Ultraconserved elements resolve the phylogeny of a globally distributed genus, *Butorides* (Aves: Ardeidae)

A Thesis submitted to the faculty of San Francisco State University in partial fulfillment of the requirements for the Degree

Master of Science

In

Biology: Ecology, Evolution, and Conservation Biology

by

Ezra Zachary Mndales

San Francisco, California

May 2023
I certify that I have read Ultraconserved elements resolve the phylogeny of a globally distributed genus, *Butorides* (Aves: Ardeidae) by Ezra Zachary Mendales, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Science in Biology: Ecology, Evolution, and Conservation Biology at San Francisco State University.

__________________________
Jaime Chaves, Ph.D.
Assistant Professor,
Thesis Committee Chair

__________________________
John P. Dumbacher, Ph.D.
Professor

__________________________
John McCormack, Ph.D.
Professor, Occidental College
Abstract

Globally distributed species give us insight into biogeographic patterns of diversification. *Butorides* herons (Aves: Ardeidae) include the globally distributed Striated Heron (*B. striata*) as well as the North and Central America Green Heron (*B. virescens*). These two species hold many similarities across their ecology and morphology. One useful character for identification is the rufous neck of the Green Heron compared to the slate colored neck of the Striated Heron. However, across the globe within the species Striated Heron, their plumage also shows remarkable variation. Additionally, one taxon (*B. s. sundevalli*), endemic to the Galápagos archipelago, continues to be inconsistently treated as subspecies of *B. striata* and as a full species. Despite their broad geographical range and widely varying morphology, there has been no genetic investigation into the biogeographic structure of the clade, how morphology relates to this structure, and the taxonomic placement of *B. s. sundevalli*. Here, we present a phylogeny of the genus based on thousands of ultraconserved elements (UCEs) generated from target capture during library construction and massively parallel sequencing. We additionally investigated morphological data collected from hundreds of live and preserved specimens across the globe. Our genetic and morphological results suggest a biogeographic split between Old and New World *Butorides* herons as well as within the Americas. Finally, we also find evidence counter to the presumption that *B. s. sundevalli* shares a common ancestor with the Striated Heron, but rather it is shared with the Green Heron. This evidence, along with its unique ecology, adds to our understanding of the Lava Heron’s evolutionary history and future, and provides more justification to elevate its species rank.
Acknowledgements

This research would not be possible without the collaboration of researchers and institutions across the United States. Tissue was approved for destructive sampling from the Chicago Field Museum, American Museum of Natural History (AMNH), the Smithsonian Institute, and California Academy of Sciences. Funding was provided by the Chaves-Dumbacher Lab, San Francisco State University, the AMNH Chapman Award Committee, and the AMNH Ornithology Collections Study Grant.

Parque Nacional Galápagos awarded us incredible collaboration during our field work on the archipelago, supporting us throughout our time there in early 2022. Their park rangers ensured the safe capture and release of this species, in particular to Jason Castañeda for his field skills and support. Thank you to Jim Henderson at the California Academy of Sciences for his data pipeline in capturing usable mitochondrial data post-sequencing. Additionally, special thanks to Pooneh Kalhori for their hands-on support in my field, lab, and bioinformatic work.
List of Tables ........................................................................................................................................ vii

List of Figures ..................................................................................................................................... viii

Introduction ....................................................................................................................................... 1

Methods ............................................................................................................................................ 4
  2.1 Taxon sampling 4
  2.2 DNA extraction, UCE enrichment, and sequencing 4
  2.3 Genetic dataset assembly 5
  2.4 Phylogenetic analysis 6
  2.5 Morphometric analysis 7

Results .............................................................................................................................................. 8
  3.1 Phylogenetic results 8
  3.2 Morphometric results 9

Discussion ......................................................................................................................................... 11
  4.1 Butorides phylogeographic structure 11
  4.2 Butorides morphological structure 11
  4.3 Taxonomic placement of the Lava Heron 13

References ....................................................................................................................................... 21
List of Tables

Table 1. Genetic sample data

Error! Bookmark not defined.
List of Figures

Figure 1. Biogeographic distribution and sample distribution map ...................................... 17
Figure 2. Morphometric results ................................................................................................. 18
Figure 3. Phylogeny of Butorides from UCE data ................................................................... 19
Figure 4. Phylogeny of Butorides from mitochondrial data ................................................... 20
1. Introduction

Only a handful of land-based bird species are naturally distributed throughout the globe. The metapopulations of these species can give us a unique perspective on speciation as well how phenotype can relate to genotype. Though neighboring modern populations are often the most closely related ones, cases such as the Galápagos Finches, Galápagos Mockingbird, American Flamingo, as well as non-avian taxa indicate that this is not a rule (Thornton, 1971; Grehan, 2001; Burns et al., 2002; Arbogast et al., 2006; Funk and Burns, 2018).

The Striated Heron (*Butorides striata*) is a land-based species that is categorized as cosmopolitan, yet has not been thoroughly investigated to understand questions that come with a global distribution. Their wide morphological variation within the species poses the question of how this group should be defined. Throughout the tropics and subtropics of the world, they are found on large expanses of land (eg. South America, Africa) as well as in restricted distributions across large islands (eg. Sundas Archipelagos, Madagascar), and small, isolated archipelagos as well (eg. Chagos Islands, French Polynesia). They occupy freshwater rivers, oceanic coasts, mangrove forests, mountain foothills, as well as desert regions (Kushlan and Hancock, 2005). Their plumage ranges from deep green, to sandy brown, to dark gray. This variation in phenotype led to 29-33 uniquely identified subspecies to be joined within *B. striata* depending on the checklist used (Clements et al., 2022; Gill et al., 2023).

In Venezuela, *B. striata* overlaps with the southernmost range of their sister species, the Green Heron (*B. virescens*). The territory of *B. virescens* moves northward throughout most of the US and Caribbean. In addition to their geographic niche, their ecological niche overlaps strongly; they occupy a similar territory size, lay a similar clutch size, are both ambush predators
on invertebrates and small vertebrates, and are mostly similar in their morphology (Kushlan and Hancock, 2005; Kushlan, 2009). The defining physical difference between the two species is the chestnut neck and breast of the Green Heron, compared to the often slate colored neck and breast of the Striated Heron. However, we see similar plumage differences between neighboring populations in other parts of the world. So, for a species that mostly does not migrate, we ask why, for example, a population in Brazil is labeled as more closely related to a population in India, than those in Venezuela.

A third group found only on the Galápagos Islands, the Lava Heron (B. s. sundevalli), is currently treated as subspecies to the Striated Heron (Payne, 1974; Martínez-Vilata and Motis, 1992; Kushlan and Hancock, 2005), but has also been argued as a separate species sister (Peters, 1931; Harris, 1982). The Lava heron is diagnosed by its dark and uniform plumage across its body, in contrast to other Butorides populations that have the defined cap, throat, and breast. This unique plumage has been hypothesized to be adaptive for their unique ecology on the lava strewn coastlines of the Galápagos archipelago (Snow, 1975). Still, their coloration is not monolithic, Payne (1974) published findings on all possible intermediate plumage colors. This group has been hypothesized to be most closely related to the Striated Heron mostly because this B. striata phenotype is also found on the archipelago in small densities. Without further investigation, we cannot know if this phenotype is a recent vagrant population, a separate low density resident, or whether it is an additional state in the currently established cline of phenotypes within the Lava Heron.

This subspecies represents the state of the Butorides complex as a whole; do the wide ranging phenotypes represent shallow morphological differences, or are we seeing evidence for
speciation? Most importantly, what are the actual relationships among all these disparate populations? During a 2002 classification committee on South American taxa, it was proposed to split *B. striata* into two species, *B. striata* and *B. sundevalli*. Ultimately, this proposal was rejected for lack of data, while also questioning the relationship among Old and New World Striated populations (Remson, 2002).

In this study, we use a genome-wide analysis as a first for this genus as well as a comparative morphological analysis of the different global populations. These data will be used to infer their phylogenetic relationships. We enrich our genetic data for ultraconserved elements (UCEs), regions found throughout the genome that are highly conserved across large clades such as tetrapods (Stephen et al., 2008; Faircloth, 2012, 2013). UCEs gather many phylogenetically informative regions and can be sequenced at high coverage without the cost of whole genome sequencing (Crawford et al, 2012; Raposo do Amaral et al., 2015; Branstetter et al., 2017; Chen et al., 2021; Brady et al., 2022).
2. Methods

2.1 Taxon sampling

The genetic dataset is made of 33 specimens (distribution of samples shown in Figure 1). Nine non-lethal blood samples were collected in the field by the researchers in this study. Notably, genetic sample S07 was taken from a Striated Heron morph on San Cristóbal, Galápagos. This sample is critical to understanding the genetic relationship between the two phenotypes. For the remaining samples, we requested all available tissue from several natural history museums across the US for destructive sampling. We included one sample of *Nycticorax nycticorax* (Black-Crowned Night Heron) which was added as an outgroup.

The morphometric dataset comprises 308 *Butorides* specimens from 28 recognized subspecies from the IOC World Bird List v.31.1 (Gill et al., 2023), as well as from one undefined subspecies population in Vietnam (distribution of samples shown in Figure 1). Five described subspecies populations are unrepresented mostly pertaining to remote island chains in the Indian Ocean. The widespread geographic distribution of both the genetic and morphometric samples provide a strong dataset to answer what the taxonomic structure of this genus is.

2.2 DNA extraction, UCE enrichment, and sequencing

DNA from fresh blood and tissue was extracted using Qiagen’s DNEasy kit and protocol. Extracted DNA was then sheared to a maximum size of 550bp using a Covaris sonicator and protocol (75 W peak incident power, 10% duty factor, 200 cycles per burst for 40 seconds). Samples were then cleaned with QuantaBio sparQ Magbeads. Following this, Kapa Biosystems HyperPrep kits were used for end-repair, adapter ligation, and initial amplification using 6-8
cycles of PCR depending on sample DNA concentration. New England Bioscience’s NEBNext®
Multiplex Oligos for Illumina® were used as primers. The samples were then hybridized to baits
from Arbor Bioscience MyBaits Tetrapods kit v5.2 to target 5060 UCE loci and then amplified
for another 10-12 cycles (Faircloth et al., 2012). Libraries were then pooled and amplified once
more to increase concentration for 3 more PCR cycles. The resulting pools were sent to
Novogene facilities for sequencing using NovaSeq PE150.

2.3 Genetic dataset assembly

Raw reads were downloaded and processed with FASTP to filter, trim, and remove
duplicates and adapters. The resulting files were assembled using SPAdes (Nurk et al., 2013).
The assembled contigs were then mapped to the UCE probes through the PHYLUCE pipeline
(Faircloth, 2016). The extracted UCE loci were aligned and edge trimmed using MAFFT (Katoh
and Standley, 2013). This removed UCE loci from the alignment if they were associated with
less than three taxa. 4,438 loci were available for analysis after this step. We chose a 75% data
matrix for 33 taxa as this restricted UCE loci to those where at least 24 taxa were represented per
locus, and still maintained a large dataset of 3,463 loci. The alignment length was 2.17 Mb with
a mean locus length of 595 bp and range of 290-1569 bp. This dataset used an average of 28
taxon included for each locus.

We built an additional single nucleotide polymorphism (SNP) dataset using a PHYLUCE
pipeline extension created by Andermann et al. (2019). This extension includes a python script
that extracts SNPs from the final dataset created from the original UCE workflow and creates an
MSA FASTA file. Not all samples could be included in the SNP dataset, as missing data were
prevalent where SNPs occurred. Our final dataset consisted of samples with 45% or less missing data (n=25, 517 SNPs per sample).

Despite using sequencing data from genetic material enriched for UCEs, we assembled the mitochondrial genome from substantial mitochondrial byproducts that were then recorded in the sequencing process. Genetic materials sourced from blood that were enriched for UCEs did not contain a substantial amount of mitogenomic information. Several of these taxa had to be removed entirely. However, the genome of S18 was sequenced separately and the mitogenome was able to be retrieved from this sample. The resulting dataset contained a total of 28/33 samples included in the mtDNA dataset. Contigs were assembled from the raw reads using SPAdes and mapped to a reference genome (NCBI Reference Sequence: NC_025922.1) using BWA-MEM. An additional outgroup taxon was added to improve rooting in phylogenetic analysis (NCBI Reference Sequence: LC541454.1). After sorting and indexing the output BAM files, they were concatenated and aligned using MAFFT into a multiple sequence alignment (MSA). The final MSA contained 19,096bp for analysis.

2.4 Phylogenetic analysis

UCE and mitogenomic data were run though both maximum likelihood (ML) and Bayesian analysis. The ML trees were made using IQ-TREE by generating 10,000 samples for ultrafast bootstrap (Minh et al., 2020). The best fit models were determined using ModelFinder through Bayesian Information Criterion (Kalyaanamoorthy et al., 2017). We used the program MrBayes to conduct a Bayesian analysis of the UCE data and mitogenomic data (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The analysis set four chains to run for 3.5-4 million generations and sampled the chains every 500 generations with a 25% burn-in.
We used SNAPP for our Bayesian analysis of SNP data (Bryant et al., 2012). SNAPP runs as a package add-on to BEAST software (Bouckaert et al., 2019). The XML file that runs the SNAPP template was created using BEAUti. The chain length in the analysis was 5 million with samples being taken every 1000 produced. We checked the log files for quality assurance using Tracer. The resulting trees were analyzed using TreeAnnotator where the first 20% of trees were discarded. We were able to view the consensus tree with FigTree v1.4.4.

2.5 Morphometric analysis

To answer how morphology plays into the taxonomic structure of *Butorides*, we started our morphometric analysis using each described subpopulation as a discrete unit. This allowed us to test whether any group’s measured trait deviated from the global average of that trait. Measurements were recorded from preserved (n=300) and live specimens (n=8). Culmen length, width, and height, taken from the nares as defined by Borras et al. (2000) were measured. Additionally, we measured tarsus length, and wing length (unflattened). Where possible, sex and life stages were recorded as well as collection location.

Once we had phylogenetic results, we could set biogeographic boundaries based on these when analyzing morphological characters. This showed whether morphometric structure aligned with the genetic structure. Further, a discriminant function analysis was used to determine which traits held the greatest weight for species delimitation.
3. Results

3.1 Phylogenetic results

UCE data reveal strong evidence for biogeographic isolation and potentially speciation within *Butorides*. All major clade nodes have 100% bootstrap support in the Bayesian consensus tree while the RAxML tree agrees on the major topology (Figure 3). The first split within the genus occurs between Old and New World Striated Herons. Following this there is evidence for splits between the African group (*Af*) and Australasian Group (*AA*) within the Old World Group. Bayesian results confidently place North African specimens with *AA*, however the ML results do not corroborate this. According to the Bayesian results, the Madagascan sample (434743) is outgroup to the mainland African samples (455344, 8107, 642965) despite the mainland having larger geographic disparity from each other compared to the large island. This tells us that the oceanic distance between Madagascar and the mainland presents a stronger genetic barrier than the rest of the continent. Other than this, results at fine scale mostly corroborate with geographic distance. For instance, sample 12587 fell within the overlap range of two populations in *AA* and is sister to those groups.

Within the New World herons, there is evidence counter to common hypotheses regarding the placement of the Galápagos (*G*) Lava Heron, *B. s. sundevalli*. Our results show they are most closely related to the North/Central American (*NCA*) *B. virescens*, rather than the South American (*SA*) *B. striata*. To confirm this topology, a likelihood ratio test was used in IQ-TREE by constraining different topologies. The organization of the Lava Heron as sister to *SA* was rejected when *NCA* was outgroup.
The mitochondrial data reveal conflicting results from the nuclear dataset. Figure 4 shows a polytomy in the consensus Bayesian tree between AA, Af, and SA/G clades with the NCA clade as an outgroup with high confidence. Additionally, rather than G clade as sister to NCA, SA branches within G. This would be more consistent with prior hypotheses of the placement of the Lava Heron, however the phylogeny shows that G is more basal to the mainland SA clade. North Africa and Sub-Saharan Africa group confidently together in the mitochondrial phylogeny rather than North Africa with AA as in the UCE phylogeny, though now the Asian B. s. javanica groups with the rest of Af.

3.2 Morphometric results

Sex was not a determining factor for the physical size of any population. Generally, few populations deviate from the global average of measured specimens, although there are some notable results. These include the Lava Heron (B. s. sundevalli) that deviates from the global culmen width and height (Figure 2B). This has been noted previously due to their increased consumption of hard-shelled crustaceans compared to other Butorides spp. (Kushlan, 2009). Two populations have mean wing sizes larger than the standard deviation of the global average; B. s. amurensis, and B. v. anthonyi (Figure 2C). These two are also the only migratory populations within Butorides. There are three populations that have smaller wing sizes than average; B. s. crawfordi, B. s. degens, B. v. bahamensis. These three are island bound populations.

Five biogeographic groups were used based on the results of the phylogenetic analyses: SA, NCA, G, Af, and AA. A discriminant function analysis of these biogeographic groups had similar results in four of the five groups. The analysis was mostly unable to accurately predict
which biogeographic group a specimen belonged to except for $G$, which was accurately predicted among 80.0% of the samples. $AA$ was also accurately predicted 91.3%, however this diagnosis is highly sensitive as the $SA$, $NCA$, and $AF$ groups also were inaccurately predicted as $AA$ more than 52% of the time. When this was graphed as a canonical discriminant function, the group centroids revealed structure among the groups. Figure 2D shows grouping on the Y-axis (Function 2) of New World groups ($NCA$ and $SA$) and Old World ($Af$ and $AA$). The dominant character for Function 2 was tarsus length. Along the X-axis (Function 1), there is a clear separation of Galápagos specimens and the rest of the world. This corroborates the results of Figure 2B.
4. Discussion

4.1 Butorides phylogeographic structure

Butorides herons require a taxonomic reorganization. According to the genetic data there is evidence to split Old and New World Butorides striata herons. This point tells us that if Old World Striated Herons are kept conspecific with New World Striated Herons, that the America Green heron should be conspecific as well given that they share a common ancestor with South American Striated Herons.

Within the Old World Herons there is an argument to split the Sub-Saharan African and Australasian bioregions, though there should be more investigation before this is determined. The placement of North African populations cannot be resolved from this study between Sub-Saharan and South Asian populations. Further, we cannot make conclusions on the topology of the Butorides terminal branches within this sub-clade. Future studies will need much greater genetic coverage to represent the potential diversity within Australasia. The region also has highly variable plumage-based clines that deserve investigation.

4.2 Butorides morphological structure

The morphological analysis results show certain traits can play a stronger role to indicate species structure than others for Butorides herons. According to the canonical discriminant function, we see that tarsus length can discriminate between Old and New World Herons. From the same results we see that culmen height and width is a strong indicator for Lava Heron discrimination from all other herons. This is also apparent in the analyses of the Butorides population comparisons against the global average (Figure 2A). This is likely due to the unique feeding ecology of the Lava Heron compared to the rest of the world. Rather than feeding on
majority small fish and invertebrates, the Lava Heron forages on the coastlines of the archipelago where they consume much higher proportions of thick shelled crustaceans (Kushlan, 2009), which is considered adaptive for capture and handling of hard-shelled prey (Kushlan, 1978; Kushlan and Hancock, 2005).

Though wing-size is not an indicator that aligns with genetic structure, our results do show some significance that align with certain ecotypes, which warrants further genetic investigation. The wingsize of *B. v. anthonyi* and *B. s. amurensis* are the only two that significantly deviate above the global average and they are the only populations that are recorded to migrate regularly. Additionally, there is an apparent island miniaturization effect of wing size on remote island inhabitants, *B. s. degens* and *B. s. crawfordi*. The average wing size of *B. s. crawfordi* (155mm) is 13% smaller on average than the global average (176mm) - well outside the lower standard deviation of the global average, which falls at 167mm.

Plumage, however, does not appear to coincide with phylogenetic signals. As denoted by Payne (1974), the Galápagos Lava Heron dark plumage is not a monolith. An important question concerning the Lava Heron species delimitation was the presence of the mainland Striated morph on the islands. This study sequenced the DNA of a Striated Heron morph (S07) and it groups confidently with the Lava Herons. This indicates that all Galápagos *Butorides* herons are of one polymorphic species. Further, the presence of a dark morphotype in other parts of the world tells us that this could be a common mutation that has become nearly fixed in Galápagos. Both Green and Striated Herons in other parts of the world have melanistic individuals reported. Melanism in the Green Heron was first reported in Honduras (Drucker et al., 2018). *Butorides striata brevipes* has been documented in Djibouti in a melanistic form sympatric with the common Striated
Heron morph (Hering, 2014). We also know that melanism can become fixed in populations with limited gene flow in avian groups such as the Chestnut-bellied Monarch, Vermilion Flycatcher, and Tropical Peewee (Campagna et al., 2022; Smith, 2016).

4.3 Taxonomic placement of the Lava Heron

UCEs resolve a novel phylogeny for the Lava Heron. Despite prior assumptions on the topology of New World *Butorides* herons, Lava and Green Heron are sister to each other, while the Striated Heron sits as the outgroup. The purported presence of the Striated Heron on the Galápagos archipelago is apparently inaccurate according to our results, though a confirmatory analysis of this polymorphism should be done in the future. If true, this would imply that the Lava Heron is strongly isolated from other *Butorides* genotypes.

However, the mtDNA shows different results from the nuclear data. Mito-nuclear discordance often appears due to the haploid nature of mitochondrial cells. Every four copies of nuclear DNA that are passed to the next generation, only one copy of mtDNA is passed (Ballard and Whitlock, 2004). Because of this, mito-nuclear discordance can be caused by a historical or recent introgression event and often it is difficult to distinguish which event may be the case (Nielsen and Wakeley, 2001).

The position of Striated Heron as the ingroup within the topology of the mtDNA consensus tree implies Striated Heron is derived from the Lava Heron. However, a better way to interpret this is that the mitochondrial genome of the Striated Heron ancestor coalesces with the Lava Heron ancestor prior to the Lava Heron’s isolation. Toews and Brelsford (2012) describe several instances of this type of discordance without isolation and secondary contact. The result of this is likely adaptive resulting in mitochondrial capture from selected traits. Due to the lower
effective population size of mitochondrial genomes than nuclear, fixation of new genes can occur much more quickly, and even more so for small populations (Hudson and Turelli, 2003). Alternatively, a recent hybridization event from secondary contact after isolation could also lead to a similar outcome, particularly if there is a sex-based asymmetrical migration or vagrancy (Funk and Omland, 2003). As previously stated, determining which one can prove difficult.

Based on these genetic and morphometric results, there is strong evidence to elevate the Galápagos Lava Heron. As it currently sits, the Lava Heron is subspecies to the Striated Heron, which according to our phylogeny, is not its closest living relative, but rather the Green Heron. Additionally, this study shows that there does not appear to be current potential for introgression of the Striated Heron, as the morphotype that was presumed to be from the mainland, is actually a polymorphism of the Lava Heron.

Given their unique evolutionary history, morphological and ecological uniqueness, and continued strong isolation that promotes a unique evolutionary future, we feel this justifies the split of *B. striata* (South American Striated Herons) from the North/Central American Green Heron (*B. virescens*) and the Galápagos Lava Heron. This would add another endemic and protected species to the Ecuadorian Galápagos archipelago - an ecologically important and globally protected site.

The apparent mito-nuclear discordance shows that there is a unique and interesting history to the emergence of the Lava Heron and deserves more research and insight to reveal their origins. This study can only speculate how the Lava Heron is sister to a species that is further geographically than another *Butorides* species. It is possible that rare migratory behavior within this genus could play a role. Our morphological analysis confirms that the two migratory
populations are those with the largest wingspan. *B. v. anthonyi* migrates yearly from the Western US into Central America for breeding. This reported movement is much larger than the populations of Striated Heron in South America and could lead to vagrancy into novel habitats as was discovered with other endemic Galápagos avian taxa eg. Galápagos Finches and Mockingbird (Funk and Burns, 2018).
Table 1. Genetic sample data

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Institution</th>
<th>Identifier</th>
<th>Source</th>
<th>Location</th>
<th>Cleaned Reads</th>
<th>No. UCE Contigs</th>
<th>Mean config length</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. s. atricapilla</td>
<td>FMNH</td>
<td>455344</td>
<td>Tissue</td>
<td>Nikolokota, Malawi</td>
<td>1550001</td>
<td>4152</td>
<td>809.16</td>
</tr>
<tr>
<td>B. s. atricapilla</td>
<td>AMNH</td>
<td>DOT-8107</td>
<td>Tissue</td>
<td>Central Africa Republic</td>
<td>1560231</td>
<td>3981</td>
<td>722.72</td>
</tr>
<tr>
<td>B. s. atricapilla</td>
<td>USNM</td>
<td>642965</td>
<td>Tissue</td>
<td>Estuaires Province, Gabon</td>
<td>2173780</td>
<td>3229</td>
<td>521.61</td>
</tr>
<tr>
<td>B. s. brevipes</td>
<td>USNM</td>
<td>647929</td>
<td>Tissue</td>
<td>Djibouti Region, Djibouti</td>
<td>2262255</td>
<td>2776</td>
<td>468.6</td>
</tr>
<tr>
<td>B. s. brevipes</td>
<td>USNM</td>
<td>647949</td>
<td>Tissue</td>
<td>Djibouti Region, Djibouti</td>
<td>3354899</td>
<td>3120</td>
<td>522.5</td>
</tr>
<tr>
<td>B. s. carcinophila</td>
<td>FMNH</td>
<td>352219</td>
<td>Tissue</td>
<td>Sibuyan, Philippines</td>
<td>4302280</td>
<td>4314</td>
<td>908.33</td>
</tr>
<tr>
<td>B. s. javanica</td>
<td>AMNH</td>
<td>DOT-9681</td>
<td>Tissue</td>
<td>Singapore</td>
<td>3353160</td>
<td>2543</td>
<td>433.28</td>
</tr>
<tr>
<td>B. s. macrorhynca</td>
<td>AMNH</td>
<td>DOT-2419</td>
<td>Tissue</td>
<td>New South Wales, Australia</td>
<td>2831289</td>
<td>3099</td>
<td>503.48</td>
</tr>
<tr>
<td>B. s. rutenburgi</td>
<td>FMNH</td>
<td>434743</td>
<td>Tissue</td>
<td>Tamatave, Madagascar</td>
<td>3867535</td>
<td>3015</td>
<td>517.08</td>
</tr>
<tr>
<td>B. s. solomonensis</td>
<td>AMNH</td>
<td>DOT6615</td>
<td>Tissue</td>
<td>Solomon Islands</td>
<td>3357379</td>
<td>3500</td>
<td>538.18</td>
</tr>
<tr>
<td><em>B. s. spp</em></td>
<td>AMNH</td>
<td>DOT-12587</td>
<td>Tissue</td>
<td>Indonesia</td>
<td>3323120</td>
<td>4016</td>
<td>727.16</td>
</tr>
<tr>
<td>B. s. striata</td>
<td>FMNH</td>
<td>458417</td>
<td>Tissue</td>
<td>Amazonas, Brazil</td>
<td>3356633</td>
<td>3312</td>
<td>567.45</td>
</tr>
<tr>
<td>B. s. striata</td>
<td>USNM</td>
<td>659836</td>
<td>Tissue</td>
<td>Guyana</td>
<td>3490850</td>
<td>3849</td>
<td>673.11</td>
</tr>
<tr>
<td>B. s. striata</td>
<td>AMNH</td>
<td>DOT-2960</td>
<td>Tissue</td>
<td>Bolivar, Venezuela</td>
<td>3918563</td>
<td>3175</td>
<td>560.32</td>
</tr>
<tr>
<td><strong>B. s. “striata”</strong></td>
<td>No Voucher</td>
<td>S07</td>
<td>Blood</td>
<td>San Cristóbal, Galápagos</td>
<td>2001734</td>
<td>4149</td>
<td>808.6</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S09</td>
<td>Blood</td>
<td>San Cristóbal, Galápagos</td>
<td>4157015</td>
<td>4058</td>
<td>826.32</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S10</td>
<td>Blood</td>
<td>Santa Cruz, Galápagos</td>
<td>1725219</td>
<td>3849</td>
<td>735.59</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S12</td>
<td>Blood</td>
<td>Santa Cristóbal, Galápagos</td>
<td>1473235</td>
<td>4106</td>
<td>817.68</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S13</td>
<td>Blood</td>
<td>San Cristóbal, Galápagos</td>
<td>1879390</td>
<td>4095</td>
<td>811.51</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S15</td>
<td>Blood</td>
<td>Santa Cruz, Galápagos</td>
<td>2877133</td>
<td>4205</td>
<td>830.17</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S18</td>
<td>Blood</td>
<td>Santa Cruz, Galápagos</td>
<td>1684279</td>
<td>3078</td>
<td>500.5</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S19</td>
<td>Blood</td>
<td>Santa Cruz, Galápagos</td>
<td>1300834</td>
<td>2782</td>
<td>529.50</td>
</tr>
<tr>
<td>B. s. sundevalli</td>
<td>No Voucher</td>
<td>S24</td>
<td>Blood</td>
<td>Santa Cruz, Galápagos</td>
<td>2153971</td>
<td>3436</td>
<td>626.88</td>
</tr>
<tr>
<td>B. v. anthonyi</td>
<td>CAS</td>
<td>96552</td>
<td>Tissue</td>
<td>California, USA</td>
<td>2590790</td>
<td>3370</td>
<td>593.99</td>
</tr>
<tr>
<td>B. v. anthonyi</td>
<td>CAS</td>
<td>98481</td>
<td>Tissue</td>
<td>California, USA</td>
<td>1630209</td>
<td>3594</td>
<td>629.45</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>USNM</td>
<td>507836</td>
<td>Tissue</td>
<td>Bocas Del Toro, Panama</td>
<td>1938284</td>
<td>4319</td>
<td>962.5</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>FMNH</td>
<td>508643</td>
<td>Tissue</td>
<td>Illinois, USA</td>
<td>1101833</td>
<td>3343</td>
<td>579.8</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>AMNH</td>
<td>DOT-23977</td>
<td>Tissue</td>
<td>New Jersey, USA</td>
<td>3686102</td>
<td>4336</td>
<td>992.22</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>AMNH</td>
<td>DOT-13747</td>
<td>Tissue</td>
<td>Florida Keys, USA</td>
<td>2046712</td>
<td>4254</td>
<td>909.15</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>USNM</td>
<td>646729</td>
<td>Tissue</td>
<td>La Paz, El Salvador</td>
<td>2920262</td>
<td>4267</td>
<td>864.33</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>USNM</td>
<td>646807</td>
<td>Tissue</td>
<td>La Paz, El Salvador</td>
<td>1779001</td>
<td>3381</td>
<td>605.36</td>
</tr>
<tr>
<td>B. v. virescens</td>
<td>AMNH</td>
<td>DOT-7359</td>
<td>Tissue</td>
<td>Florida, USA</td>
<td>1978673</td>
<td>3626</td>
<td>545.69</td>
</tr>
<tr>
<td>G. Gossei</td>
<td>NCBI</td>
<td>LC541454</td>
<td>Tissue</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>N. nycticorax</td>
<td>CAS</td>
<td>90391</td>
<td>Tissue</td>
<td>California, USA</td>
<td>1883641</td>
<td>3655</td>
<td>584.78</td>
</tr>
</tbody>
</table>

Table 1. Data associated with the genetic samples used in this study including the institution they were sourced from, museum ID code, collection location, as well as sequencing information. (*) This sample is from a region that has overlap between B. s. javanica and B. s. carcinophila. (**) This sample presents morphologically as the mainland South American B. s. striata, despite coming from the Galapagos.
Figure 1. World map representing the different biogeographic regions the genus *Butorides* occupies. The color of the polygon is representative of the biogeographic region. The black borders of the polygons represent different populations/subspecies within a biogeographic region. The circles represent different samples that were included in this study, both morphometric and genetic.
Figure 2. Morphometric results

Figure 2. (A) - (C) Box and whisker plot of the tarsus length (mm), culmen height (mm), and wing length (mm) of the globally defined populations measured from preserved and live specimens. Points marked as red boxes are outliers. The area represented in grey is the dataset average +/- the standard deviation. Regions are shaded in reference to the biogeographic region they were collected in. Biogeographic regions are shown at the bottom of the figure. Circles inside each region is a sample that was measured in this study. (D) shows the Canonical Discriminant Function group centroids color coded by biogeographic region. The dominant characteristic in Function 1 is culmen height and width. The dominant characteristic in Function 2 is tarsus length and culmen width.
Figure 3. Bayesian consensus tree of ultraconserved elements (UCEs) in the genus *Butorides*. Colors represent the biogeographic region of the samples. Maps on the right of the phylogeny contain points labeled with the sample ID and shows the precise collection location. Nodes that are circled in red are those with a posterior value lower than 0.99 and are labeled with their value. Unlabeled nodes have a posterior value of 0.99 - 1.0.
Figure 4. Phylogeny of *Butorides* from mitochondrial data

Figure 4. Bayesian consensus tree of mitochondrial genome in the genus *Butorides*. Colors represent the biogeographic region where the genetic sample was collected. Maps on the right of the phylogeny contain points labeled with the sample ID and shows the precise collection location. Nodes that are circled in red are those with a posterior value lower than 0.99 and are labeled with their value. Unlabeled nodes have a posterior value of 0.99 - 1.0.
References


