cristatellus* relative to *D. punctatus* and *D. fimbriatus,* two
364   widespread species on the Greater Sunda Shelf. Importantly, a sample (TNHC 56763) collected
365   in 1996 by JAM provides phylogenetic evidence for a third *D. fimbriatus* group species on
366   Borneo, with the natural question being whether this sample is conspecific with the name-
367   bearing holotype specimen of *D. cristatellus* housed in the British Museum of Natural History.
368   Several species composition outcomes were possible, all of which were considered in a
369   hypothesis-testing framework. In Hypothesis 1, *D. cristatellus* and *D. punctatus* are synonyms,

370   together representing a single species distinct from *D. fimbriatus* and TNHC 56763. In
371   Hypothesis 2, *D. cristatellus* and *D. fimbriatus* are synonyms. In Hypothesis 3, *D. cristatellus* is
372   a species distinct from *D. punctatus* and *D. fimbriatus* but conspecific with TNHC 56763.
373   Finally, in Hypothesis 4, *D. cristatellus*, *D. fimbriatus*, *D. punctatus*, and TNHC 56763 all
374   represent distinct species, with TNHC 56763 representing a fourth sympatric species on Borneo.
375   The only way to conclusively test these alternative hypotheses was to obtain informative genetic
376   data from the holotype specimen of *D. cristatellus* for comparison with TNHC 56763, and
377   representative specimens of *D. fimbriatus* and *D. punctatus*.
378   We initially believed that genetic data would easily be retrieved from the *Draco*
379   *cristatellus* holotype. The holotype was prepared before the advent of formalin-fixation, and we
380   consequently had reason to believe that the specimen was originally fixed with ethanol and had
381   never been exposed to formalin. Because tissue samples collected for genetic analysis are
382   routinely stored in ethanol, we were confident that the holotype would still hold high molecular-
383   weight DNA suitable for genomic sequencing. Our hope was to perform exome-capture with this
384   sample and include it in a larger *Draco* phylogenomic data set. However, our initial attempts at
385   extracting DNA from the sample using methods appropriate for historical and formalin-fixed
386   tissues failed, forcing us to adjust both our approach and our expectations. Fortunately, our
387   alternative hypotheses proved testable without comprehensive genomic data from the holotype.
388   Indeed, analysis of the initial 125 bp of mitochondrial ND2 data identified when the holotype
389   sequence data was subjected to GenBank BLAST allowed us to reject hypotheses 1, 2, and 4 in
390   favor of hypothesis 3. The additional 422 bp of mitochondrial data obtained via mapping of
391   holotype contigs to the reconstructed mitochondrial genome for TNHC 56763 simply provided
392   additional confirmation that *D. cristatellus* and *D. punctatus* are each valid species, and that our
393   specimen TNHC 56763 from Santubong, Sarawak is indeed a true *D. cristatellus* exemplar. This
394   finding was a best-case scenario because all future specimens for which tissue samples are
395   obtained can now be compared with known *D. cristatellus*, *D. punctatus*, and *D. fimbriatus*
396   samples for genetic identification.
397   What lessons can be learned from our attempt to obtain genetic data from the *Draco*
398   *cristatellus* holotype? First, even when initial attempts at extraction and quantification of DNA
399   suggest that none is present, small numbers of DNA molecules may survive in the sample. For
400   questions of simple species identification involving old and highly degraded samples, it may
401   only be necessary to obtain limited data — even a few hundred base pairs of mitochondrial data
402   may be sufficient to address the question. Our study shows that this is indeed possible even when
403   initial assessments suggest that DNA in a tissue sample has been highly degraded. Obtaining
404   data in these instances will likely require highly specialized extraction procedures such as the
405   silica-column based extraction methodology utilized here, followed by short-fragment
406   sequencing. Finally, we believe that the difficulty we confronted with this ethanol-fixed sample
407   — which is consistent with the problems experienced by Ruane and Austin (2016) with their
408   presumed alcohol-fixed samples (4 of 5 of which failed to sequence) — suggests that hydrolyzed
409   ethanol-fixed tissues might be more problematic than formalin-fixed samples for genomic
410   sequencing efforts (see Handt et al. 1994 for a description of DNA hydrolysis). This has
411   important implications for curatorial practices. Not only is potential hydrolytic damage cause for
412   concern with whole fluid specimens stored in ethanol, but it could be particularly problematic for
413   tissue samples stored in ethanol that are not maintained in sub-zero degree conditions.
414   Evaporation of ethanol from tissue vials might render even modern tissue samples virtually
415   unusable for genetic analysis.

416
417 **CONCLUSIONS**
418
419 The development of HTS has revolutionized biological research by making genome-scale data
420 readily available at a reasonable cost, even for non-model organisms.
421 Systematists have fully embraced these advances in data acquisition for freshly sampled
422 specimens, but are just beginning to harness HTS for the millions of fluid-preserved historical
423 samples housed in natural history collections around the world. As we have shown here,
424 acquiring genetic data from old museum specimens will sometimes present special challenges,
425 but the information that can be gleaned from such specimens may be the only way to
426 conclusively resolve previously intractable evolutionary and taxonomic questions.
427
428 **ACKNOWLEDGEMENTS**
429
440
441
442 **REFERENCES**
443
444 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, & D. J. Lipman. 1990. Basic local alignment
445   search tool. *Journal of Molecular Biology* 215:403-410.
446 Arbeli, Z. & C. L. Fuentes, 2007. Improved purification and PCR amplification of DNA from
447   environmental samples. *FEMS Microbiology Letters* 272:269–275.
448 Boulenger, G. A. 1900. Descriptions of new batrachians and reptiles from the Larut Hills, Perak.
449   *Annals and Magazine of Natural History* Series 7, Volume 6:186-193.
450 Cotoras, D. D., G. G. R. Murray, J. Kapp, R. G. Gillespie, C. Griswold, W. B. Simison, R. E.
451   Green, & B. Shapiro. 2017. Ancient DNA resolves the history of *Tetragnatha* (Araneae,
452   Tetragnathidae) spiders on Rapa Nui. *Genes* 8:403.
453 Dabney, J., M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N.
454   Garcia, S. Paabo, J.-L. Arsuaga, & M. Meyer. 2013. Complete mitochondrial genome
455   sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments.
456   *Proceedings of the National Academy of Sciences*, 110:15758–15763.
457 Fulton, T. L. 2011. Setting Up an Ancient DNA Laboratory. *Ancient DNA Methods and*
458   *Protocols* (ed. by M. Shapiro, B.; Hofreiter), pp. 1–11. Humana Press, New York.
459 Grandison, A. G. C. 1972. The Gunong Benom Expedition 1967. 5. Reptiles and amphibians of
460   Gunong Benom with a description of a new species of *Macrocalamus*. *Bulletin of the British*
461   *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 <u>*Draco*</u>).
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: <u>19505943</u>]
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 <u>*Draco*</u> 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.

507   Palkopoulou, E., S. Mallick, P. Skoglund, J. Enk, N. Rohland, H. Li, A. Omrak, S. Vartanyan, H.
508   Poinar, A. Götherström, D. Reich, & L. Dalén. 2015. Complete genomes reveal signatures of
509   demographic and genetic declines in the woolly mammoth. *Current Biology* 25:1395-1400.
510   Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud,
511   P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu1, M. Kircher, M.
512   Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P.
513   Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson,
514   H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L.
515   V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin,
516   D. Reich, Janet Kelso, & S. Pääbo. 2014. The complete genome sequence of a Neandertal
517   from the Altai Mountains. *Nature* 505(7481):43-49.
518   Ruane, S., & C. C. Austin. 2016. Phylogenomics using formalin-fixed and 100+ year-old
519   intractable natural history specimens. *Molecular Ecology Resources* doi:10.1111/1755-
520   0998.12655.
521   Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
522   Version 4. Sinauer Associates, Sunderland, Massachussetts.
523   Tin M. M.-Y., E. P. Economo, & A. S. Mikheyev. 2014. Sequencing Degraded DNA from Non-
524   Destructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun
525   Phylogenetics. *PLoS ONE* 9(5): e96793. doi:10.1371/journal.pone.0096793.
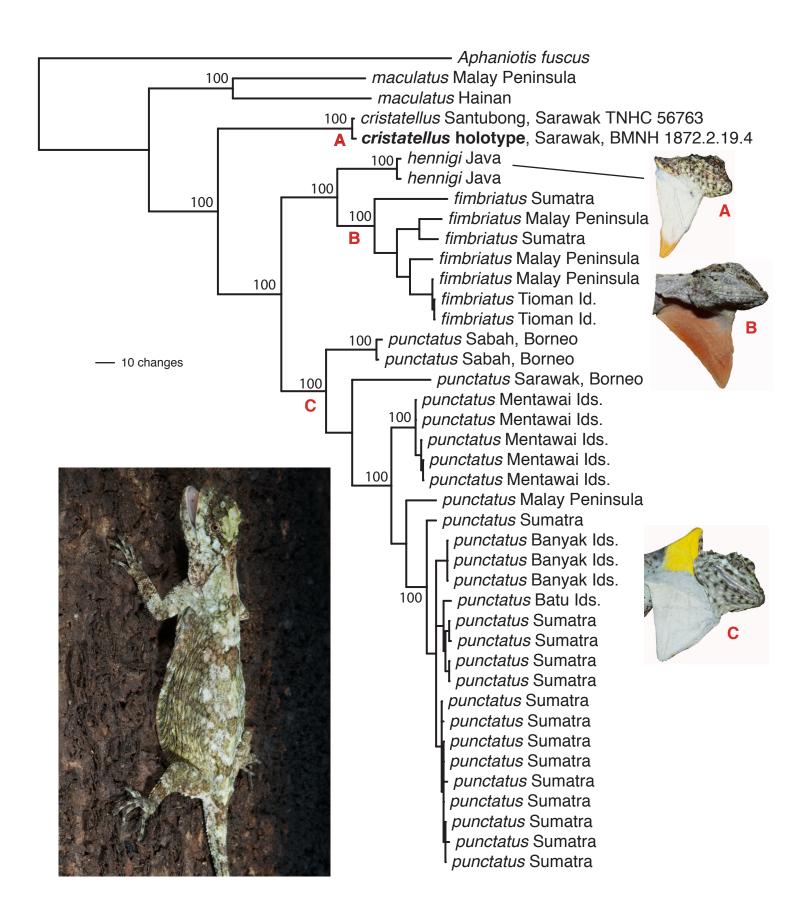526   Uetz, P. 2006. The EMBL Reptile Database CD-ROM edition, March 2006.

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*.

Aphaniotis fuscus

100 *maculatus* Malay Peninsula

maculatus Hainan

100 *cristatellus* Santubong, Sarawak TNHC 56763

A *cristatellus* **holotype**, Sarawak, BMNH 1872.2.19.4

100 *hennigi* Java

hennigi Java

100

100 *fimbriatus* Sumatra

100 *fimbriatus* Malay Peninsula

B *fimbriatus* Sumatra

fimbriatus Malay Peninsula

fimbriatus Malay Peninsula

fimbriatus Tioman Id.

fimbriatus Tioman Id.

100 *punctatus* Sabah, Borneo

punctatus Sabah, Borneo

100 *punctatus* Sarawak, Borneo

100 *punctatus* Mentawai Ids.

punctatus Mentawai Ids.

100 *punctatus* Mentawai Ids.

punctatus Mentawai Ids.

100 *punctatus* Mentawai Ids.

punctatus Malay Peninsula

punctatus Sumatra

punctatus Banyak Ids.

punctatus Banyak Ids.

punctatus Banyak Ids.

punctatus Batu Ids.

100 *punctatus* Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

punctatus Sumatra

— 10 changes

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	0.9%
punctatus	414/2908	14.2%
fimbriatus	200/1566	12.8%
hennigi	111/810	13.7%

2
3
4
5
6
7
8