University of North Carolina Asheville
Journal of Undergraduate Research
Asheville, North Carolina
May 2022

# Phylogenomic Data Provide Resolution to the West Indian Boa (genus *Chilabothrus*) Tree of Life

Willow Ogburn
Biology Department
The University of North Carolina Asheville
One University Heights
Asheville, North Carolina 28804 USA

Faculty Advisor: Dr. Graham Reynolds

## Abstract

Ten years ago, nine species of West Indian Boas (genus *Chilabothrus*) were recognized from the Greater Antilles and Lucayan Archipelago. Since then, an additional five species have been described or elevated, necessitating a revision of the phylogeny for the genus. I obtained DNA samples from every species of *Chilabothrus*, including the newly described *C. ampelophis*, and generated genomic data using ultraconserved elements sequencing. I used bioinformatic approaches to clean and analyze the dataset, then used phylogenomic analyses to construct a new genomic-scale tree of life for the genus. This new phylogeny resolves some previously recalcitrant nodes in the phylogeny, and includes the placement of *C. ampelophis*, thus providing a much improved and more accurate view of the evolutionary relationships of these boas. This phylogeny will be useful in upcoming studies of the evolution of specialization and historical biogeography in the genus.

## 1. Introduction

The genus *Chilabothrus* (Figure 1) is comprised of 14 extant species distributed in the Greater Antilles (Jamaica, Hispaniola, Cuba, and Puerto Rico), and the Lucayan Archipelago[1,2]. This clade dates back to a Miocene colonization of the Greater Antilles from South America, followed by further diversification within the West Indies[1]. Species within *Chilabothrus* exhibit significant morphological and ecological variation, with evidence for the repeated evolution of small specialist and large generalist species throughout the islands[2]. Large-bodied species are ecological generalists, using both terrestrial and arboreal habitats and preferring a variety of prey such as mammals, large ectotherms, and birds[2]. Small-bodied species are either terrestrial or arboreal and are almost exclusively saurophagus, feeding primarily on *Anolis* lizards[2]. *Chilabothrus* species are nocturnal and occupy a wide range of habitats, including xeric scrub and montane rainforest. Large and small species are distributed across the West Indies, and no single island has more than one large species[2].

Figure 1. The Southern Bahamas boa (*Chilabothrus chrysogaster*). This is one of five species of the genus that occupy the Lucayan Archipelago. This individual is from a new population that was discovered by Dr. Reynolds in March 2022 on a remote island in the Turks and Caicos Archipelago. Photograph by R. Graham Reynolds.

A study by Reynolds et. al. (2013) detailed the first multilocus phylogenetic analysis of the West Indian Boas[1]. Prior to this study, phylogenetic investigations into their origin of diversification had been contradictory[3,4], and further evidence was needed to support the idea that *Chilabothrus* arose from one colonization event with among-island diversification. While previous analyses used morphological data[3] or a single mitochondrial gene[4], the Reynolds et al. (2013) analysis used 10 genes, two mitochondrial and eight nuclear, to examine the relationships between 12 species and 4 subspecies within the genus. The phylogeny was constructed using a Bayesian framework, and historical timing of island colonization was estimated using fossil-calibrated divergence time analyses. This study included tissue samples from (at the time) all described species of *Chilabothrus,* with the exception of the Hispaniolan *C. gracilis*. From these analyses, the divergence of the genus *Chilabothrus* was found to be approximately 25.3 Mya, and the Cuban boa *C. angulifer* was shown to represent the basal extant lineage (Figure 2). The phylogeny showed that the Puerto Rican bank clade, consisting of *C. inornatus* and *C. monensis*, diverged shortly after, followed by the Jamaican clade (Figure 2). However, the pattern of divergence for the Bahamian and Hispaniolan boas was shown to be more complicated. A possible paraphyletic relationship in *C. striatus* was found, where *C. exsul* was the closest relative of the Bahamian *C. striatus*, whereas the western Great Bahama Bank subspecies was found to be a separate lineage and given the name *C. strigilatus*. This study was essential in beginning to define relationships within the genus and provided evidence for a single dispersal event from mainland Central/South America into the Proto-Greater Antilles. It also defined the name for the genus (formerly *Epicrates*), identified an additional species (*C. strigilatus*) for conservation prioritization, and suggested evidence for repeated evolution of small body forms.

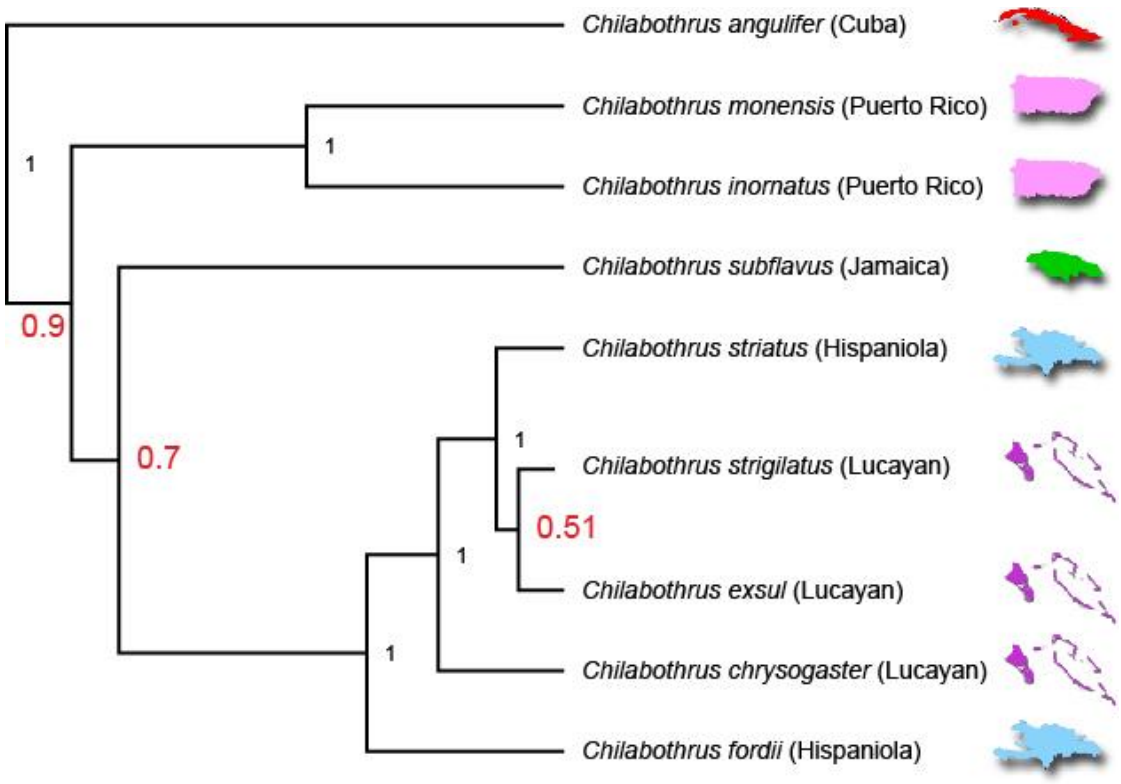


Figure 2. First multilocus phylogeny of the genus *Chilabothrus*[1].

Figure 2 Modified phylogeny from Reynolds et al. 2013, showing the topology as well as the recalcitrant nodes (in red)[1]. Note that *C. angulifer* is shown as the sister lineage to all other *Chilabothrus*.

This evidence was expanded upon in a 2016 paper by Reynolds et. al. which aimed to further investigate head shape and body size diversification within *Chilabothrus*[2]. The genus contains two main morphotypes: large-bodied generalists and small-bodied specialists, both of which are distributed across the West Indies. Using phylogenetic analyses, the authors tested association between size evolution, cranial morphology, and ecological specialization. Additionally, they expanded upon the phylogeny previously created in 2013, resolving the placement of the Hispaniolan *C. gracilis*. This study used a subset of two individuals from each species from the previous study as well as newly generated sequence data, using PCR to amplify 10 genes, to create a phylogeny for 11 species of *Chilabothrus*. The authors generated morphological measurements from live-captured animals as well as intact collection individuals. Ultimately, no phylogenetic signal in the evolution of body size was found, demonstrating repeated evolution of small body size. Transitions to small body size were also associated with an increase in habitat specialization as well as increased diversification of head shape, suggesting a reduction in size allows for increased specialization and raising the possibility of adaptive evolution within the genus. Although the placement of *C. gracilis* was resolved, showing it to be the sister taxon of *C. fordii*, some nodes, such as the placement of *C. angulifer*, were still problematic.

West Indian boids have often been cited as an example of island-induced diversification[2]; however, the evolutionary relationships within the genus *Chilabothrus* have not yet been fully clarified. The last multilocus tree was published in 2013[1], and many nodes lacked resolution, particularly *C. strigilatus* and *C. exsul* in the Bahamian clade. This phylogeny indicated that *C. strigilatus* was an independent species from *C. striatus*, and emphasized the need for further phylogenetic resolution. A 2018 paper by Reynolds et. al. using mitochondrial DNA also showed many areas of low support, again in regards to *C. strigilatus* and *C. exsul*. While this phylogeny suggested the placement of the newly described *C. schwartzi* and *C. argentum*, additional analyses are important to solidify their placement within the clade. Further trees using mtDNA showed differing topologies compared to the 2013 phylogeny and suggested that the Cuban Boa (*C. angulifer*) was sister to the Puerto Rican clade[5], contradicting previous findings that stated it was sister to all other West Indian boids[1] (Figure 2). Additionally, 5 new species in the genus *Chilabothrus* have been

described in the last 10 years, a 55% increase in species diversity. Properly defining the evolutionary relationships within the genus have important implications for future evolutionary research, as well as conservation efforts.

Phylogenomic studies using large amounts of sequence data are critical in understanding species-level relationships between many taxa[6] and have previously proved useful in resolving evolutionary relationships among squamate reptiles[7]. This study uses ultraconserved elements (UCEs), a class of genetic marker that provide conserved priming sites within the tetrapod genome and allow for the generation of expansive phylogenomic datasets[8]. UCEs have a high level of sequence conservation across the tetrapod genome, which is useful for identification and alignment[8]. Priming sites are conserved, making primer "bait" kits useful across tetrapods, but the flanks of the conserved sites are variable. Thus, phylogenetically informative data are found by sequencing the flanks of these UCEs and can be used to examine variability between taxa.

Here I describe the generation and analysis of phylogenomic data to estimate a robust phylogenetic hypothesis for the West Indian Boas.

## 2. Methodology

### 2.1 Sample collection

I gathered one sample from each of the 14 species of *Chilabothrus*. Tissue samples consisted of either tail or scale clips preserved in 95% ethanol and stored at -80 degrees C. Samples were previously extracted using the Wizard SV® Kit (Promega, Madison, WI), to extract whole genomic DNA from these samples, then stored at -20°C.

### 2.2 Genomic Library Prep and Sequencing

I used an electrophoresis gel check of fragment lengths and fluorometric quantitation of extractions using a Qubit 3.0® (Thermo Fisher Scientific, Waltham, MA) with dsDNA BR Assay Kit to examine the quality of DNA extractions. I rejected extractions with a sample concentration below 2.5 ng/ml. I sequenced all boa samples at the ~1100 base pair (bp) mtDNA locus cytochrome B (CytB) to confirm extraction quality and to provide Sanger sequence data to compare to our genomic data. After selecting the extractions with the highest quality and concentration of DNA, I sent extraction aliquots to RAPiD GENOMICS ® LLC (Gainesville, FL) for UCE sequencing and sequence capture following the original UCE protocol[6]. Aliquots were first normalized to equal concentration, then libraries were constructed using the myBaits® UCE Tetrapods 5kv1 primer set, which has the potential to bait up to 5,000 UCE loci across both the nuclear and mitochondrial genomes. The libraries were purified on a Pippin Prep, and unique barcode adapters were ligated to each library (one library from each sample) to allow for subsequent identification of samples. Libraries were then pooled together creating one master library, and Illumina adapters were ligated to each DNA fragment to allow for fragment hybridization to the Illumina flow cells. The library was then sequenced on a single-lane Illumina® HiSeq 2500 run using 250 bp paired-end chemistry.

### 2.3 Bioinformatics

Illumina sequencing produced resulting .fastq files which contain both base pair reads and a quality score associated with each read. I batched these raw fastq files together and performed preliminary quality control visualization using FASTQC 0.10.1[9]. FASTQC allows for estimates of read quality at the level of individual reads to examine read length and quality dropoff at the ends of the sequencing reactions via a set of graphical visualizations and tables (Figure 3).

Following this, I batched raw fastq files into the Linux *phyluce* pipeline, a toolkit of conda-packaged Python scripts to assist in UCE extraction and identification, to process raw fastq reads[10]. I used ILLUMIPROCESSOR[11], a Python script that calls TRIMMOMATIC[12] to trim adapters and low-quality sequence reads. I then used the Velvet algorithm[13] with *kmer*=51 to assemble contigs from cleaned reads. I matched contigs to the UCE Tetrapods 5kv1 probe set to extract UCE loci from the contigs file ((https://github.com/faircloth-lab/uce-probe-sets/blob/master/uce-5k-probe-set/uce-5k-probes.fasta). I aligned these UCE loci using the MAFFT algorithm[14] with edge trimming. I also removed prepended locus names from the alignment. I created a 50% complete data matrix, retaining aligned UCE loci with at least 50% of taxa represented for every locus. I performed UCE bioinformatics on the *CHILABOTHRUS* Dell® PowerEdge© Server (16c, 128gb RAM) running Ubuntu 20.04 at the University of North Carolina Asheville. I further

visualized the dataset using R in RStudio[15], including using the package ggplot2[16] to graph the number of loci recovered from each sample following bioinformatic analysis.

### 2.4 Phylogenomic Analyses

I used our concatenated 50% UCE data matrix to estimate a maximum-likelihood tree using the RAxML algorithm implemented in the RAxML HPC[17,18] executable within the *phyluce* package for LINUX[8]. I first ran a search for the best ML tree, then matched this tree with 1000 bootstrap repetitions in accordance with the GTRGAMMA model. I visualized the resulting tree in FigTree[19].

## 3. Results

### 3.1 Sequencing and Bioinformatics

I obtained 150-bp raw sequence reads (Figure 3) with a range of $1.3x10^6$-$3.1x10^6$ reads in each direction (forward and reverse) per individual, with a mean of $1.9x10^6$. All 14 *Chilabothrus* species were sampled, as well as a *Boa constrictor* outgroup (from an introduced population in Puerto Rico). I retained a range of $2.5x10^6$-$6.2x10^6$ reads after cleaning and adapter trimming, with a mean of $3.8x10^6$. The total base pairs for each individual ranged from $3.7x10^8$-$1.0x10^9$, with a mean of $5.8x10^8$ base pairs.

### 3.2 UCE Phylogenomic Analyses

The total number of UCE loci per individual ranged from 1,379-3,210, with a mean of 2,741 loci per sample, compared to a target enrichment of 5,000 loci (Figure 4). Per sample, total base pair number ranged from $3.0x10^5$-$1.7x10^6$. Each UCE loci had a mean length of 219-515 base pairs. Because no UCE loci had all samples represented, I reduced the dataset to a 50% complete data matrix that contained $1.1x10^6$ total base pairs. In the entire matrix, samples had between 29-83% missing data, with a range between 195,318-815,737 base pairs.

The ML tree I generated had high bootstrap values across all nodes, with a lowest bootstrap value of 96 (Figure 5). Certain relationships within my phylogeny differ from previously hypothesized relationships. In this topology, *C. angulifer* is placed as sister to all other *Chilabothrus* rather than within the Puerto Rican clade. *C. subflavus*, the Jamaican boa, is placed with high confidence as sister to all other *Chilabothrus* species. Additionally, the topology of the Bahamian clade is resolved within this ML tree. *C. exsul* and *C. strigilatus* are shown to be sister taxa, and are closest to the group composed of *C. schwartzi* and *C. argentum*. This topology also supports two colonizations of the Bahamas from Hispaniola. One of these colonizations resulted in the divergence of *C. exsul*, *C. strigilatus*, *C. schwartzi*, and *C. argentum*, while the second lineage, *C. chrysogaster*, colonized the Turks and Caicos islands. The newest discovered species within the genus, *C. ampelophis*, is placed as the sister taxa to *C. fordii* within the Hispaniolan clade. Other relationships within the tree are consistent with previously generated phylogenies[1,2,3].

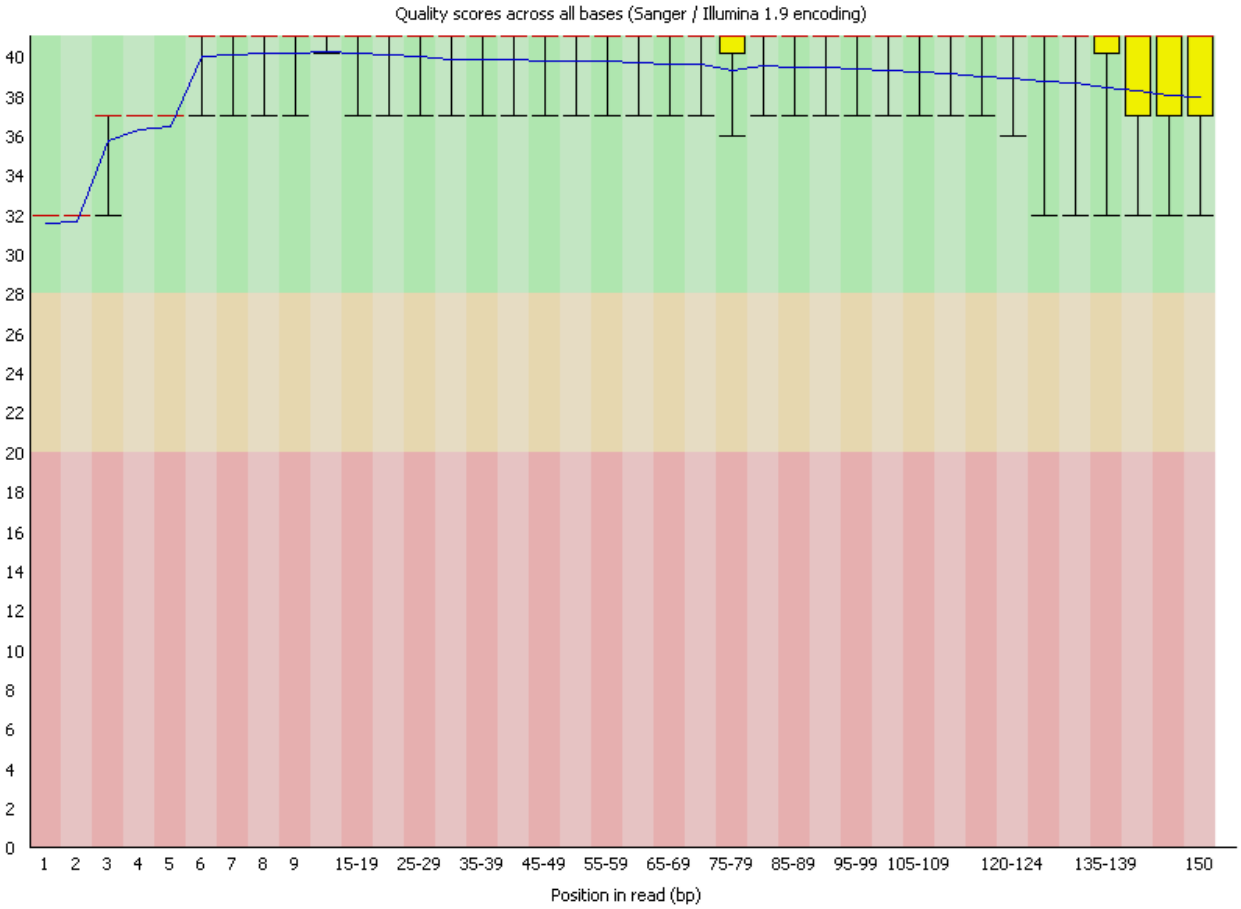


Figure 3. Example output from FastQC[7] of a 150bp sequence read.

Figure 3 All positions in every 150bp sequence read for a species (*Chilabothrus argentum*, in this case) was of high quality. Quality scores are on the left, and scores above 28 are considered high quality. All samples showed very similar profiles.

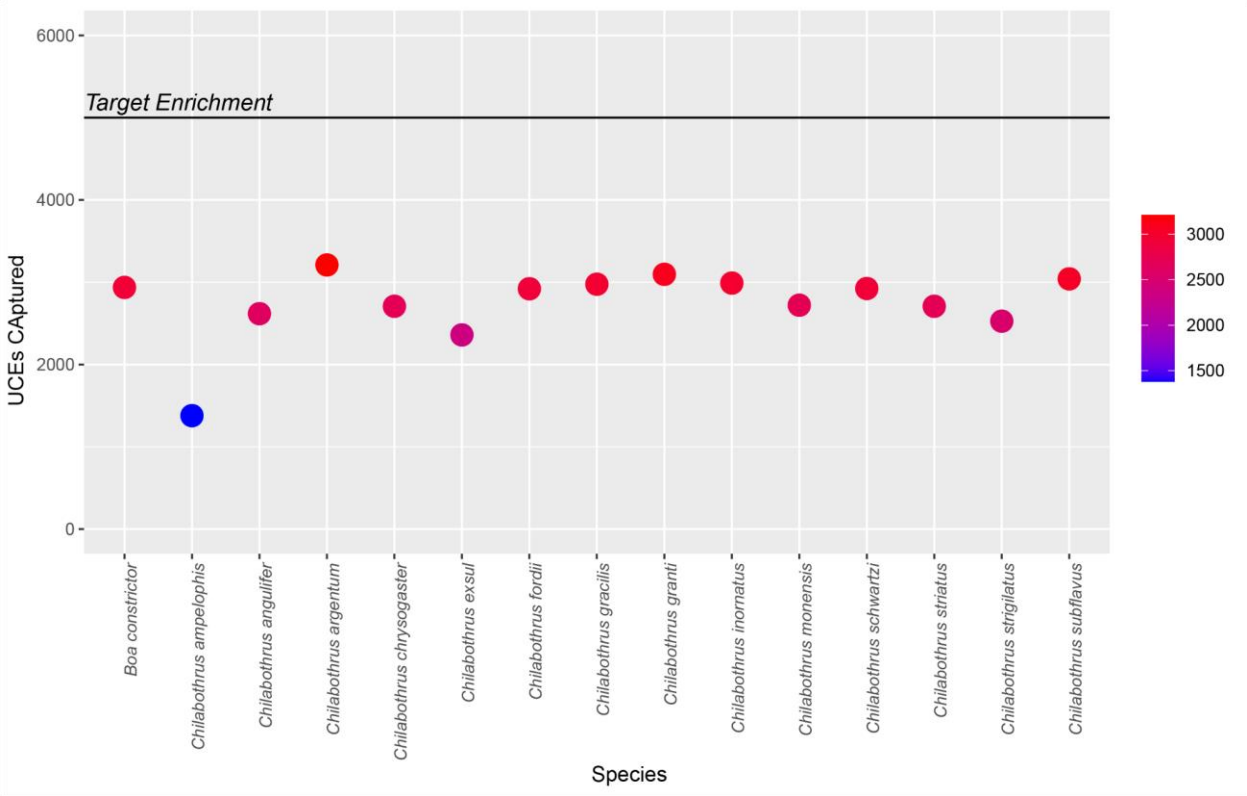


Figure 4. UCE loci enrichment of all taxa.

Figure 4 All species used in this study had enrichment above 2,000 loci, indicating a successful enrichment, cleaning, and aligning procedure. The exception is *C. ampelophis*, which still had over 1,500 loci captured. The black line indicates the maximum capture of 5,000 loci.

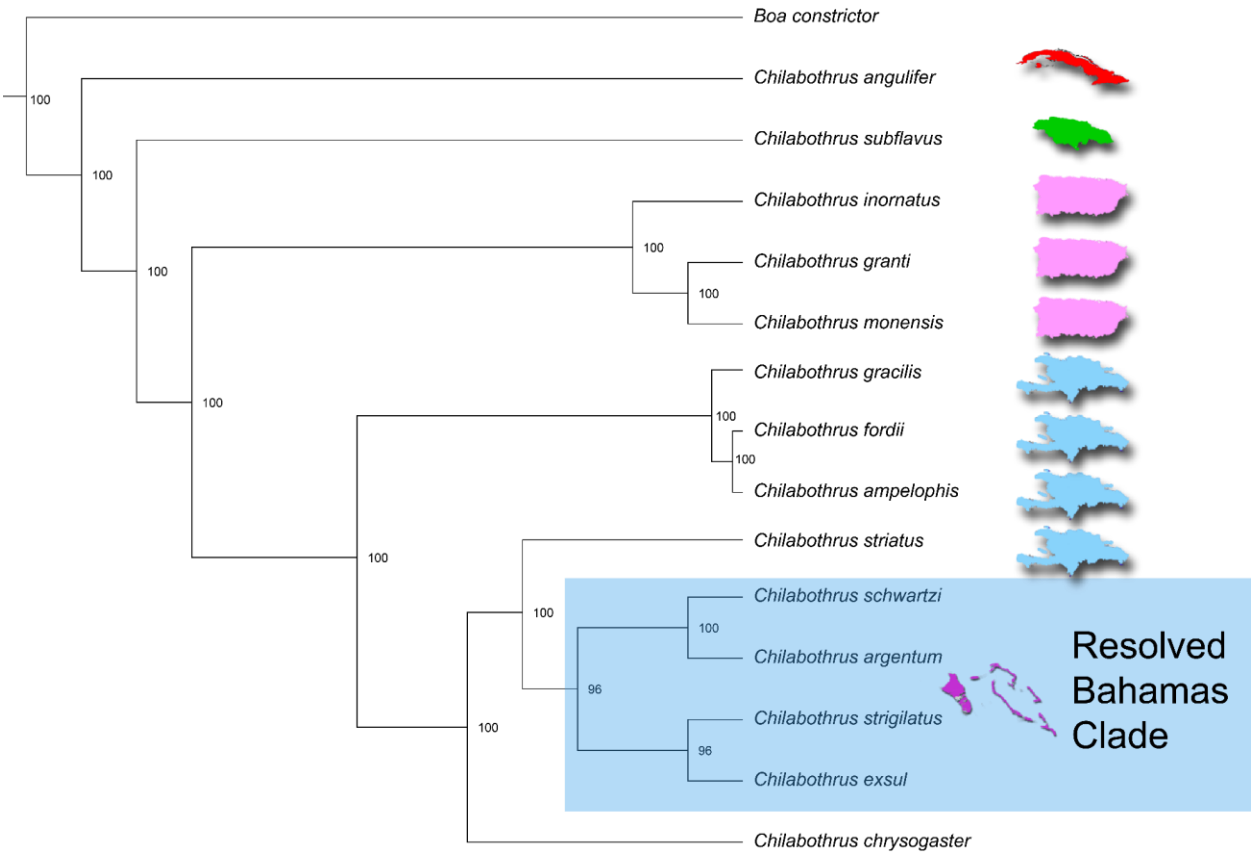


Figure 5. Maximum likelihood phylogeny of 14 *Chilabothrus* species and a *Boa constrictor* outgroup.

Figure 5 This phylogeny was generated from the 50% UCE data matrix using RaxML HPC[17]. The numbers at the nodes represent bootstrap support values, and the outlines of the islands the samples came from are on the right. Red = Cuba, Green = Jamaica, Pink = Puerto Rico Bank + Mona, Blue = Hispaniola, Purple = Bahamas. Note that the bottom lineage *C. chrysogaster* is from the Lucayan Archipelago as well.

## 4. Discussion

My phylogenetic analyses support many of the previously established relationships within the genus, with a few notable exceptions. The topology created from our UCE analyses (Figure 5) suggests that *C. angulifer* is the sister taxon to all other extant *Chilabothrus*, refuting the findings from previous topologies using mtDNA which suggested *C. angulifer* was sister to just the Puerto Rican clade[2,5]. Additionally, it clarifies the relationships between *C. exsul*, *C. strigilatus*, and *C. striatus*. Previous mitochondrial studies suggested up to three independent colonizations of the Lucayan Archipelago from Hispaniola[2,5]. However, both the maximum likelihood phylogeny and Bayesian phylogeny generated from my UCE analyses indicate that there were at most two independent colonizations. Within the Bahamian boas clade, *C. argentum* and *C. schwartzi* are found to be most closely related, while *C. exsul* and *C. strigilatus* are sister lineages.

Other relationships within the genus are consistent with previously hypothesized placement. For example, the placement of *C. schwartzi*, which was described in 2018, is concordant with the topology suggested by the mtDNA tree generated by Reynolds et al. 2018. It is also preserved as the sister species to the newly-discovered *C. argentum* from the central Bahamas. These two species together represent a central/eastern lineage of boas distinct from the central/northern lineage comprising the species *C. strigilatus* on the Great Bahama Bank and *C. exsul* on the Little Bahama Bank. The newest discovered *Chilabothrus* species, *C. ampelophis*, was described in 2021, and an mtDNA phylogeny was generated to suggest its placement as sister to *C. fordii*[20]. The phylogeny I generated affirms this topology with high confidence.

My phylogeny also helps to resolve some recalcitrant nodes deeper within the tree. For example, the node at which *C. subflavus* diverges has been problematic, with relatively low support and contested placement in previous phylogenies (Figure 2). The topology generated by Reynolds et. al. 2013 suggested that *C. subflavus* was the sister taxon of the Hispaniolan and Lucayan Archipelago clade. However, both the maximum likelihood and Bayesian phylogenies (Figure 5) that I generated suggest that *C. subflavus* diverged after *C. angulifer*, and is sister to all other lineages. This relationship was supported by a high posterior probability value (0.98) in comparison to previous topologies[1], and is the first time *C. subflavus* has been placed with high confidence values.

One potential limitation of using UCEs to construct this phylogeny is that allele variability (in diploids) is not assembled using this technique[21]. Instead, reads are assembled into contigs that only contain canonical nucleobases, ignoring variable positions[21]. The use of allelic sequences may improve phylogenetic hypotheses, particularly at more recent timescales, but there are few existing computational pipelines to integrate allelic information in phylogenetic analyses[21]. Additionally, it is important to note that despite this potential limitation, the methods I used in this study use the same statistical models (maximum likelihood and Bayesian) to create this topology as previous phylogenetic studies of the genus *Chilabothrus* and other New World boas[1,2,5,20]. The difference between my phylogeny and previous methods ultimately comes from the amount of data used. For example, the phylogeny created by Reynolds et. al. 2013 used ten loci and a few thousand base pairs total across taxa compared to the mean of 2,741 loci and over 800,000 base pairs per sample I obtained. Because of this, I was able to create a topology with more robust support in comparison to previously generated trees.

Additionally, my work helps to emphasize the usefulness of UCE data in large-scale phylogenomic studies. By using an ultraconserved elements sequencing pipeline, I was able to generate a large amount of phylogenetically informative data to help resolve previously recalcitrant nodes in the topology. The use of UCEs has previously been used to successfully resolve evolutionary relationships in other groups of squamate reptiles[6]. Species-level topological definition generated via UCEs also holds important conservation implications. Within the genus *Chilabothrus*, two species are classified as endangered and one as critically endangered on the IUCN Red List, and many face threats such as habitat loss and encroachment. A well-supported phylogeny for the genus is therefore important for informing both future genetic and conservation work regarding the West Indian Boas.

## 5. Acknowledgements

I thank Dr. Meigs and Dr. Helm for serving on my committee. Funding for this work was provided by the Reynolds Lab at UNC Asheville.

## 6. References

1. R. Graham Reynolds et al. “Molecular phylogeny and historical biogeography of West Indian boid snakes (*Chilabothrus*),” *Molecular Phylogenetics and Evolution* 68 (2013): 461-470.
2. R. Graham Reynolds et al. “Ecological specialization and morphological diversification in Greater Antillean boas,” *Evolution* 70 (2016): 1882-1895.
3. Peter J. Tolson. “Phylogenetics of the boid snake genus *Epicrates* and Caribbean vicariance theory,” *Occas. Pap. Mus. Zool. Univ. Mich.* 715 (1987): 1-68.
4. Paula C. Rivera et. al. “Species Delimitation in the Continental Forms of the Genus *Epicrates* (Serpentes, Boidae) Integrating Phylogenetics and Environmental Niche Models,” *PLoS ONE* 6 (2011), https://doi.org/10.1371/journal.pone.0022199.
5. R. Graham Reynolds et al. “Rediscovery and a redescription of the Crooked-Acklins boa, *Chilabothrus schwartzi* (Buden, 1975), Comb. Nov.,” *Brevoria* 558 (March 2018): 1-16.
6. Yuchi Zheng and John J. Wiens. “Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species,” *Molecular Phylogenetics and Evolution* 94 (2016): 537-547.
7. Jesse L. Grismer et. al. “The Eurasian invasion: phylogenomic data reveal multiple Southeast Asian origins for Indian Dragon Lizards,” *BMC Evolutionary Biology* 16 (Feb 2016): 1-11.
8. Brant C. Faircloth et. al. “Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales,” *Syst. Biol.* 61 (2012): 717-726.
9. Simon Andrews. *FastQC: a quality control tool for high throughput sequence data.* PC. 2014.

10. Brant C. Faircloth. “PHYLUCE is a software package for the analysis of conserved genomic loci,” *Bioinformatics* 32 (2016): 786-788.
11. Brant C. Faircloth. *Illumiprocessor: a Trimmomatic wrapper for parallel adapter and quality trimming.* faircloth-lab. PC. 2014.
12. Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. “Trimmomatic: a flexible trimmer for Illumina sequence data,” *Bioinformatics* 30 (2014): 2114-2120.
13. Daniel Zerbino and Ewan Birney. “Velvet: algorithms for de novo short read assembly using de Bruijn graphs,” *Genome Res* 18 (2008): 821-829.
14. Kazutaka Katoh and Daron M. Standley. “MAFFT Multiple sequence alignment software version 7: improvements in performance and usability,” *Mol. Biol. Evol.* 30 (2013): 772-780.
15. R Core Team. *R: A language and environment for statistical computing*. V.4.2.0. R Foundation for Statistical Computing. PC. 2018.
16. Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. PC. 2016.
17. Alexandros Stamatakis. “RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models,” *Bioinformatics* 22 (2006): 2688-2690.
18. Alexandros Stamatakis, Paul Hoover, and Jacques Rougemont. “A fast bootstrapping algorithm for the RAxML web-servers,” *Syst. Biol* 57 (2008): 758-771.
19. Andrew Rambaut. *Figtree*. V 1.4.4. Andrew Rambaut Group. PC. 2018.
20. Miguel A. Landestoy T, R. Graham Reynolds, and Robert W. Henderson. “A small new arboreal species of West Indian boa (Boidae; *Chilabothrus*) from Southern Hispaniola,” *Brevoria* 571 (August 2021): 1-20.
21. Tobias Andermann et. al. “Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements,” *Syst. Biol* 68 (2018): 32-46.