MOLECULAR PHYLOGENETICS OF AMBLYCORYPHA (ORTHOPTERA: TETTIGONIIDAE): A MOLECULAR MORPHOMETRIC AND MOLECULAR TAXONOMIC APPROACH

A thesis presented to the faculty of the Graduate School of
Western Carolina University in partial fulfillment of the
requirements for the degree of Master of Science in Biology

By

Charles Benedict Sither

Director: Dr. Kefyn Catley
Professor, Department of Biology

Committee Members: Dr. Tim Forrest, Department of Biology
Dr. Katherine Mathews, Department of Biology
Dr. Brian Byrd, School of Health Sciences

June 2018

ACKNOWLEDGEMENTS

First, I would like to express my greatest appreciation to Dr. Kefyn Catley for his continuing support, guidance, and encouragement throughout my graduate career at WCU. He has graciously imparted knowledge and experiences that I will continually utilize in my future studies and research. I would not have been in this position without the support and mentorship from Dr. Brian Byrd, Dr. Bruce Harrison, and Brittania Bintz throughout my baccalaureate and postbaccalaureate adventures. I am deeply indebted to their unwavering patience, direction, and expertise that they have so selflessly shared throughout the years. I would also like to extend a very special thanks to Dr. Katherine Mathews and Dr. Tim Forrest. Without their continuous help and support in the forms of specimens, technical knowledge, and mentorship this project would not have possible.

To my friends and fellow lab mates and graduate students, the wonderful memories crafted and shared during my time at WCU has created a lasting experience that I will hold close for the rest of my life. Lastly, I want to give my deepest thanks to my parents, Ellen, Mary, and Tom Sither. All of whom nourished and cultivated my scientific curiosity and intrinsic drive. A grand thank you all for graciously taking the time to shape me into the person I am today. These brief sentences do not begin to express how deeply everyone mentioned here has impacted my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC – Akaike Information Criterion
BI – Bayesian Inference
bs – non-parametric bootstrapping
bp – base pair
BPP – Bayesian Phylogenetics and Phylogeography
CBC – compensatory base change
CO1 – cytochrome oxidase subunit 1
dG – delta Gibbs free energy
dsDNA – double-stranded deoxyribonucleic acid
ESS – effective sample size
I – invariant
ITS1 – internal transcribed spacer 1
ITS2 – internal transcribed spacer 2
G – gamma
gDNA – genomic deoxyribonucleic acid
GTR – general time reversible
K2P – Kimura-2 parameter
MBG – molecular biology grade
MCMC – Markov Chain Monte Carlo
ML – Maximum Likelihood
MPS – massively parallel sequencing
mtDNA – mitochondrial deoxyribonucleic acid
my – million years
myo – million years old
NNI – nearest-neighbor interchange
nuDNA – nuclear deoxyribonucleic acid
OTU – operational taxonomic unit
PCR – polymerase chain reaction
pp – posterior probability
rDNA – ribosomal deoxyribonucleic acid
rjMCMC – reversible jump Markov Chain Monte Carlo
RNA – ribonucleic acid
rRNA – ribosomal ribonucleic acid
SNP – single nucleotide polymorphism
Spp – multiple species
Sp – single species
ssDNA – single-stranded deoxyribonucleic acid
Taq – polymerase isolated from *Thermus aquaticus* bacteria
USC – unified species concept

ABSTRACT

MOLECULAR PHYLOGENETICS OF AMBLYCORYPHA (ORTHOPTERA: TETTIGONIIDAE): A MOLECULAR MORPHOMETRIC AND MOLECULAR TAXONOMIC APPROACH

Charles Benedict Sither, M.S.

Western Carolina University (June 2018),

Director: Dr. Kefyn Catley

Genus *Amblycorypha* (Orthoptera: Tettigoniidae) is comprised of 14 nominal species exhibiting highly similar morphologies. Three major morphologically similar species complexes exist in *Amblycorypha* – the *uhleri*, *oblongifolia*, and *rotundifolia* complexes. While each species is morphological similar, the songs that males use to attract mates differ drastically among species. Recently collected male and female mating songs suggest multiple undescribed species exist within the *rotundifolia* complex. Using molecular techniques, I aim to delimit species groups within *Amblycorypha* and attempt to reconstruct their evolutionary histories. The ITS1 (~461 bp), 5.8S (174 bp), and ITS2 (240 bp) nuDNA regions and a partial CO1 (523 bp) mtDNA gene were sequenced using massively parallel sequencing technologies. The CO1 mtDNA region was the most variable (10.1% overall mean distance), followed by ITS2 (1.1% mean distance), ITS1 (0.9% mean distance), and 5.8S (0.02% mean distance). A single nucleotide polymorphism was present in 5.8S uniting the *uhleri* complex as a clade. K2P interspecific differences had large overlap in both nominal species groups and unknown species groups. ML and MSC phylogenetic analyses recovered the *uhleri* complex as monophyletic, while the *oblongifolia* and *rotundifolia* complexes were polyphyletic. Additionally, 6 distinct clades of 'unknown specimens' were recovered in ML and MSC analyses using all gene targets. Finally, *A. bartrami* may represent a species complex based on the molecular evidence presented here. This study represents the first molecular phylogeny for genus *Amblycorypha*. While incomplete, this study supports additional cryptic species within the *rotundifolia* complex that were initially detected based on male songs.

CHAPTER 1: SEQUENCING AMBLYCORYPHA SPECIMENS

Introduction

Taxonomy, Species Recognition, and Isolating Mechanisms

The family Tettiigonidae (Insecta: Orthoptera), commonly referred to as "katydids", represent one of the most noticeable nocturnal acoustic insect taxa in the United States. Acoustic signaling in katydids provides a means for long distance communication between sexes. Tettiigonids have long been recognized as an enigmatic taxon due to convergent morphological similarities and inconsistent results from DNA-based phylogenies (Legendre et al. 2010, Mugleston et al. 2013). Katydid species, like many insects, were classified by their morphological features with the assumption that different morphology indicated reproductive isolation from sister species. Many entomologists were hesitant to classify morphologically indistinguishable populations based on ecological factors like song-type or habitat preferences, preferring species classifications based on their ability to be distinguished "on the pin", i.e., by distinct morphological characters that all individuals in a population share, which can be referenced in a museum or collection (Hubbell 1954, Hubbell 1956). However, in many cases new morphological features were not discovered until after acoustic and ecological differences were used to cluster individuals from a population into similar groups. This was first demonstrated in *Nemobius* spp. (Orthoptera: Gryllidae; Alexander 1957) and *Chorthippus* spp. (Orthoptera: Acrididae; Perdeck 1958), independent of each other. Alexander (1957) noted that even though songs differed among *Nemobius* spp., there were no morphological differences in the lengths and tooth spacing of their stridulatory files. The behavioral differences were chiefly based on different rhythms of the songs. Perdeck (1958) demonstrated that by combining the length of the forewing, the number of stridulatory teeth, and width of wing costal areas, each *Chorthippus* spp. could be differentiated morphologically thus indicating a physical component to song differences. Both studies led to a currently recognized theory that song-types in orthopoterans are always different between species occurring in the same space and time (Drosopoulos and Claridge 2005).

The subfamily Phaneropterinae is one of the nineteen subfamilies found within Tettigoniidae (Mugleston et al. 2013). Phaneropterine katydids engage in pair-forming duet systems where a female responds acoustically to a male call. Both male and female alternate their signals while moving towards each other until they reach immediate proximity of one another and begin other premating rituals (Spooner 1968, Spooner 1995). Competitive interactions occur during mate pair formations between a male-female duet and adjacent males, where nearby males 'eavesdrop' on a female signal and employ acoustic tactics intended to disrupt neighboring male communications and pair forming duets with a nearby female (Drosopoulos and Claridge 2005, Hammond and Bailey 2003, Bailey and Field 2000). These competitive interactions may act as a driver of speciation in Phaneropterine katydids, thereby contributing to their enormous diversity (Spooner 1968).

A member of the subfamily Phaneropterine, the genus *Amblycorypha* contains 14 described species within 3 recognized complexes – the *rotundifolia*, *uhleri*, and *oblongifolia* groups (Table 1). The *oblongifolia* complex comprises 5 different species with relatively simple songs. Rehn and Hebard (1914) originally separated *Amblycorypha* into groups I and II based on morphological differences in the shape of the humeral sinus of the lateral lobes. Group I possessed a strongly rectangulate humeral sinus while the humeral sinus on group II was either not rectangulate or subobsolete. Group I contained the *oblongifolia* and *uhleri* complexes while group II represented what is now known as the *rotundifolia* complex.

Within the *oblongifolia* complex, *A. floridana*, *A. huasteca*, and *A. insolita* were reported as being morphologically similar but occupying distinctly different habitats and geographic ranges (Figure B1). Rehn and Hebert (1914) recognized that *A. insolita* occupied an arid desert habitat ranging from Big Bend, Texas into the Sonoran Desert with a northern terminus south of Flagstaff, AZ, *Amblycorypha huasteca* was known from Texas, Oklamhoma, Kansas, southeastern New Mexico, and northeastern Mexico occupying prairie and mixed forest habitats. It was also noted that *A. floridana* and *A. carinata* were found to differ based on the length and size of the female ovipositor and male stridulatory field, with *A. carinata* possessing a broader longer ovipositor and a broader stridulating field. *Amblycorypha*

floridana was restricted to Florida and the Florida panhandle while *A. carinata* ranges from southern Georgia to New England along the eastern seaboard and into the Gulf Coast regions as far west as Texas. It wasn't until Walker (2004) found that *A. carinata* was elevated to specific status based on distinct differences in male songs between *A. floridana* and *A carinata.* Both species have overlapping distributions, yet they do not appear to hybridize. Lastly, *A. oblongifolia* has a nearly ubiquitous distribution in the eastern and Midwestern United States and Canada. *Ambycorypha oblongifolia* can be morphologically distinguished from other *Amblycorypha* males and females by the stridulating field length being larger than that of the pronotal disc and the female ovipositor being arcuate with margins of teeth.

Distinctions were made between the *oblongifolia* complex and *uhleri* complex based on overall size and body shape. Walker (2004) later identified the "virtuoso katydids" describing 5 species in the *uhleri* complex based on song-types, morphological features, and geographic distributions (Figure B2). Each species was first identified based on distinctive male songs with morphological features identified later after populations were separated. *Amblycorypha uhleri* can be distinguished based on geography, occurring only on the Edwards plateau in Texas, and from syntopic species such as *A. rivograndis* by the length of the tegmina. Habitat information is lacking for *A. uhleri* but specimens were collected by Walker (2004) in roadside weeds and overgrown hillsides in Texas. The other Texas species, *A. rivograndis* has been found in thickets and "jungle-like" environments in southern Texas and in oak stands in the northern parts of its distribution. *Amblycorypha longinicta*, and *A. cajuni* have syntopic distributions and are found in herbaceous vegetation in old fields and woodland edges. However, *A. cajuni* is found more often in lowland habitats with moist vegetation while *A. longinicta* can be found in mesic habitats with better drainage. *Amblycorpha arenicola* occurs in longleaf pine and other sandhill communities that are xeric or mesic habitats throughout the southeastern United States. All three species, *A. cajuni*, *A. longinicta*, and *A. arenicola* can be morphologically separated based on the number of stridulatory teeth and file length in males and shape and size of female ovipositors (Walker 2004).

The group II members from Rehn and Hebert (1914) represented the *rotundifolia* complex (geographic distributions in Figure B3). The *rotundifolia* members are mostly flightless and contain 4 nominal species. *Amblycorypha parvipennis* is the only *rotundifolia* complex member that can be readily identified based on morphology; the tegmina conceal the hindwings at rest while in the other 3 species the hindwings protrude beyond the ends of the tegmina approximately 2-6mm. Also, *A. parvipennis* is geographically isolated from other *rotundifolia* complex members, since it is distributed west of the Mississippi river and found in prairie and old pasture habitats (Isley 1941, Walker et al. 2003). The other three members, *A. rotundifolia, A. bartrami*, and *A. alexanderi* are solely distinguished based on male song types. Habitat differences exist for *A. bartrami*, which is found in xeric and mesic scrub oak sandhill habitats. However, *A. alexanderi* and *A. rotundifolia* are syntopic across most of their geographic distributions (Walker et al. 2003). Limited dispersal due to flightlessness and allopatric barriers in the western North Carolina mountains may reduce gene flow between populations belonging to the *rotundifolia* complex. This is evidenced by recent song and morphological data collected by Forrest et al. (unpublished data) supporting the hypothesis that additional cryptic species remain undescribed within *Amblycorypha*.

Table 1. *Amblycorypha* Valid Species List

Complex	Genus/Species	Author/Publication Date
uhleri Complex	*Amblycorypha arenicola*	Walker, 2004
	Amblycorypha cajuni	Walker, 2004
	Amblycorypha longinicta	Walker, 2004
	Amblycorypha rivograndis	Walker, 2004
	Amblycorypha uhleri	Stål, 1876
rotundifolia Complex	*Amblycorypha alexanderi*	Walker et al., 2003
	Amblycorypha bartrami	Walker et al., 2003
	Amblycorypha parvipennis	Stål, 1876
	Amblycorypha rotundifolia	(Scudder, 1862)
oblongifolia Complex	*Amblycorypha carinata*	Rehn and Hebard, 1914
	Amblycorypha floridana	Rehn and Hebard, 1905
	Amblycorypha huasteca	(Saussure, 1859)
	Amblycorypha insolita	Rehn and Hebard 1914
	Amblycorypha oblongifolia	De Geer, 1773

Massively Parallel Sequencing

Massively parallel sequencing (MPS) approaches are any of several high-throughput sequencing technologies that can concurrently sequence from millions of individual DNA molecules. MPS contrasts from traditional capillary sequencing technologies, like Sanger sequencing, in which sequence data is derived from a population of DNA (Voelkerding et al. 2009). MPS technologies are being slowly adopted in the fields of phylogenetics and phylogeography. Difficulties in adopting MPS technologies have centered around statistical methods that previously required homologous gDNA regions to infer phylogenetic relationships. However, MPS is inherently a cost-effective approach for multiplexing and library preparation (McCormack et al. 2013). Recent advances in small nucleotide polymorphism (SNP) coalescent methods (Boucaert et al. 2014, Leaché et al. 2015) and cost reductions in sequencing, e.g., human whole genome sequencing approaching $1,000 (https://www.genome.gov/sequencingcosts/), has created opportunities for phylogeneticists and systematists to investigate hypotheses like the timing of

insect radiations (e.g, Misof et al. 2014) and deep-level phylogenies (e.g., Song et al. 2017) that would have previously been prohibitively expensive. Many MPS technologies rely on the use of adapter sequences that bind to a solid support on which sequencing occurs. These short oligonucleotide sequences can be added onto the 5' ends of targeted PCR amplification primers to enable cost-effective, amplicon-based sequencing of a limited number of gene targets (Illumina®, Inc. 2016). MPS technologies for amplicon-based sequencing have promise as an inexpensive means of sequencing multiple amplicons from hundreds of individuals in a few runs. Such technology is less laborious and typically more effective when generating sequences from multiple loci than traditional Sanger sequencing methods.

The ITS Cistron and Cytochrome Oxidase subunit 1

The internal transcribed spacer region 1 (ITS1) and 2 (ITS2) (Figure B4) separate the ribosomal DNA (rDNA) subunits. ITS1 separates 18S and 5.8S, while ITS2 separates 5.8S and 28S. The ITS regions are non-coding functional RNA that promote proper folding and maturation of 5.8S and 28S rRNA by promoting cleavage domains and binding sites for proteins involved in ribosome maturation (Nazar 2003, Nazar 2004, de la Cruz et al. 2015). ITS2 is thought to possess a four- to six- helicoidal ring secondary structure that is highly conserved across a broad range of eukaryotes (Wolf et al. 2005, Coleman 2007, Coleman 2009). Coleman (2009) suggested that relatively conserved ITS2 regions are stabilized by selective pressures ensuring proper rRNA processing. Lastly, the ITS cistron (ITS1, 5.8S, and ITS2) is thought to undergo coevolution with ITS1 and ITS2 exhibiting similar amounts of G/C content and possessing similar rates of change (Torres et al. 1990). This highly conserved secondary structure of ITS1 and ITS2 provides a unique opportunity for comparison at multiple taxon levels. Both ITS1 and ITS2 primary nucleotide sequences are heavily utilized as gene markers for differentiating cryptic insect species due to their locations between highly conserved rDNA regions, variability found within the primary sequence structure, and numerous repetitive copies found within each cell (Ullrich et al. 2010, Li et al. 2010, Coleman 2009).

Additionally, ITS1 and ITS2 folded secondary structures are typically conserved in lower taxon levels (e.g., genus and species) with little variation between groups, but less conserved between Families

and Orders (Coleman 2007) allowing for elucidation of evolutionary relationships at different levels with either primary sequence or secondary structure information. Lastly, RNA secondary structures are stabilized by molecular hydrogen bonding and only certain nucleotides can form thermodynamically stable hydrogen bonds. Due to these interactions nucleotides that form Watson-Crick base pairs tend to co-vary within an ITS RNA molecule. Compensatory base changes (CBCs) are mutations that occur in Watson-Crick binding of a stem structure. Wolf et al. (2013) proposed a CBC species concept by determining the likelihood that one CBC predicted two different species based upon Mayr's biological species concept.

In addition to ITS2, the mtDNA gene cytochrome oxidase subunit 1 (CO1) is a barcoding gene heavily utilized in molecular systematics due to its divergence among closely related species and location between highly conserved mtDNA regions (Herbert et al. 2003, Folmer et al. 1994). Phylogenetic analyses based on mtDNA regions typically produce robust phylogenetic results but are not always representative of a species tree (Pamilo and Nei 1988, Nichols 2001, Wahlberg et al. 2003). Using mtDNA regions alone in phylogenetic analyses can produce results that disagree with nuclear DNA (nuDNA) and unite taxa due to introgressive hybridization and mitochondrial capture (McGuire et al. 2007). However, in conjunction with nuDNA regions, robust phylogenies can be developed for testing species delimitation and reconstructing evolutionary histories (Toews and Brelsford 2012).

Aim 1

Given the lack of morphological differences among cryptic *rotundifolia* complex members, but who do differ in song phenotypes, and recent successes using ITS1, 5.8S, ITS2, and CO1 in generating phylogenetic hypotheses, it is reasonable to assume that these are viable gene targets for reconstructing the evolutionary histories and delimiting closely related cryptic *Amblycorypha* species. Thus, this study aims to 1) design novel primers with Illumina® R1 and R2 adapter sequences covalently attached to the forward and reverse primers for multiplex sequencing of ITS1, 5.8S, and ITS2 gene regions, along with a partial CO1 gene using massively parallel sequencing, and 2) compare sequence variation between and within each group to evaluate their utility for delimiting species and reconstructing phylogenies.

Methods

Specimen Acquisition and DNA Extraction

Specimens included in this study were 143 *Amblycorypha* spp. and 1 *Atlanticus* sp. as the outgroup that were identified based on song-type and morphological characteristics. All specimens were either field collected or obtained from personal museum collections (Table A1; includes sequenced specimens only). Each specimen has either field and/or laboratory recordings of song-type made by Forrest et al. used for species identifications. Hind femora were removed from field collected specimens shortly after preservation, i.e., being frozen for pinning. Each leg was kept frozen at -80^{o}C prior to genomic DNA (gDNA) extraction. Genomic DNA was extracted from either the proximal half of the hind femur or the entire hind femur. A Qiagen DNeasy® tissue kit (Qiagen, Valencia, CA) was used by Scobie (2013) for gDNA extraction. While an Invitrogen PureLink® Genomic DNA Mini kit (Invitrogen, Waltham, MA) following the Human Buccal Swab Lysate procedure detailed in the PureLink® Genomic DNA Kit user guide (revision 2.0) was used during study. Afterwards, *Amblycorypha* gDNA extracts were stored at -80^{o}C for long term storage.

Primer Design and PCR

Katydid specific nested primers were designed using Primer3 (Koressaar and Remm 2007, Utergasser et al., 2012). *Amblycorpyha* ITS1, 5.8S, and ITS2 sequences (Forrest et al., unpublished data) were aligned in ClustalX version 2.1 (Larkin et al. 2007) to an annotated *Scudderia furcata furcata* Brunner von Wattenwyl (GenBank Acession: AM888963.1) ITS1, 5.8S, and ITS2 rDNA sequence. Homologous sequence regions were identified across all samples. Prospective primers were checked for compatibility in Primer3. Four primers were selected based on amplicon length (<600 bp) and annealing temperatures (Figure 1; Table 2).

Nextera® XT DNA library preparation uses transposomes to simultaneously fragment and incorporate adapter sequences into the 5' end of DNA molecules. Adapters are used for incorporating sample dependent index sequences incorporate (for sample identification) and to enable binding of template DNA to a flow cell for facilitation of MPS sequencing. Adding Nextera® XT adapter sequences

onto the ends of PCR primers allows the transposome incorporation step to be skipped. Here, nested forward and reverse primers were designed with Nextera® XT (Illumina® Inc., San Deigo, CA) transposase adapter read 1 and 2 sequences on the 5' ends of each primer sequence, respectively.

Initial PCR amplification cycles for amplifying a ~3,700 bp rDNA region (the exact size for *Amblycorypha* species is unknown) using conserved primers LR7 (Vigalys and Hester 1990) and NS19b (Bruns Lab, UC Berkeley, unpublished primer; https://nature.berkeley.edu/brunslab/tour/primers.html) consisted of an initial denaturation step at 95ºC for 5 minutes, followed by 35 cycles of 95ºC for 60 seconds, 55ºC for 60 seconds, and 72ºC for 5 minutes. A final 72ºC extension was 10 minutes. Each 25 µL reaction consisted of 22.5 µL Platinum® PCR SuperMix High Fidelity (ThermoFisher Scientific, Waltham MA) – which consists of 22 U/mL complexed recombinant *Taq* DNA polymerase, *Pyrococcus* sp. GB-D thermostable polymerase, and Platinum® *Taq* antibody; 66 mM Tris-SO_4 (pH 8.9), 19.8 mM $(NH_4)_2SO_4$, 2.4 mM $MgSO_4$, 220 µM dNTPs, and stabilizers - 200 pM concentration of forward and reverse primers, and 1.5 µL of unquantified gDNA template PCR amplicons that were visualized by ethidium bromide-stained 1% agarose gels buffered with 1X TBE.

Nested PCR amplification cycles for katydid specific primers with Nextera® XT transposase adapter reads 1 and 2 (Table 2) consisted of an initial denaturation step at 95ºC for 5 minutes, followed by 35 cycles of 95ºC for 60 seconds, 53ºC for 45 seconds, and 72ºC for 45 seconds. A final 72ºC extension was 7 minutes Each 25 µL reaction consisted of 22.5 µL Platinum® PCR SuperMix High Fidelity (ThermoFisher Scientific, Waltham MA) – which consists of 22 U/mL complexed recombinant *Taq* DNA polymerase, *Pyrococcus* sp. GB-D thermostable polymerase, and Platinum® *Taq* antibody; 66 mM Tris-SO_4 (pH 8.9), 19.8 mM $(NH_4)_2SO_4$, 2.4 mM $MgSO_4$, 220 µM dNTPs, and stabilizers - 200 pM concentration of forward and reverse primers, 0.5µL of nuclease-free H_2O, and 1 µL of uncleaned and unquantified PCR amplicon template. Nested PCR amplicons were visualized by ethidium bromide-stained 1% agarose gels buffered with 1x TBE.

Massively Parallel Sequencing

Nested PCR amplicons were quantified using a Qubit 2 dsDNA High Sensitivity kit (ThermoFisher Scientific, Waltham MA). PCR products were subsequently diluted in molecular biology grade (MBG) H_2O resulting in a final 40 pg/µL PCR product concentration. Following normalization, diluted PCR amplicons were cleaned using Agencourt AmPure XP PCR purification (Beckman Coulter, Inc.). Afterwards, AmPure cleaned PCR amplicons were quantified using an Agilent Bioanalyzer (Agilent Technologies) to assure that PCR amplicons obtained from all samples were at equimolar concentrations. Next, Illumina® i7 and i5 Nextera® indices were added onto Nextera® XT transposase adapter sequences using an 8-cycle PCR reaction consisting of an initial 95ºC denaturation at 3 minutes, followed by 8 cycles of 95ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 30 seconds. A final 72ºC extension was 5 minutes. Each 20 µL reaction consisted of 10 µL 2x KAPA HiFi HotStart Ready Mix (KAPA Biosystems, Inc.). Illumina® indices are unique, 8 nucleotide sequences which are incorporated onto the 5' and 3' ends of each PCR amplicon acting as unique identifiers for each sample. With 8 different 5' sequences and 12 different 3' sequences, 96 samples can be uniquely labeled based on their index combination. Indexed PCR amplicons were cleaned using Agencourt AmPure XP PCR purification beads (Beckman Coulter Inc., Pasadena CA) and eluted into 10 µL of MBG H_2O. Amplicon peaks were quantified using an Agilent Bioanalyzer DNA 1,000 kit (Agilent Technologies). Indexed amplicons were diluted to a 4 nM concentration. Diluted indexed PCR amplicons (5 µL) were pooled together to create a final normalized library.

DNA denaturation was performed by adding 5 µL of 0.2 N NaOH to 5 µL of 4 nM normalized library, followed by a 5 minute room temperature incubation. Afterwards, 990 µL of pre-chilled HT1 buffer was added resulting in a 20 pM denatured library with a 1 mM NaOH concentration. Two MiSeq® FGx sequencing runs were conducted with a 6 pM final library concentration (180 µL of a 20 pM denatured library into 420 µL of pre-chilled HT1 buffer). One MiSeq® FGx sequencing run was performed with an 8 pM final concentration (240 µL of a 20 pM denatured library into 360 µL HT1 buffer). A PhiX control was prepared and diluted into the same respective picomolar concentration as the

final sample library concentration. A total of 30 µL denatured and diluted PhiX control was spiked into 570 µL of diluted library. Afterwards, combined library and PhiX control was incubated at 96ºC for 2 minutes then immediately placed in an ice-water bath for 5 minutes and loaded onto the MiSeq® reagent cartridge. The MiSeq FGx™ Forensic Genomics System instrument (Illumina® Inc., San Diego CA) was set for a "*.Fastq only" run with 301 cycles.

Sequence Assembly and Alignment

Sequence reads were assembled and analyzed in CLC Genomics Workbench 11.0.0 (Qiagen, Inc., Hilden, Germany). A workflow was designed for separating each PCR amplicon read based on size, sequence similarity, and sequence quality. First, paired-end reads were imported with any failed samples immediately discarded and primary QC reports generated for each sequence file prior to paired end reads being processed through a workflow. Paired-end reads were subsequently merged based on two separate categories, 1) predicted amplicon sizes with CS_Kay_F2:CS_Kay_R1 primer pair amplicons ~565 bp indicating a ~17 bp overlap between each read direction, thus a minimum overlap score of 30 with a mismatch cost of 2 and a gap cost of 3 was set and 2) CS_Kay_F3:CS_Kay_R2 primer pair amplicons predicted to be 473 bp with ~63 bp overlap between each read direction – a minimum overlap score of 100 was set with a mismatch cost of 2 and gap cost of 3. Merged reads were then trimmed based on a quality limit of $p \leq 0.05$. Any reads with more than 2 ambiguous bases were removed. Afterwards, adapter sequences and indexing primers were removed from the 5' end of all filtered reads. Trimmed sequence reads were mapped to non-target reference sequences, e.g., ITS2 sequences were mapped to ITS1 and CO1 *S. furcata* reference sequence, with a match score of 1, mismatch score of 2, insertion/deletion cost of 3, and length/similarity fraction set at 0.7 and 0.8, respectively. Non-target read mapping acted as a noise elimination measure preventing primer-dimer or short-read sequences from mapping to a random gene target.

CLC Microbial Genomics Module *de novo* OTU clustering tool was used for clustering and extracting similar sequences. Sequences were clustered together if they shared ≥97% similarity. Afterwards, clustered sequences were filtered based on the total number of reads (>10), or a proportion of

the reads present in a table (>0.5%), whichever was greater; this filter removes errant OTUs caused by sequence rarefication. Consensus sequences were extracted from OTU cluster tables and annotated. Since CO1 sequences do not overlap, unmerged reads were trimmed and mapped to a *Amblycorypha oblongifolia* De Geer CO1 sequence (GenBank: KM536809.1) and a major vote consensus sequence was extracted for each sample with ambiguous nucleotides placed in areas of low coverage (<10 reads). Sequences were aligned in CodonCode Aligner 7.1.2 (CodonCode Corporation, Centerville MA) under default assembly conditions. Afterwards, sequences were trimmed to remove all ambiguous nucleotide positions. Since all 658 bp of CO1 were not recovered, only nucleotide sequences are analyzed for CO1 because I could not be sure that any ambiguous nucleotides removed shifted the codon reading frame and thus changed the actual amino acid sequence.

ITS1, 5.8S, ITS2, and CO1 Sequence Validity

The NCBI BLAST (Altschul et al. 1990) function was used for comparing sequences generated in this study to sequences generated by other researchers. Additionally, if ITS1 and ITS2 are authentic sequences under functional and selective constraints rather than pseudogenes then they should contain similar GC content (Mullineaux and Hausner 2009). GC content and nucleotide diversity (π) for ITS1 and ITS2 were determined in MEGA7 software (Kumar et al. 2016).

ITS2 Secondary Structure

CBCs and ITS2 secondary structure variation between species were identified based on a secondary structure model that consisted of a four to five finger open-handed ITS2 secondary structure, containing a pyrimidine-pyrimidine mismatch present on helix II, the presence of helix IIa, and conserved sequence motif present on the terminal 5' end of helix II as outlined by Coleman (2007). Models of ITS2 secondary structures were developed using the Mfold web server RNA folding platform (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) under default folding conditions (Zuker 2003). ITS2 sequences were first folded without constraints then examined for conserved structures. Constraints were added to prevent binding between select nucleotides at the distal 3' and 5' ends. Afterwards, ITS2 sequences were homology folded in the ITS2 Ribosomal RNA Database

(http://its2.bioapps.biozentrum.uni-wuerzburg.de/) based on a structural model identified in Mfold (Schultz et al. 2005, Wolf et al. 2005, Selig et al. 2007, Koetchan et al. 2009, Koetschan et al. 2012, Ankenbrand et al. 2015). Consensus secondary structures for ITS2 were redrawn as radial structures in VARNA version 3.9.3 (http://varna.lri.fr/) (Darty et al. 2009). Compensatory nucleotides present in all helix and stem structures were identified based on VARNA radial illustrations. CBC locations and position of each nucleotide pair in the primary nucleotide sequence were recorded.

Pairwise Comparisons and Interspecific Distance

Overall mean distance for each gene target and all genes (rDNA + mtDNA) were estimated using MEGA7 (Kumar et al. 2016). An estimate for pairwise distance for individual specimens, pairwise distance between each group, and pairwise distance within each group was calculated using a Kimura 2-parameter (K2P) model plus Gamma (G) with both transitions and transversions included. A 1,000 bootstrap replicate was used to estimate standard error with gaps/missing data treated as complete deletions, i.e., sites with ambiguous nucleotides are removed before the analysis begins. While a K2P model is not optimal for the CO1 mtDNA region, it does not affect the overall result when compared to a p-distance model using nucleotide sequence data. Groups were categorized based on collection locations, i.e. ‘unknown’ specimens, or formal species names. CO1, ITS1, and ITS2 K2P distance line plots were determined by plotting the intraspecific and interspecific nucleotide sequence differences in the R package SPIDER (Brown et al. 2014).

Table 2. Primer pairs and expected annealing temperatures (°C) for nested PCR and sequencing

Primers	Sequence 5' → 3'	Annealing (°C)	%GC	Predicted Amplicon Size	Publication
28s and 18s rDNA Primers					
LR7	TACTACCACCAAGATCT	53.6	41.2	~3,700 bp	Vigalys and Hester (1990)
NS19b	CCGGAGAGGGAGCCTGAGAAC	68.9	66.7		Bruns Lab, UC Berkeley
Nested ITS1, 5.8S, and ITS2 rDNA					
CS_Kay_F2[a]	ATTCGCCCTTTCCGTAGGT	60.8	52.6		
CS_Kay_R1[a]	CAATTTGCTGCGTTCTTCAT	58.9	40.0	~565 bp	Novel primers
CS_Kay_F3[a]	AGTACAACCCTGAACGGTGG	59.8	55.0	~473 bp	
CS_Kay_R2[a]	CCCTTTCCTCCGCTTATTG	59.6	52.6		
CO1 mtDNA					
LCO1490	GGTCAACAAATCATAAAGATATTGG	59.7	32.0	658 bp	Folmer et al. (1994)
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	64.5	34.6		Folmer et al. (1994)

[a] Primers designed for this study

[b]Predicted amplicon sizes based on tf-21 sequence alignment for the ITS cistron and known sizes of partial CO1 amplicon

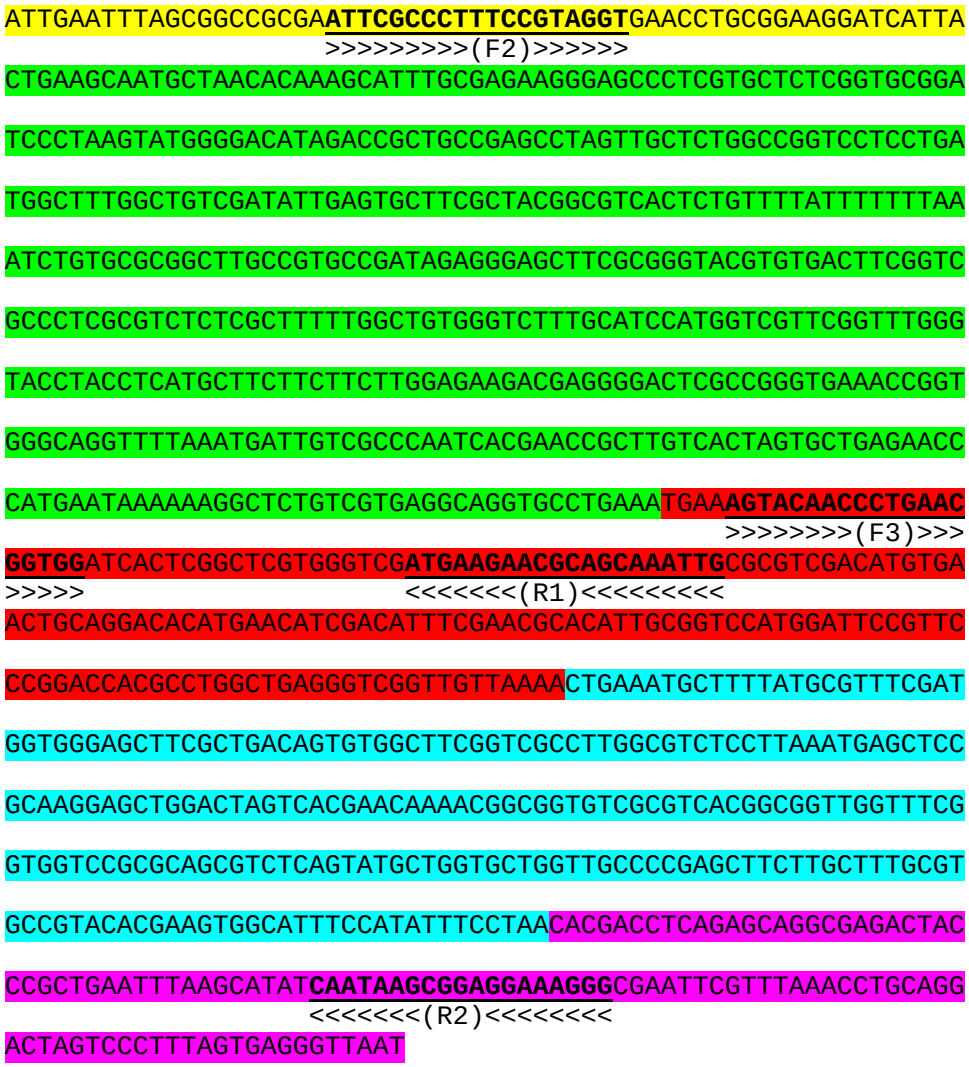

Figure 1. Annotated sequence (5' → 3') for sample tf-21 (BDB, unpublished sequences). Novel primer pair binding sites for CS_Kay_F2:CS_Kay_R1 (565 bp amplicon) and CS_Kay_F3:CS_Kay_R2 (473 bp amplicon) indicated with directional arrows for either forward or reverse primers. The 18S (yellow), ITS1 (green), 5.8S (red), and ITS2 (blue), and 28S (purple) are highlighted to show approximate lengths and primer annealing locations.

RESULTS

DNA Amplification and Massively Parallel Sequencing

A total of 142 of the 144 katydid samples had gDNA successfully extracted and confirmed either by a Qubit® Fluorometer dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, Waltham MA) or using PCR with ITS 1 and ITS 4 forward and reverse primers, with an expected 1,100 bp amplicon (White et al. 1990). A single *Amblycorypha* sp. and the *Atlanticus* sp. outgroup failed to have any extracted gDNA. Katydid-specific primers successfully amplified ITS1, 5.8S, and ITS2 rDNA regions with an average concentration of PCR product of 54.6 (SD = 10.57), 122.9 (SD = 33.7), and 76.2 (SD = 25.7) nM of PCR amplicons produced by primer pairs CS_Kay_F2:CS_Kay_R1, CS_Kay_F3:CS_Kay_R2, and Folmer et al. (1994) CO1 primers respectively.

Overall, 47 of the 142 samples were successfully sequenced with expected sequence coverage for ITS1, 5.8S, and ITS2, along with partial coverage for CO1. Full length MPS reads for primer pairs CS_Kay_F2:CS_Kay_R1 (ITS1 and 5' end of 5.8S) and CS_Kay_F3:CS_Kay_R2 (3' end of 5.8S and ITS2) were successfully recovered, while partial reads were recovered for CO1 from Folmer et al. (1994) primer pairs in sequencing run 1 (Table 3). Partial CO1 recovery with a minimum of 56 bp missing in the center of the amplicon was expected since a 658 bp amplicon would have no overlap with 301 cycles on the Illumina® MiSeq FGx. OTU extraction in CLC Genomics Workbench 11.0.0 enabled successful recovery and assembly of full length reads for ITS1, 5.8S, and ITS2 rDNA gene targets for most samples. Low read coverage in sequencing run 3 prevented full coverage for ITS1 gene targets (approximately 19% to 94% sequencing coverage for a 565 bp PCR product depending on the sample) in 27 samples. Error rates for each sequencing run varied greatly ranging from 2.46% to 10.35%. Q-scores >30 for each run varied from 19.99% to 76.20%. Technical laboratory errors caused sequencing run 2 to fail since <1 pM final library concentration was loaded onto the Illumina® MiSeq FGx. Additionally, technical errors in library preparation may explain the high error rates and low coverage seen in sequencing run 3. Due to low quality, high error rates, and low coverage, sequencing runs 2 and 3 were not used in any further analyses.

ITS1, 5.8S, ITS2, and CO1 Sequence Validity and ITS2 Secondary Structures

Top BLAST hits for all rDNA gene targets were their respective gene region from Phaneropterinae katydid members including *S. furcata furcata* (Genbank Accession: AM88963.1), *Poecilimon thessalicus* von Wattenwyl (Genbank Accession: AM888998.1), and *Barbitistes serricauda* Fabricius (Genbank Accession: AM888955.1). For CO1, the top 3 hits were all of *A. oblongifolia* (Genbank Accession: KR144595.1, KM536809.1, KM532357.1) CO1 genes sequenced from the Canadian DNA barcode project (Hebert et al. 2016). All targets were highly similar (>90% identity, <0.001 E-value) to their respective sequences found in Genbank. GC content comparisons between ITS1 and ITS2 was within 2% difference and nucleotide diversities for ITS1 and ITS2 were 0.9% and 1.1%, respectively. Lastly, a conserved, four-finger ITS2 secondary structure was found for all 47 ITS2 sequences (Figure 2). Watson-Crick binding in ITS2 secondary structures was highly conserved and no CBCs were identified across all samples. A conserved motif on the 5' side of helix III (5'-GGCGGTCCGCGCAGCGT-3') was present on all *Amblycorypha* and Phaneropterinae members. A similar 5' helix III sequence was present (5'- GGYGGTCCGCGCAGCGT – 3') on all other Tettigoniidae ITS2 sequences investigated in this study, including *Tettigonia caudata* (Charpentier) KT358357.1, *Acrometopa syriaca* Brunner von Wattenwyl KM981976.1*, Baritistes serricauda* en Lorraine AM888955.1*, Conocephalus fuscus* (Fabricius) KT823230.1*, Ducetia japonica* (Thunberg) KM981988.1*, Poecilimon luschani* Boztepe et al. AM889012.1*, Phaneroptera falcata* Poda AM888980.1*, Tylopsis lilifolia* (Fabricus) AM888956.1, and *S. furcata furcata* AM88963.1.

Characteristics of the Nucleotide Dataset

The aligned nucleotide dataset contains 48 samples with a total of 67,248 nucleotides and no missing or ambiguous sites (1,401 nucleotides per sample) from 47 different *Amblycorypha* specimens and a single *S. furcata furcata* obtained from GenBank. Trimming CO1 sequences of all ambiguous nucleotides resulted in 523 nucleotide sites present for each sample. The mean overall pairwise distance for the entire concatenated character matrix (excluding outgroups) was d = 0.041 ± 0.003 S.E.

Nuclear gene targets contained less overall mean distance compared to the mitochondrial gene target. The aligned 5.8S rDNA region comprised 175 bp along with the lowest overall mean distance of all gene targets at (d = 0.002 ± 0.002 S.E), along with a nucleotide diversity of (π) = 0.002 and Tajima's *D* = 0.38. Additionally, a single nucleotide polymorphism in 5.8S at position 2 of 175 (G → A) separated the *uhleri* complex from the *oblongifolia* and *rotundifolia* complexes. This was the only nucleotide polymorphism found within all sequenced *Amblycorypha* specimens. The *oblongifolia* and *rotundifolia* complexes had identical 5.8S rDNA sequences. The aligned ITS2 rDNA dataset contained 240 bp and an overall mean distance of 0.011 ± 0.003 S.E., genetic diversity of (π) = 0.011, and Tajima's *D* = -1.30. Two hundred and twenty nucleotide sites were conserved across all sequences with 13 parsimony informative (5.4%) sites and 7 singletons. The aligned ITS1 rDNA dataset had 463 bp with an overall mean distance of 0.009 ± 0.002 S.E., genetic diversity of (π) = 0.009, and Tajima's *D* = -1.38. Of the 463 bp, 32 bp were variable, 21 were parsimony informative (4.5%), along with 11 singletons. Lastly, the aligned CO1 mtDNA dataset had 523 bp with an overall mean distance of 0.101 ± 0.01 S.E., a nucleotide diversity of (π) = 0.085, and Tajima's *D* = 0.87. A total of 361 nucleotides were conserved across all sites with 135 parsimony informative sites and 27 singleton bps. Most of the variability within CO1 exists within the 3rd codon position. Tajima's *D* for all gene targets was between 0 ± 2, which is unlikely to be a significant value indicating that all sequences are possibly evolving under a random process (Tajima 1989).

Intra- and Interspecific Differences

Intraspecific distances for most species for all gene targets were <1.0%. However, certain species groups have intraspecific distances >1.0% for CO1 including *A. longinicta*, *A. bartrami*, *A. oblongifolia*, *A. parvipennis*, and *A. floridana* (Table 4). Line plots of K2P distance for CO1, ITS1, and ITS2 revealed that certain individuals were more closely related to heterospecific members than to their closest conspecific (Figure 3a,b,c). Due to a 2-step PCR approach, individual polymorphic sequences were not analyzed since I could not be sure if the polymorphism was biologically real, due to a rarefication event, or some other artifact from PCR amplification.

Table 3. Sequencing run quality scores and coverage metrics for each gene target.

	Sequencing Run 1	Sequencing Run 2	Sequencing Run 3
Reads	301 cycles + 16 cycles for indices	301 cycles + cycles for indices	301 cycles + 16 cycles for indices
Library Conc. (pmol)			
Clusters PF (%)	96.43 ± 0.29	55.76 ± 4.00	62.16 ± 3.44
Cluster Density	555 ± 16	81 ± 2	971 ± 28
Error Rate (%)	2.46	10.35	8.64
%≥Q30	76.2	19.99	36.56
Used in Dataset?	Yes	No	No
No. of Samples	49	27	96

Table 4. Intraspecific mean pairwise distances for CO1, ITS1, 5.8S, and ITS2 gene regions. Standard deviation is estimated by 1,000 bootstrap replicates.

Species	CO1 K2P %Mean Distance	ITS1 K2P %Mean Distance	5.8S K2P %Mean Distance	ITS2 K2P %Mean Distance
*A. alexanderi**	-	-	-	-
A. arenicola	0.5 ± 0.2	0.1 ± 0.1	0.01 ± 0.01	0.3 ± 0.2
A. bartrami	4.6 ± 0.7	0.7 ± 0.3	0.0 ± 0.0	1.0 ± 0.4
A. carinata	0.2 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.4 ± 0.4
A. floridana	1.9 ± 0.6	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
A. longinicta	8.9 ± 1.0	0.1 ± 0.1	0.01 ± 0.01	0.0 ± 0.0
*A. '*nrbartrami'	1.8 ± 0.4	1.2 ± 0.4	0.0 ± 0.0	1.5 ± 0.6
A. oblongifolia	7.7 ± 1.1	0.7 ± 0.4	0.0 ± 0.0	0.8 ± 0.6
A. parvipennis	6.7 ± 1.0	0.4 ± 0.2	0.0 ± 0.0	2.3 ± 0.8
*A. rotundifolia**	-	-	-	-
A. 'Unknown Big Creek'*	-	-	-	-
A. 'Unknown Conley Creek'*	-	-	-	-
A. 'Unknown Coweeta'*	-	-	-	-
A. 'Unknown Craggy'	0.4 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
A. 'Unknown Greenbrier Picnic'*	-	-	-	-
A. 'Unknown Joyce Kilmer'*	-	-	-	-
A. 'Unknown Max Patch'	0.7 ± 0.4	0.2 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
A. 'Unknown Rich Mountain'	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
A. 'Unknown Standing Indian'	0.7 ± 0.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
A. 'Unknown Georgia'	0.3 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.2

*singleton species could not have intraspecific calculated

**highlighted CO1 sequences have >2% intraspecific variation

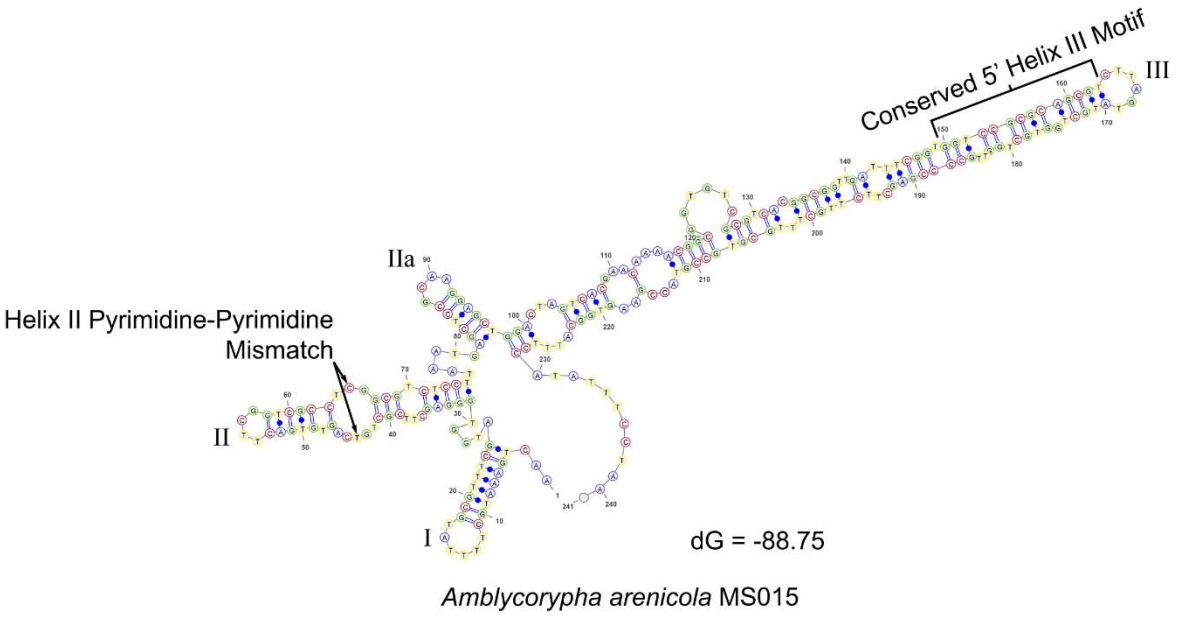

Figure 2. *Amblycorpyha arenicola* (MS015) ITS2 secondary structure model. Helices are labeled by roman numerals. A pyrimidine-pyrimidine mismatch (C→T mismatch here) exists on helix II along with the presence of helix IIa (between helix II and helix III). A conserved motif on the 5' side of helix III (5'-GGTGGTCCGCGCAGCGT-3') was present on all *Amblycorypha* and Phaneropterinae members. A similar helix III was present (either an exact match or with 1 or 2 transitions) on all other Tettigoniidae members in GenBank.

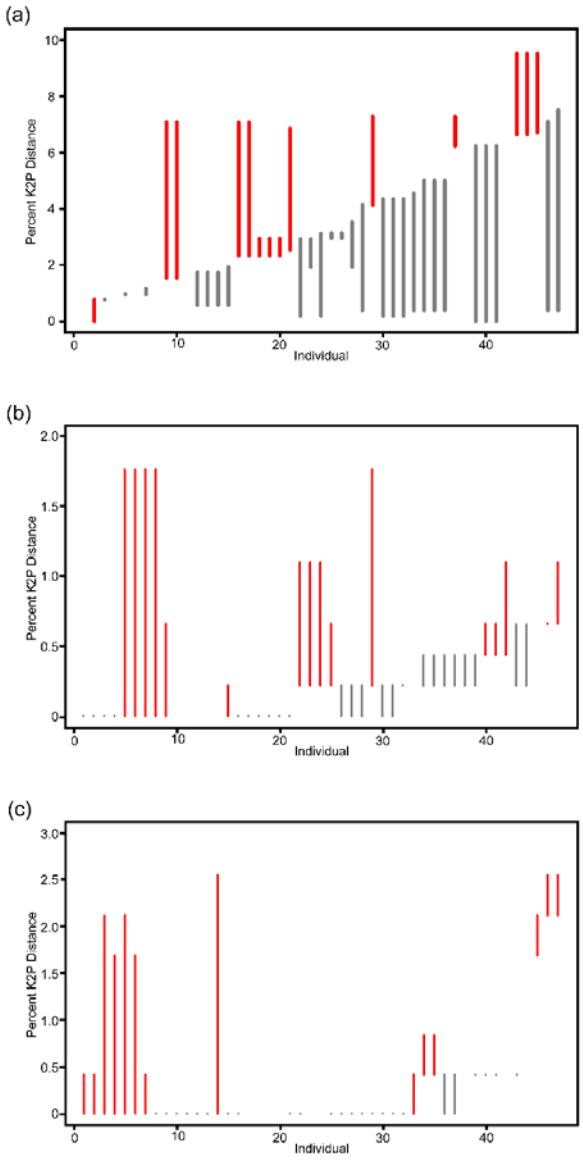

Figure 3. Line plots of intraspecific and interspecific distances for *Amblycorypha* CO1 (a), ITS1 (b), ITS2 (c) nucleotide datasets. Each line represents an individual within the dataset. The top of the line represents the closest interspecific distance while the bottom of the line represents the furthest intraspecific distance. Grey lines are individuals that are more closely related to their conspecifics while red lines are where that relationship is reversed, and the closest non-conspecific (bottom of the line) is closer than its nearest conspecific (top of the line). Individuals represented on the x-axis are sorted by distance from least to greatest. Individual identities cannot be determined here since the SPIDER package reorders individuals from least to greatest intraspecific distance.

DISCUSSION

Interspecific ITS2 Similarity

The ITS1 and ITS2 rDNA regions were more conserved between taxa than expected. Previous studies using ITS2 to delimit or identify dipteran, lepidopteran, and hymenopteran cryptic species lineages yielded large polymorphic differences (Li and Wilkerson 2005, Li et al. 2010). However, studies involving ITS1 and ITS2 in the Tettigoniidae reported highly similar sequences. For instance, Ullrich et al. (2010) used ITS1 and ITS2 sequences to infer the phylogenetic relationships among *Poecilimon* (Orthoptera: Tettigoniidae) members. ITS1 was the only phylogenetically informative target – ITS2 had reportedly no influence on ingroup tree topology. Similarly, Snyder et al. (2009) used ITS2 partial sequences along with other nuclear markers for temperate and tropical *Neoconocephalus* (Orthoptera: Tettigoniidae) members – ITS2 and other rDNA targets had low phylogenetic utility at a species level. It was only when combined with mtDNA sequences that a robust species level phylogeny was able to be reconstructed. *Amblycorypha* ITS2 exhibited a near homogenous secondary structure across all individuals. The peculiarity of ITS2 secondary structure homogeneity in Phaneropterine katydids was also seen in Ullrich et al. (2010), whereas no ITS2 variation was seen in ingroup taxa for the Barbitistini (Orthoptera: Tettigoniidae). Homology modeling with other ITS2 sequences from Phaneropterine katydid taxa showed nearly 100% helical structural similarities within the subfamily. Katydid taxa outside the Phaneropterine had a ITS2 secondary structure with helix IV present. It is possible that the helix IV is a derived loss in Phaneropterine katydids.

ITS1, 5.8S, ITS2, and CO1 Phylogenetic Utility

This study demonstrates the successful sequencing of the ITS1, 5.8S, and ITS2 rDNA regions and a partial sequence of the CO1 mtDNA region using novel katydid specific primers with Illumina® adapter sequences. ITS1, 5.8S, and ITS2 have low intra- and interspecific variation indicating that they may not provide phylogenetic resolution on a species level but could help resolve deeper nodes. Additionally, a partially recovered CO1 mtDNA region exhibited high intra- and interspecific nucleotide sequence variation. Hebert et al. (2003) demonstrated that CO1 sequences from 13,320 metazoan species

pairs ranged from 0% to 53.7% across a broad range of metazoan taxa with 98% of species pairs showing >2% CO1 sequence divergence. It is expected that CO1 sequences from the same species would have <2% intraspecific sequence, and in most cases, intraspecific differences in CO1 is <1%.

The identification and specific status for *Amblycorypha* members may be more nuanced now with the addition of molecular information accompanying morphological and song-type datasets. Four nominal taxa (*A. bartrami*, *A. longinicta*, *A. oblongifolia*, and *A. parvipennis*) had greater than 2% intraspecific distance in CO1 based on 2 or more individuals per species, which could represent lineages of cryptic species (Table 4) (Hebert et al. 2003). Higher intraspecific divergences in CO1 may be related to mitochondrial variants that were previously geographically isolated representing historical gene pool fragmentation. However, mtDNA can be under a non-neutral selection model and undergo selective sweeps mediated by mitochondrial variants that have more effective nuclear genome interactions (Ballard 2000, Gerber et al. 2001, Ballard and Whitlock 2004). Furthermore, intracellular symbionts have been shown to reduce mtDNA diversity across geographic ranges by rendering conspecifics with different intracellular symbionts infertile, or infecting and killing both hosts (Shoemaker et al. 1999, Jiggins 2003, Shoemaker et al. 2004, Hurst and Jiggins 2005). Lastly, male songs in katydids are typically under selective pressures to increase syllable rates, song length, or other song elements for female attraction (Tuckerman et al. 1993). Increases in syllable rates and song lengths are positively correlated with oxygen consumption and increased energetic costs, presumably this would be associated with more efficient nuclear-mitochondrial interactions (Bailey et al. 1993, 1995). If nuclear-mitochondrial interactions are not favored, then selective pressures from non-random mating would remove mitochondrial variants resulting in stabilizing selective pressures for efficient mitochondrial haplotypes within a species. Thus, these 5 species should be investigated further for differences in morphological, behavioral, and ecological factors that may support the possibility of unrecognized cryptic species.

The inclusion of ecological, behavioral, or other data related to the biology of cryptic species is compulsory because divergent sequences could be produced from fully interbreeding populations that gained diversity through introgression or ancestral polymorphisms (Nettel et al. 2008, Charlesworth 2010,

Toews and Brelsford 2012). The necessity of requiring additional evidence is exemplified in *Mechanitis* (Lepidoptera: Nymphalidae) where 8 distinct mtDNA haplotypes in 4 nominal species were detected but amplified fragment length polymorphism genotyping supported only one of the four new haplotypes as a distinct species (Dasmahapatra et al. 2009). Lastly, all 4 gene targets should be used in further phylogenetic analyses. Each target has a certain degree of variation ranging from the highly diverse mtDNA region (>10% overall mean diversity) to 5.8S with only a single nucleotide polymorphism (<1% mean diversity) that is a synapomorphy for the *uhleri* complex.

Findings and Implications

This study was the first time the ITS cistron and the CO1 mtDNA region have been evaluated for intra- and interspecific variation in the genus *Amblycorypha*. Furthermore, it is the first time ITS2 secondary structures are described and homology modeled for *Amblycorypha*. The initial goals of this study are limited by failed sequencing runs (47 of the 142 samples sequenced) due to technical laboratory errors. However, preliminary findings are promising and these initial gene targets should provide greater resolution in reconstructing the evolutionary histories of *Amblycorypha* and delimiting potential cryptic species found within the genus. Lastly, intraspecific and interspecific differences in many genes for certain nominal species may represent cryptic species lineages that were previously undetected, which is further investigated in chapter 2.

CHAPTER TWO: THE MOLECULAR PHYLOGENETICS OF GENUS *AMBLYCORYPHA*

Introduction

Molecular Phylogenetics and Species Delimitation

Species are the basic taxonomic unit in many systematic studies in the fields of ecology, evolution, biogeography, and conservation (Whelan et al. 2001). Despite being a fundamental unit of evolutionary biology, species concepts are numerous and far from being agreed on. For example, katydid species have been classified under the mate recognition species concept (Paterson 1985, Bultin and Ritchie 1989) where shared specific mate recognition (e.g., male songs), and/or differences in genital morphology forming species specific fertilization systems, act as pre-zygotic mating barriers providing a means for classifying species (Spooner 1968, Walker et al. 2003, Walker 2004). However, *Amblycorypha alexanderi* and *A. bartrami* were described based on behavioral differences alone (Walker et al. 2003). Mate recognition common garden experiments can gauge how interested males and females are to another population's song type, but may not be able to assess the amount of gene flow within a population since only one migrant into a new population per generation is necessary to prevent divergence between two or more populations (Mills and Allendorf 1996). Furthermore, assessing whether gene flow is occurring between two populations (i.e., by successful mating and producing viable offspring) is not easily measured with ecological experiments, especially in *Amblycorypha* spp., which can lay eggs that remain dormant for 2 years or more before emerging (Drosopoulos and Claridge 2005). Therefore, genetic studies can help overcome experimental difficulties involving ecological manipulations (Hebert et al. 2003, Janzen 2004, Fišer et al. 2018), as well as phylogenetic studies that assess evolutionary independence for each lineage (Mallet 1995, Fujita et al. 2012).

Maximum Likelihood (ML) and Bayesian Inference (BI) are powerful tools used to reconstruct phylogenetic lineages (Felsenstein 1981, Huelsenbeck and Crandall 1997, Huelsenbeck and Ronquist 2001). ML seeks to find the parameter values that maximize the likelihood, given the data (Felsenstein 1981). In a phylogenetic context, ML seeks to find the evolutionary tree which has the highest probability

of fitting the model of evolution in a particular dataset, while BI methods seek to find the most likely set of trees given the data (Huelsenbeck and Ronquist 2001). While sounding similar in their approaches, ML treats the optimality criterion as a constant and does not allow the injection of *a priori* beliefs resulting in a single, most likely phylogenetic tree. In contrast, BI approximates the posterior distribution of a set of parameters relying on the use of *a priori* beliefs to influence the outcome providing a set of most likely trees for the data and set of priors.

The phylogenetic species concept is based upon the idea of reciprocal monophyly, where two clades of a single taxon are more closely related to each other than clades from another taxon. When gene trees are estimated by ML or BI, then it is expected that two clades will coalesce within each taxon before a coalescence event occurs between taxa (Rosen 1979, Mishler 1985, Donoghue 1985, Mallet 1995). However, difficulties in species delimitation under a phylogenetic species concept can occur when interpreting trees reconstructed by BI and ML methods. Tree reconstruction methods may have difficulties reconstructing or representing highly structured populations, recent divergences, incomplete ancestral lineage assortment, and mitochondrial capture from secondary contact. All four scenarios can lead to the reconstructions of a gene tree rather than a species tree, sometimes resulting in improper boundaries being placed around a genetic lineage rather than a species lineage (Pamilo and Nei 1988, Maddison 1997). Moreover, non-parametric bootstrapping and Bayesian posterior probabilities (pp) have trouble providing confidence for rapid and/or recent radiations, along with species groups that have signals for introgression and incomplete lineage sorting in ancestral populations (Alfaro et al. 2003).

Recently, species delimitation has become increasingly rigorous relying on statistical methods for testing alternative models of evolution (Fujita et al. 2012). Multispecies coalescence is a method of species delimitation built on the Kingman coalescent theory (Kingman 1982, Heled and Drummond 2009, Yang and Rannala 2010). Coalescence utilizes the neutral drift theory, which predicts that most of the of variation within a population and between species does not occur from natural selection, rather it occurs from genetic drift. Thus, the method models the time it takes for random genetic drift to cause alleles of a locus to coalesce back to a single common ancestor based on an estimated population size (Futuyma

2005). Multispecies coalescence (MSC) extends this idea to multiple species determining the probability that a locus is independently evolving within populations, indicating a lack of gene flow between populations (Futuyma 2005, Yang and Rannala 2010). These methods incorporate demographic parameters like population size and lineage divergence times to determine the shape and patterns in tree topologies (Allegrucci et al. 2011, Sağlam et al. 2014).

Differences in heritable rates and effective population sizes impact demographic parameters for each locus under an MSC model. For example, mtDNA is expected to have ¼ the effective population size as a nuclear DNA target (Hudson and Turelli 2003), while maternal or paternal sex linked genes have ¾ or ¼ effective population sizes, respectively, when compared to nuDNA (Charlesworth 2009). Molecular phylogenetics has played an important role in comparative analyses involving trait evolution within the Orthoptera and Tettigoniidae (Drosopoulos and Claridge 2005, Snyder et al. 2009, Mugleston et al. 2013). However, few species level phylogenies exist for the subfamily Phaneropterinae (Ullrich et al. 2010, Grzywacz et al. 2014, Kensinger et al. 2017) and no published species level phylogenies using genetic data have yet been developed for *Amblycorypha* species. Previous studies have relied upon morphological and/or acoustical differences as a means for describing new *Amblycorypha* members (Walker 2004, Walker et al. 2003). Thus, a robust species level phylogeny is necessary to understand and test hypotheses regarding the evolution of acoustic communication and complexity in *Amblycorypha* along with delimiting potentially cryptic species that are morphologically indistinguishable from non-conspecifics.

Aim 2

The aim is to reconstruct the phylogenetic relationships within *Amblycorypha* thus providing a tool for studying the evolutionary histories and delimiting species under a MSC model found within this group. The specific aims are 1) to investigate various evolutionary models for my data and, 2) reconstruct hypothetical evolutionary histories of *Amblycorpyha* spp. and, 3) to delimit unknown *Amblycorypha* species and confirm monophyly for nominal *Amblycorypha* members.

Methods

Gene Tree Reconstructions

Three datasets (CO1 only; ITS1+5.8S+ITS2 only; and all genes) were analyzed using ML inference in MEGA7 v7.0.21 (Kumar and Tamura 2016). Nucleotide substitution models were selected based on the lowest Akaike Information Criterion (AIC) corrected measure generated in MEGA7. A General Time Reversible (GTR) model plus 5 discrete Gamma (G) parameters were selected as the best fitting model for CO1 (AIC = 5,707.90), while a K2P+G (5 discrete G parameters) was the best model for the ITS cistron (AIC = 4,005.802). Lastly, since MEGA7 cannot partition different models of evolution for different gene targets in a concatenated dataset, a GTR+G (5 discrete G parameters) plus invariant sites (I) model (I = 41.04% of sites) was used for the concatenated all genes dataset (AIC = 10,264.823). Since it is nearly impossible to exhaustively search all tree combinations in a ML phylogenetic tree to find the most likely tree, a heuristic search can be implemented in order to find a global optimum, i.e., the most likely tree on a particular ML landscape. Initial trees were constructed by NJ algorithms then a ML heuristic method of subtree-pruning-regrafting, which detaches a subtree then "grafts" it elsewhere on the tree and calculates a new likelihood, was implemented. Additionally, 1,000 bootstrap replicates were implemented as a non-parametric test for internodal support on all datasets. The *S. furcata furcata* outgroup was chosen as one of the closest relatives to *Amblycorypha* that had sequences for ITS1, 5.8S, ITS2, and CO1 present in GenBank. Additional outgroups were considered but after multiple searches in GenBank no other Phaneropterine katydids had all 4 gene targets completely sequenced.

Species Delimitation

The coalescent-based software Bayesian Phylogenetics and Phylogeography (BPP) v3.1 (Yang and Rannala 2014) which implements a reversible-jump Markov Chain Monte Carlo (rjMCMC) method (Yang and Rannala 2010) allows for movement between species delimitations when the underlying guide tree is fixed. This method was used to test if candidate species suggested by the all-gene tree were supported under a MSC model. Using rjMCMC sampling, posterior probabilities are determined for

differing models of speciation, i.e., collapsing a node for no speciation or expanding a node for speciation events.

Two key prior assumptions for Bayesian species delimitation include the ancestral population size (θ) and root age (τ). The priors set for either parameter can impact posterior probabilities for species delimitation for speciation models (Yang and Rannala 2010). In this instance, adequate knowledge of prior parameters was lacking. While small populations and recent speciation are hypothesized, a combination of four different parameters were tested [represented as a Gamma distribution $G(\alpha,\beta)$] to simulate either small or large ancestral populations $\theta \sim G(2,2000)$ and $\theta \sim G(1,10)$, respectively, and shallow or deep divergence times as $\tau \sim G(2,2000)$ and $\tau \sim G(1,10)$, respectively following the methods implemented by Leaché and Fujita (2010), along with Maddock et al. (2017) for species delimitation under rjMCMC.

The all-gene tree was fixed and each individual sample coded as a distinct species. Since BPP will not split a species but will always attempt to lump species together, an initial run assuming a small ancestral population and deep divergence times was used to find nodes with well supported posterior probabilities (<0.95). Individuals under well supported nodes were lumped together and tested under the 4 different priors. BPP v3.1 was run twice for each combination of θ and τ priors, using 1,000,000 generations, a burn-in of 200,000 and sampling frequency of 2. Locus and heredity rates were estimated with an initial locus rate generated by a Dirichlet distribution = 2 and initial heredity rate of G(4,4). Internodes that were supported with a pp > 0.99 were considered candidates for species. Additionally, nodes that did not recover monophyletic groups for nominal species (e.g., *A. bartrami*, *A. oblongifolia, A. parvipennis*) were tested under the same four prior parameters but using nearest-neighbor interchange (NNI). rjMCMC under NNI conditions allows exploration of such different lineages combinations. This acts as a check to ensure that tree branches, which guided the rjMCMC algorithm recovered by ML were not improperly reconstructed (e.g., long branch attraction or some other bias) failing to cluster more closely related individuals, like both *A. bartrami* lineages.

Estimated Species Tree

Bayesian species delimitation analysis using *BEAST2 version 2.5 (Ogilvie et al. 2017) was conducted using all sampled loci and guided by supported internodes with pp > 0.99 determined by BPP v3.1. The Markov Chain Monte Carlo (MCMC) algorithm ran for $3.0x10^8$ generations, with sampling every 5,000 generations and a preburn-in of $3.0x10^6$ generations. The dataset was partitioned by locus with ITS1 and ITS2 loci linked using a K2P+G with 3 gamma parameters model of evolution. A GTR+G+I with 3 gamma parameters and 50% of sites invariable was used for CO1 data. 5.8S utilized a K2P+I with 90% of sites considered invariable. Preliminary runs were conducted for detecting any variation in evolutionary rates for all genes. Additionally, strict global clock rates were investigated in MEGA7 by comparing the ML value for a given topology with and without a molecular clock constraint under either a K2P+G model (rDNA targets) or GTR+G (mtDNA target). For nuclear genes, the null hypothesis that equal rates of evolution were occurring throughout the tree failed to be rejected ($p = 0.86$. For mitochondrial genes, the null hypothesis of equal rates of evolution was rejected ($p = 0.03$). However, *BEAST2 preliminary runs using a strict global clock for CO1 allowed for adequate sampling (ESS > 200) for all parameters. Thus, a strict clock model was deemed sufficient for all genes. A Yule-process prior was used for the species tree along with a piecewise linear function and constant-root population-size model (similar methods were implemented by Maddock et al. (2017) for generating a multilocus coalescent species tree). Tracer v1.6 (Rambaut and Drummond 2003) was used for verification of adequate sampling for all parameters (ESS > 200). Posterior probability values were calculated for the majority-rule consensus of samples trees using TreeAnnotator v2.5, part of the BEAST2 package.

Results

Nuclear and Mitochondrial ML Gene Trees

The phylogenetic analyses based on ITS1, 5.8S, and ITS2 rDNA alone provides little resolution for interspecific relationships (Figure 4). Clades identified in the CO1 ML gene tree (Table 5; Figure 5) were largely incongruent with the rDNA ML gene tree. All groups that were identified as clades based on CO1 were recovered as poly- or paraphyletic using rDNA alone and not well supported by bootstrap values (bs < 50 in nearly all cases). Only the *A. longinitca* and *A. arenicola* clades were supported (bs = 72% and 62%, respectively) by rDNA alone. Additionally, the *uhleri* complex was recovered as monophyletic (bs = 99%) with rDNA. Additionally, *A. longinicta* and *A. arenicola* were monophyletic in the CO1 gene tree (bs = 99% and 100%, respectively) indicating that they are distinct clades and have undergone complete lineage assortment. Since a strict clock model was not rejected for rDNA targets (p = 0.86) the incongruence and seemingly random assortment of taxa could be just that, a random assortment of nucleotide polymorphisms that have yet to undergo complete assortment in lineages. Essentially, the rDNA ML tree may be thought of as a SNP tree without the power necessary (a significant number of SNPs) to infer relationships.

The ML all-gene tree (Figure 6) produced a result largely congruent with the CO1 ML tree. All groups of 'unknown' specimens recovered in the CO1 ML tree were recovered in the ML all-gene tree as being closely related and typically with increased bootstrap support. Group A support in the ML CO1 tree (bs = 60%) and the ML all-gene tree (bs = 60%) were equally supported. The 'unknown Rich Mountain' represented a clade within group A and was well supported (bs = 99%, 100%) in both the CO1 and all-gene tree, respectively. Alternatively, four groups – B, C, 1, 2 – were well supported in both CO1 and all-gene trees (all bs ≥ 98%). Groups D, E, and 'nrbartrami' were not well supported, or moderately supported in both the CO1 and all-gene trees with bootstrap values ranging from 19% (group D CO1 ML tree) to 62% ('nrbatrami' ML all-gene tree). Group D was also recovered as paraphyletic to *A. alexanderi* in the CO1 ML tree, but monophyletic in the all-gene tree. Furthermore, the *uhleri* complex was recovered as monophyletic (bs = 99) with both *A. arenicola* (bs = 100%) and *A. longinicta* (bs = 100%)

strongly supported in the all-gene tree. The *rotundifolia* and *oblongifolia* complexes were paraphyletic in all ML trees.

Species Delimitation

Although sampling in this study was limited, BPP species delimitation provided strong support for most candidate species on a fixed guide tree under all four θ and τ priors (Table 6). Overall, the highest probability existed for a 17 species model – the 16 species model supported a *A. floridana* + *A. carinata* clade. However, when subsets of the tree were examined allowing NNI over the guide tree, *A. oblongifolia* and *A. parvipennis* were not well supported by BPP speciation posterior probabilities (Table 7). In contrast, an *A. bartrami* species complex (2 potential species) was supported (pp > 0.90) in 3 of the 4 θ and τ priors; only the large ancestral population G(1,10) and shallow divergence time G(2,2000) showed little support (pp = 0.76). Moreover, *A. bartrami* group 2 was basal to a *A. bartrami* group 1+'nrbartrami' (*A. bartrami* Group 2, (*A. bartrami* Group 1, 'nrbartrami')) in the most likely species tree (pp = 0.66) generated by BPP in an analysis including only these three groups. Finally, the *A.* 'nrbartrami' group was supported under most τ and θ priors with the large ancestral population G(1,10) and shallow divergence time G(2,2000) providing the lowest pp = 0.84 (Table 8).

Species Tree

Under a multispecies coalescent model in *BEAST2, two major clades (1 and 2) were recovered (Figure 7, pp = 1), representing a division between most of the *rotundifolia* complex taxa (excluding *A. parvipennis*) and all the other nominal species. Only the *uhleri* complex was monophyletic while the *rotundifolia* and *oblongifolia* complexes were polyphyletic in relation to *uhleri* and each other. The ML all-gene tree (Figure 6) and MSC species tree (Figure 7) were congruent in their recovery for most taxa. The two phylogenies differed in their placement of *A. alexanderi*, the *A. floridana*+*A. carinata* clade, and *A. parvipennis*. In the all-gene tree, *A. alexanderi* is basal in respect to *A. bartrami* group 2, group D, and group E, while the species tree recovered *A. alexanderi* as the sister taxon to group D. This placement is similar to that recovered in the CO1 ML tree, with *A. alexanderi* being either a derived part of group D, or a sister taxon. *Amblycorypha oblongifolia* and *A. parvipennis* differed in their placement, but this is most

likely due to highly divergent sequences being grouped together. The *uhleri* complex was strongly supported by Bayesian posterior probabilities (pp = 1) in the MSC species tree and bootstrapping (bs = 99%) in the ML all-gene tree. However, bootstrapping and Bayesian posterior probabilities frequently represent confidence or reproducibility of a clade, they do not perform well in simulations where rapid or recent divergences in lineages have occurred even if complete lineage assortment has occurred (Alfaro et al. 2003). Given this, and the essentially random nature of rDNA, it can be assumed that this dataset fits a scenario where rapid lineage assortment has occurred, indicating either highly structured populations or recent speciation events.

Specific Names for 'Unknown' Clades?

The recovery of *A. bartrami* as two distinct groups is supported in the all-gene tree and species tree phylogenies. Additionally, *A. bartrami* groups 1 and 2 along with *A.* 'nrbartami' lineages were allowed to vary in BPP under a rjMCMC species delimitation algorithm (Table 8). BPP supported 3 distinct, independently evolving population structures. However, more information is required before *A. bartrami* can be represented as a species complex. Mitochondrial capture through introgression may explain the inheritance patterns represented in the all-gene tree and species tree, which has been demonstrated in many other taxa as a result of mtDNA falsely indicating cryptic species (e.g., Walberg et al. 2003). Additionally, if *Amblycorypha* is a recently radiated genus, incomplete lineage sorting can confound species tree findings under a multispecies coalescent model, especially if model parameters like population size, divergence times, and model of evolution are not realistic (Knowles and Carstens 2007).

A. floridana and *A. carinata* accessions were reciprocally monophyletic sister taxa in the CO1 and all gene trees, but they were represented together in an internally unresolved clade in the nuDNA phylogeny. BPP species delimitation did not find strong clade support for separating *A. floridana* and *A. carinata*. Rather, under most prior parameters it was supported a single species. However, *BEAST2 multispecies coalescent model species tree and the all-gene tree strongly supported (pp = 1; bs = 100%) two independent clades for *A. carinata* and *A. floridana.* Distribution maps generated from the GrylTett database by T. Walker from 1955 to 2002 (http://entnemdept.ufl.edu/walker/buzz/h00dbase.htm) show

overlapping distributions for the species in northern Florida. Depending on collection locations, introgression between them may still occur in areas with overlapping distributions influencing clade recovery in both gene and species trees (Toews and Brelsford 2012). While introgression between both species is not expected, it was not detected by previous studies (Walker 2004). Ultimately, song type data (Walker 2004), morphological differences (Rehn and Heberd 1914), and clade recovery from the all-gene and species tree, presented here, taken together do support two distinct species.

Initial groups generated from both the rDNA gene tree and mtDNA gene tree were incongruent. Incongruent phylogenies between the rDNA and CO1 datasets suggests possible introgression and incomplete lineage sorting in ancestral populations, or insufficient data. The all-gene tree largely supported clades generated by the CO1 gene tree and separated the *A. alexanderi* and '*bartrami* Group 2 from Clade D (all were combined in CO1). While some disagreement in branch arrangements between the all-gene tree and species tree exist, both are moderately supported with their respective bootstrap and posterior probabilities; uncertainty in the branch arrangements and internodal support may indicate a rapid and/or recent radiation in *Amblycorypha*. Finally, both the gene tree, species tree, and BPP rjMCMC species delimitation support at least 4 additional species within the *rotundifolia* complex. Additional evidence, especially behavioral and/or morphological data, should be examined before these clades can be considered as nominal species.

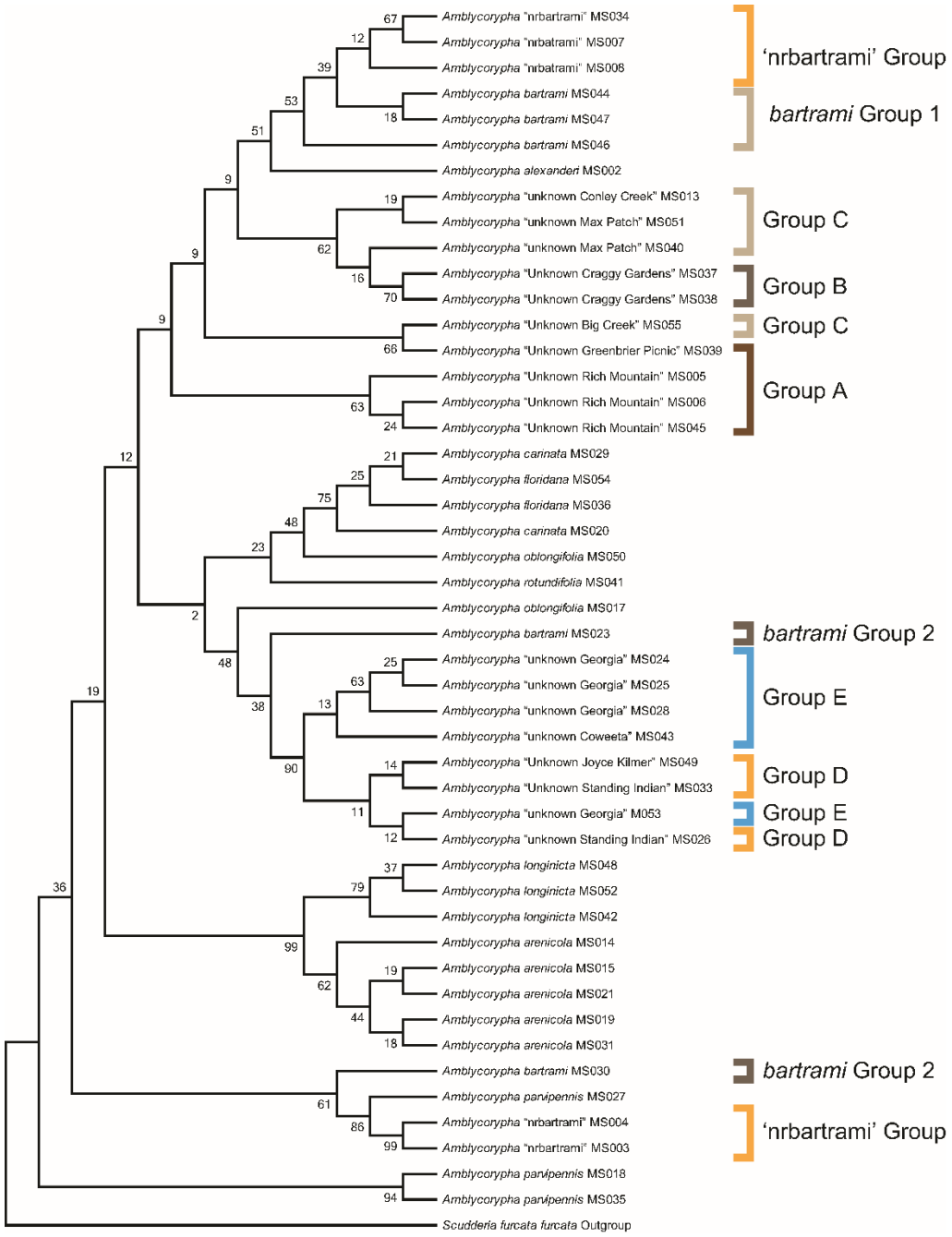

Figure 4. *Amblycorypha* evolutionary histories inferred from ITS1, 5.8S, and ITS2 gDNA using ML and K2P+G model of evolution. The highest log likelihood tree is shown here with numbers next to each tree node representing the percentage of trees in which the associated taxa were recovered after 1,000 bootstrap replicates.

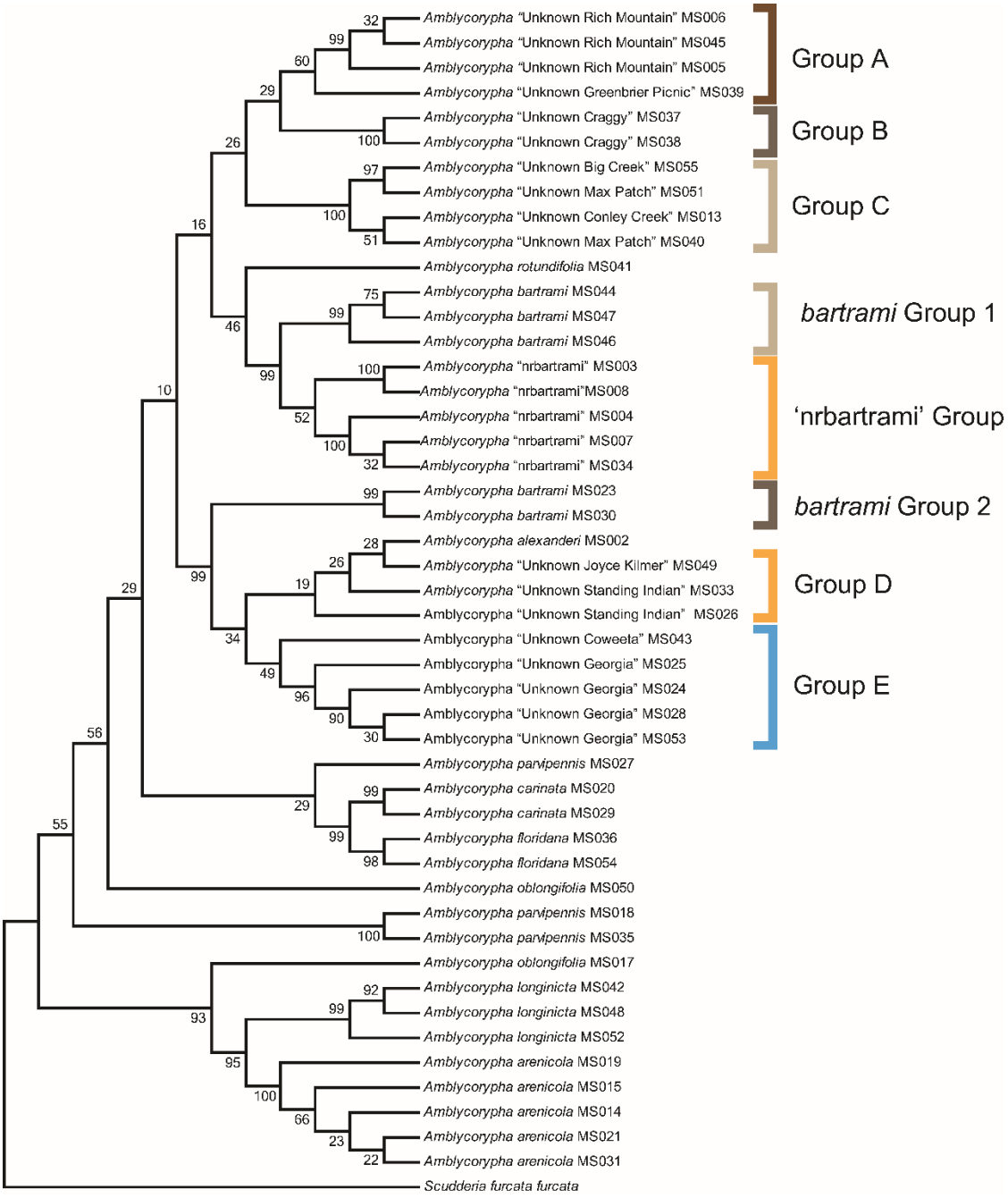

Figure 5. *Amblycorypha* evolutionary histories inferred from CO1 mtDNA using ML analysis and GTR+G model of evolution. The highest log likelihood tree is shown here with numbers next to each tree node representing the percentage of trees in which the associated taxa were recovered after 1,000 bootstrap replicates.

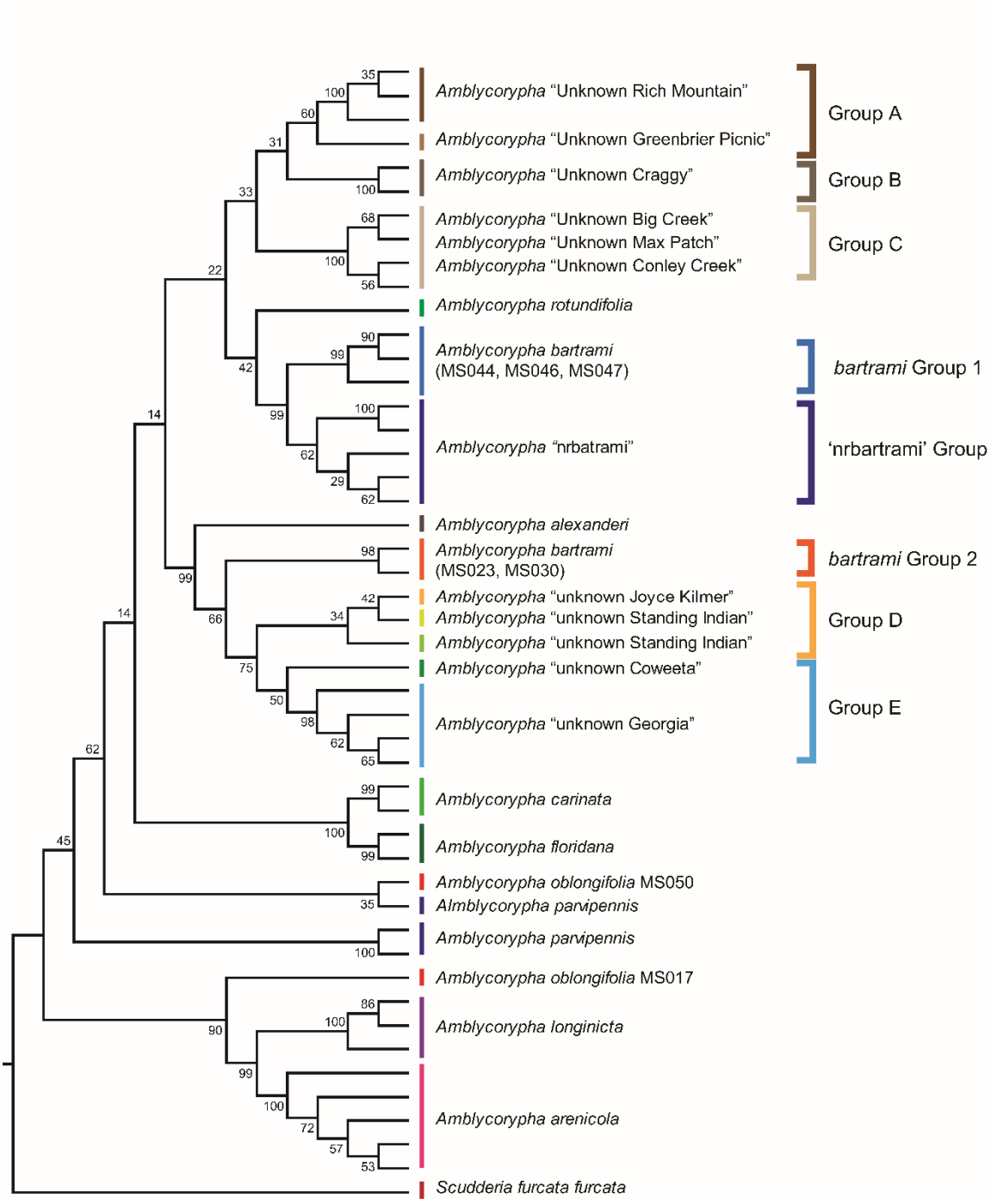

Figure 6. *Amblycorypha* evolutionary history inferred using Maximum Likelihood with a GTR+G+I model of evolution from the all-gene dataset. Here, the highest log likelihood tree is shown with numbers next to each tree node representing the percentage of trees in which the associated taxa clustered together from 1,000 bootstrap replicates.

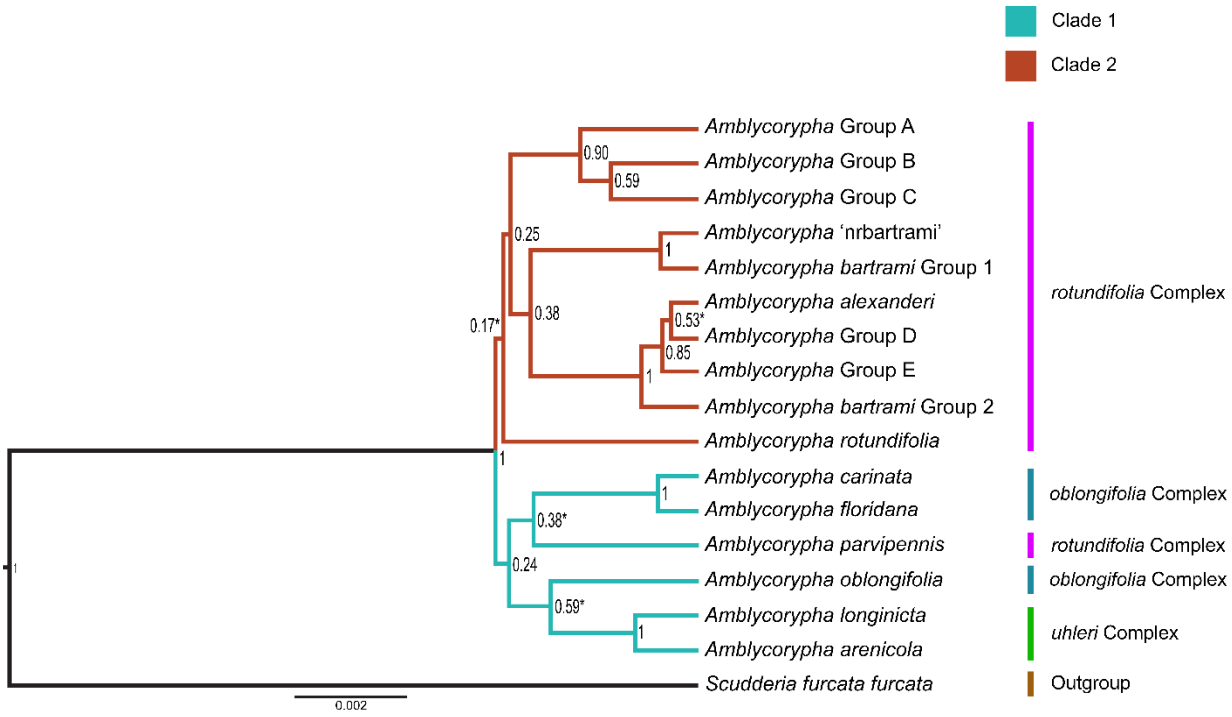

Figure 7. Total evidence tree for *Amblycorypha* evolutionary histories inferred from all 4 gene targets under multispecies coalescence in *BEAST2. Support on branches is BI posterior clade probabilities. A '*' marks where there is disagreement between the all-gene ML phylogeny and multispecies coalescent tree.

Table 5. 'Unknown' groups supported by BPP and *BEAST2.

Group Name	Species	Specimen ID
Group A	'unknown Greenbrier Picnic'	MS039
	'unknown Rich Mountain'	MS005, MS006, MS045
Group B	'unknown Craggy Gardens'	MS037, MS038
Group C	'unknown Big Creek'	MS055
	'unknown Conley Creek'	MS013
	'unknown Max Patch'	MS040, MS051
Group D	'unknown Standing Indian'	MS026, MS033
	'unknown Joyce Kilmer'	MS049
Group E	'unknown Georgia'	MS024, MS025, MS028, MS053
	'unknown Coweeta'	MS033
Group 1	*A. bartrami*	MS044, MS046, MS047
Group 2	*A. bartrami*	MS023, MS030
'nrbartrami'	*A. bartrami*	MS003, MS004, MS007, MS008, MS034

Table 6. Posterior probabilities for rjMCMC species delimitation on a fixed guide tree. Only scores from the first run are shown. Scores less than BPP ≤ 0.01 are not shown.

	Theta prior = 1 10 Tau prior = 1 10 Large Ancestral Population and Deep Divergence Time	**Theta prior = 1 10 Tau prior = 2 2000 Large Ancestral Population and Shallow Divergence Time**	**Theta prior = 2 2000 Tau prior = 1 10 Small Ancestral Population and Deep Divergence Time**	**Theta prior = 2 2000 Tau prior = 2 2000 Small Ancestral Population and Shallow Divergence Time**
P[16]*	0.05	0.05	0.08	0.07
P[17]	0.95	0.95	0.92	0.93

*All instances in which 16 species were recovered supported a (*A. floridana* + *A. carinata*) clade.

Table 7. Posterior probabilities testing the specific status of *A. oblongifolia* and *A. parvipennis* under a rjMCMC speciation model with NNI as a heuristic method on a guide tree. Only scores from the first run are shown. P[x] represents the posterior probability for different numbers of species.

	Theta prior = 1 10 Tau prior = 1 10 Large Ancestral Population and Deep Divergence Time	**Theta prior = 1 10 Tau prior = 2 2000 Large Ancestral Population and Shallow Divergence Time**	**Theta prior = 2 2000 Tau prior = 1 10 Small Ancestral Population and Deep Divergence Time**	**Theta prior = 2 2000 Tau prior = 2 2000 Small Ancestral Population and Shallow Divergence Time**
A	0.58	0.65	0.99	0.82
B	0.55	0.63	0.98	0.81
C	0.64	0.78	0.98	0.87
BC	0.047	0.06	0.01	0.03
ABC	0.30	0.11	0.005	0.08
AB	0.10	0.20	<0.01	0.08
AC	0.01	0.05	<0.01	0.02
P[1]	0.30	0.1	0.005	0.08
P[2]	0.16	0.31	0.02	0.12
P[3]	0.54	0.58	0.98	0.80

A = *A. oblongifolia* MS050

B = *A. oblongifolia* MS017

C = *A. parvipennis* MS018, MS035, MS027

Table 8. Posterior probabilities for the existence of a *Amblycorypha bartrami* complex. Only scores from the first run are shown. BBP under a rjMCMC species delimitation model with NNI as the heuristic method on the guide tree supports 3 distinct species. P[x] represents the posterior probability for different numbers of species.

	Theta prior = 1 10 Tau prior = 1 10 Large Ancestral Population and Deep Divergence Time	**Theta prior = 1 10 Tau prior = 2 2000 Large Ancestral Population and Shallow Divergence Time**	**Theta prior = 2 2000 Tau prior = 1 10 Small Ancestral Population and Deep Divergence Time**	**Theta prior = 2 2000 Tau prior = 2 2000 Small Ancestral Population and Shallow Divergence Time**
A	0.99	0.88	>0.99	>0.99
B	0.92	0.78	0.99	0.99
C	0.92	0.84	0.99	0.99
BC	0.08	0.12	0.01	0.01
ABC	<0.01	0.01	<0.01	---*
AB	<0.01	0.09	<0.01	<0.01
AC	<0.01	0.02	<0.01	<0.01
P[1]	<0.01	0.01	<0.01	---*
P[2]	0.08	0.23	0.01	0.01
P[3]	0.91	0.76	0.99	0.99

A = *A. bartrami* MS023, MS030

B = *A. bartrami* MS044, MS046, MS047

C = *A.* nrbartrami MS003, MS004, MS007, MS008, MS034

*No posterior probabilities were produced during both runs

Discussion

The species tree produced under the MSC model (Figure 7) is the most likely tree representing the evolutionary histories for *Amblycorypha* species. Since an MSC model determines whether a set of genes from a population is independently evolving in comparison to other populations, the reconstructed phylogeny is the best representation of a species tree. The ML all-gene tree and MSC species tree are mostly congruent and indicate agreement in areas where ambiguity was present prior to the species tree reconstruction (e.g., placement of *A. alexanderi* in CO1 and all-gene tree, paraphyly of *A. parvipennis* and *A. oblongifolia*). Thus, this species tree will be referenced as a best supported hypothesis of the evolutionary relationships in the genus *Amblycorypha*.

Species Delimitation Limitations

The species tree meets a unified species concept (USC) criterion that defines species as separately evolving lineages, with each lineage representing a metapopulation (de Queiroz 2007). While this definition lacks predictive capabilities for future isolation – like that of the biological species concept (Wright 1940, Mayr 1942) – it can be tested more easily, especially with genetic data. Based on BPP analyses, at least 16 clades within the ML all-gene tree represent independently evolving metapopulations. However, the rjMCMC species delimitation algorithm implemented in BPP frequently overestimates species by 5 to 13 times more than is actually present when highly structured, disjunct populations are present for a single species (Sukumaran and Knowles 2017). Despite this, population structures are predicted with extreme accuracy (Sukumaran and Knowles 2017). While each lineage meets the USC criterion and each population is highly structured according to BPP, more confidence is needed (e.g., demonstrating reproductive isolation through pre- or postzygotic barriers) before formal names can be provided for each unnamed clade. Alternatively, increased sampling (more individual specimens and more loci) along with developing more realistic ancestral population and divergence time parameters would increase confidence for formally naming each unnamed clade.

Nuclear and Mitochondrial Relationships

Several different scenarios can explain the discordance and lack of confidence displayed between the mtDNA and nuDNA gene trees. Introgression and incomplete lineage sorting in ancestral populations, particularly in rapidly diverging populations, would provide a similar signal as the one reconstructed here. Bayesian posterior probabilities or bootstrapping are expected to give low statistical support for internodes where rapid or recent radiations in lineages have occurred (Alfaro et al. 2003). Alternatively, a lack of interdependence in gene targets and nucleotide sites confounds bootstrapping and posterior probabilities if a model of evolution is not appropriately selected, or is unrealistic for the dataset at hand (Galtier 2004).

However, incomplete lineage sorting is not expected to leave a predictable biogeographic pattern (Funk and Omland 2003). In contrast, hybridization zones due to secondary contact can leave predictable patterns if hybrids are not favored (either through post- or prezygotic mating barriers) for either species but do not typically extend far behind a static hybrid zone (as opposed to a moving hybrid zone where character displacement or niche exclusion is occurring). This type of static hybridization zone is empirically demonstrated between cutthroat and rainbow trout and in New World primates (Cortes Ortiz et al. 2007, Metcalf et al. 2008). A static hybrid zone may exist for *A. floridana* and *A. carinata* in northern Florida/southern Georgia where both species overlap in their distributions. However, Walker (2004) found no evidence of hybridization in the field between both species. Ultimately, a study designed around determining if secondary contact was occurring, and where hybrid zones occur would be necessary to rule out this scenario.

Alternatively, strong selective pressures acting on ITS1/ITS2 or on mtDNA for stabilizing, directional, or diversifying selective pressures can be geographically linked. ITS2 primary sequences typically exhibit close to neutral rates of evolution and nucleotide frequencies (Coleman 2009). Furthermore, ITS2 secondary structures are highly conserved across phaneropterines, indicating strong stabilizing selection within this subfamily. Thus, selective pressures on ITS1 and ITS2 sequences is most likely not a primary cause of diversification patterns exhibited in *Amblycorypha*. The 5.8S SNP found at

position 2 of 175 united the *uhleri* complex as a clade. The *uhleri* clade also is derived within the MSC species tree in relation to the *oblongifolia* and *rotundifolia* complexes. These virtuoso katydids (*uhleri* complex) have complex multisyllable songs that sometimes do not repeat for more than 40 seconds in comparisons to the simpler, shorter songs of other *Amblycorpyha* species. This indicates that signal complexity in the virtuoso katydids is derived and possibly increasing in complexity than that of the simpler songs from other *Amblycorpha* species.

In contrast, mtDNA changes in response to environmental stressors within and between species. First, local environmental conditions can strongly favor a specific mitochondrial variant even in smooth gradient conditions (Irwin 2012). This type of selection was exhibited in sparrows along elevation gradients where nuDNA smoothly flowed between populations but mtDNA was strongly selected for depending on elevation (Cheviron and Brumfield 2009). On the other hand, mtDNA is uniparentally inherited, thus possessing a fourfold smaller effective population size (Hudson and Turelli 2003). This means that mtDNA will complete a lineage sorting process faster than any single nuDNA gene since the rate is inversely proportional to the effective population size (Funk and Omland 2003). Thus, mtDNA is more likely than single nuDNA targets to accurately reflect recent divergences. Given that the ITS1 and ITS2 are thought to be evolutionarily linked (Torres et al. 1990), mtDNA should have greater resolution at predicting recent divergences. Since all clades recovered from the CO1 mtDNA gene tree were recovered in the ML all-gene tree and species tree it can be hypothesized that complete lineage separation has occurred in most metapopulations, but nuDNA has not completely differentiated independently within each *Amblycorypha* clade. A link between song-type, fitness, and each mtDNA haplotype through a common garden transplant experiment could provide strong evidence for either allopatry or sympatry being the primary driver in gene flow restriction between each clade, or whether allopatry and sympatry are both working in concert and thus limiting gene flow.

Finally, biogeographical conclusions that may be drawn from this study are limited. However, the two major clades recovered in the MSC species tree (pp = 1) have all members of the *rotundifolia* complex, excluding *A. parvipennis*, present on the eastern side of the Mississippi river and *A. parvipennis*

on the western side. Additionally, the *A. longinicta* populations from Arkansas and Mississippi had greater intraspecific distances between each gene target than other species indicating that these allopatrically separated populations may be independently evolving. If these biogeographical patterns are related to Quaternary glaciation events, specifically the Laurentide ice sheet formation which is responsible for the present course of the Mississippi river (Dyke and Prest 1987), then two refugia on the eastern and western sides of the Mississippi river may have led to the formation of two major clades within *Amblycorypha*. From what has been discovered so far, most of the *Amblycorypha* species diversity exists in two locations – central Texas (8 taxa) and western North Carolina (5 nominal taxa plus 4 groups described here) – both of which acted as refugia during the Quaternary glacial epoch (Hewitt 2000). Lastly, coastal plains habitat may have provided refuge for other *Amblycorypha* species distributed throughout the southeastern United States (*A floridana*, *A. arenicola*, *A. longinicta,* and *A. cajuni*) and occasionally up the eastern seaboard (*A. carinata*). The inclusion of other *uhleri* complex members and southwestern *Amblycorypha* species (*A. insolita, A. rivograndis*, and *A. huasteca*) would help provide a more complete story regarding the biogeography of *Amblycorypha*.

Implications and Future Directions

This study presents one of the first species level phylogenies built on nucleotide sequence data for the genus *Amblycorypha*. The recovery of 5 distinct clades containing 'unknown' species (which have different song types than that of other morphologically similar individuals (Tim Forrest, pers. comm.)) provides additional evidence that unknown lineages exist within the *rotundifolia* complex. However, the question of precise evolutionary relationships and whether each lineage is independently evolving remain unresolved due to incomplete taxon sampling and low sample sizes. Developing a robust molecular phylogeny for *Amblycorypha* will require additional taxon sampling and additional analysis of nuDNA and mtDNA loci. For instance, including additional gene targets, especially nuDNA targets that have a faster rate of evolution than ITS1/ITS2, or targets that act as indicators of sexual selection, and sex linked genes may provide resolution in lineage assortment for each population.

Sampling from locations in which the holotype for each species was described – or sequencing from the holotype itself – may help resolve current (as evidenced by two distinct *A. bartrami* clades) and future issues regarding evolutionary histories for this taxon. Additionally, sequencing 5 nominal species that were not included in this study; *A. uhleri*, *A. cajuni*, *A. rivograndis*, *A. insolita*, and *A. huasteca* would provide clarity and may help resolve the poly- and paraphyletic relationships found within the *oblongifolia* and *rotundifolia* complexes. Especially since the latter 3 species have widespread mid- and southwestern distributions. They are particularly famous for their sky island biogeography, unique patch dynamics, and occurrence on the Edwards plateau (which is known for being a refugium during previous glaciation events). Investigating *Amblycorypha* populations in areas where suitable habitat is fragmented may lead to the discovery of additional cryptic species. Notably, *A. insolita* has been collected from many disjunct sky islands throughout the southwestern US in Big Bend National Park (TX), Mt. Lemmon (AZ), Mt. Mica (AZ), Pinal Peak (AZ), Guadalupe Mountains National Park (TX/NM), and certain locations along the Rio Grande river.

REFERENCES

Alexander RD. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). *Annals of the Entomological Society of America.* 50(6): 584-602.

Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution.* 20(2): 255-266.

Allegrucci G, Trucchi E, Sbordoni V. 2011. Tempo and mode of species diversification in *Dolichopoda* cave crickets (Orthoptera, Rhaphidophoridae). *Molecular Phylogenetics and Evolution*. 60(1): 108-121.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215(3): 403-410.

Ankenbrand MJ, Kelle, A, Wolf M, Schultz J, Förster F. 2015. ITS2 database V: Twice as much. *Molecular Biology and Evolution*. 32(11): 3030-3032.

Bailey WJ, Withers PC, Endersby M, Gaull K. 1993. The energetic costs of calling in the bushcricket *Requena verticalis* (Orthoptera: Tettigoniidae: *Listroscelidinae*). *Journal of Experimental Biology*. 178(1): 21-37.

Bailey WJ. 1995. Costs of calling in Tettigoniid Orthoptera: a case study of *Requena verticalis* (Tettigoniidae: Listroscelidinae). *Journal of Orthoptera Research*. 65-73.

Bailey, WJ, Field G. 2000. Acoustic satellite behaviour in the Australian bushcricket *Elephantodeta nobilis* (Phaneropterinae, Tettigoniidae, Orthoptera). *Animal Behaviour.* 59(2): 361-369.

Ballard JWO. 2000. Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Evolution.* 51(1): 64-75.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*. 13(4): 729-744.

Bargues MD, Marcilla A, Ramsey JM, Dujardin JP, Schofield CJ, Mas-Coma S. 2000. Nuclear rDNA-based molecular clock of the evolution of Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Memórias do Instituto Oswaldo Cruz*. 95(4): 567-573.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambuat A, Drummond, A. J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology.* 10(4): e1003537.

Brown SD, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH. 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*. 12(3): 562-565.

Butlin RK, Ritchie MG. 1989. Genetic coupling in mate recognition systems: what is the evidence?. *Biological Journal of the Linnean Society*. 37(3): 237-246.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*. 10(3): 195-205.

Charlesworth D. 2010. Don't forget the ancestral polymorphisms. *Heredity.* 105: 509-510.

Cheviron ZA, Brumfield RT. 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*. 63(6): 1593–1605.

Coleman AW. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research*. 35(10): 3322-3329.

Coleman AW. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution*. 50(1): 197-203.

Coluzzi M, Sabatini A, Della Torre A, Di Deco MA, Petrarca V. 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science*. 298(5597): 1415-1418.

Cortés-Ortiz L, Duda TF, Canales-Espinosa D, García-Orduña F, Rodríguez-Luna E, Bermingham E. 2007. Hybridization in large-bodied New World primates. *Genetics.* 176(4): 2421-2425.

Culumber ZW, Shepard DB, Coleman SW, Rosenthal GG, Tobler M. 2012. Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (*Teleostei: Xiphophorus*). *Journal of evolutionary biology*. 25(9): 1800-1814.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. *25*(15): 1974.

Dasmahapatra, KK, Elias M, Hill RI, Hoffman JI, Mallet J. 2010. Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*. 10(2): 264-273.

De Geer C. 1773. Mémoires pour servir a l'histoire des Insectes. 3: i-viii, 1-696. *Hasselberg, Stockholm*.

de la Cruz, J, Karbstein, K, Woolford Jr, JL. 2015. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annual Review of Biochemistry*. 84: 93.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology*. *56*(6): 879-886.

Donoghue MJ. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*. 172-181.

Drosopoulos S, Claridge MF. 2005. *Insect sounds and communication: physiology, behaviour, ecology, and evolution*. CRC Press.

Dyke A, Prest V. 1987. Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie Physique et Quaternaire*. 41(2): 237-263.

Eickbush TH, Eickbush DG. 2007. Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics*. 175(2): 2477-85

Felsenstein, J. (1981). A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biological Journal of the Linnean Society*. 16(3): 183-196.

Fišer C, Robinson CT, Malard F. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*. 27(3): 613-635.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for the amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*. 3(5): 294-299.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*. 27(9): 480-488.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics. 34*(1): 397-423.

Futuyma DJ. 2005. Evolution. Sinauer & Associates. *Inc., Sunderland, Massachusetts*. 226-243.

Galtier N. 2004. Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites. *Systematic Biology. 53*(1): 38-46.

Gerber AS, Loggins R, Kumar S, Dowling TE. 2001. Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes?. *Annual Review of Genetics.* 35(1): 539-566.

Grzywacz B, Heller KG, Lehmann AW, Warchałowska Śliwa E, Lehmann GU. 2014. Chromosomal diversification in the flightless Western Mediterranean bushcricket genus *Odontura* (Orthoptera: Tettigoniidae: Phaneropterinae) inferred from molecular data. *Journal of Zoological Systematics and Evolutionary Research*. 52(2): 109-118.

Hammond TJ, Bailey WJ. 2003. Eavesdropping and defensive auditory masking in an Australian bushcricket*, Caedicia* (Phaneropterinae: Tettigoniidae: Orthoptera). *Behaviour.* 140(1): 79-95.

Hebert PD, Ratnasingham S, de Waard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*. 270(Suppl 1): S96-S99.

Hebert PD, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S, Jannetta P, Jannetta P. 2016. Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B.* 371(1702): 20150333.

Heled J, Drummond AJ. 2009. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution. 27(3*): 570-580.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*. 405(6789): 907.

Hubbell TH. 1954. Entomology: The Naming of Geographically Variant Populations. *Systematic Zoology*. 3(3): 113-121.

Hubbell TH. 1956. Some aspects of geographic variation in insects. *Annual Review of Entomology*. 1(1): 71-88.

Hudson RR, Turelli M. 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*. 57(1): 182-190.

Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*. 28(1): 437-466.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8): 754-755.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*. 294(5550): 2310-2314.

Hurst GD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London B: Biological Sciences.* 272(1572): 1525-1534.

Illumina® Inc. 2016. 16S metagenomic sequencing library preparation. [accessed 2018 May 27]. Available from: https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html

Irwin DE. 2012. Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *The American Naturalist.* 180(1): 35-49.

Jiggins FM. 2003. Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics.* 164(1): 5-12.

Keller I, Alexander JM, Holderegger R, Edwards PJ. 2013. Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology.* 26(12): 2527-2543.

Kensinger BJ, Schwemm MR, Luttbeg B. 2017. Molecular Phylogeny for the *Obolopteryx* Katydids of the Southwestern United States (Orthoptera: Tettigoniidae: Phaneropterinae). *Journal of the Entomological Research Society*. 19(3): 7-14.

Kingman JFC. 1982. The coalescent. *Stochastic Processes and their Applications*. 13(3): 235-248.

Knowles LL, Carstens BC. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology.* 56(6): 887-895.

Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz, J. 2009. The ITS2 Database III—sequences and structures for phylogeny. *Nucleic Acids Research*. *38*(suppl_1): D275-D279.

Koetschan C, Hackl T, Müller T, Wolf M, Förster F, Schultz J. 2012. ITS2 database IV: interactive taxon sampling for internal transcribed spacer 2 based phylogenies. *Molecular Phylogenetics and Evolution*. 63(3): 585-588.

Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 23(10): 1289-1291.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution.* 33(7):1870-1874.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin, F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23(21): 2947-2948.

Isely FB. 1941. Researches concerning Texas Tettigoniidae. *Ecological Monographs*. 11(4): 457-475.

Leaché AD, Banbury BL, Felsenstein J, De Oca ANM, Stamatakis A. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*. 64(6): 1032-1047.

Leaché AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B: Biological Sciences.* rspb20100662.

Legendre F, Robillard T, Song H, Whiting MF, Desutter-Grandcolas, LAURE. 2010. One hundred years of instability in ensiferan relationships. *Systematic Entomology*. 35(3): 475-488.

Li C, Wilkerson RC. 2005. Identification of *Anopheles* (*Nyssorhynchus*) *albitarsis* complex species (Diptera: Culicidae) using rDNA internal transcribed spacer 2-based polymerase chain reaction primes. *Memorias do Instituto Oswaldo Cruz*. 100(5): 495-500.

Li Y, Zhou XIN, Feng G, Hu H, Niu L, Hebert PD, Huang D. 2010. COI and ITS2 sequences delimit species, reveal cryptic taxa and host specificity of fig-associated *Sycophila* (Hymenoptera, Eurytomidae). *Molecular Ecology Resources*. 10(1): 31-40.

Maddison WP. 1997. Gene trees in species trees. *Systematic biology*. 46(3): 523-536.

Maddock ST, Childerstone A, Fry BG, Williams DJ, Barlow A, Wüster W. 2017. Multi-locus phylogeny and species delimitation of Australo-Papuan blacksnakes (*Pseudechis Wagler*, 1830: Elapidae: Serpentes). *Molecular Phylogenetics and Evolution*. 107*:* 48-55.

Mallet J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution*. 10(7): 294-299.

Mayr E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*. *66*(2): 526-538.

Metcalf JL, Siegle MR, Martin AP. 2008. Hybridization dynamics between Colorado's native cutthroat trout and introduced rainbow trout. *Journal of Heredity*. 99(2): 149-156.

McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, Orange DI, Lemos-Espinal J, Riddle BR, Jaeger JR. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution*. 61(12): 2879-2897.

Mills LS, Allendorf FW. 1996. The one-migrant-per-generation rule in conservation and management. *Conservation Biology.* 10(6): 1509-1518.

Mishler BD. 1985. The morphological, developmental, and phylogenetic basis of species concepts in bryophytes. *Bryologist.* 207-214.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, ... Niehuis O. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 346(6210): 763-767.

Mugleston JD, Song H, Whiting MF. 2013. A century of paraphyly: A molecular phylogeny of katydids (Orthoptera: Tettigoniidae) supports multiple origins of leaf-like wings. *Molecular Phylogenetics and Evolution*. 69(3): 1120-1134.

Mullineux T, Hausner G. 2009. Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera Grosmannia and Leptographium. *Fungal Genetics and Biology*. 46(11): 855-867.

Nazar RN. 2003. Ribosome biogenesis in yeast: rRNA processing and quality control. *Applied Mycology and Biotechnology*. 3: 161-185.

Nazar R. 2004. Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB life*. 56(8):457-465.

Nettel A, Dodd RS, Afzal-Rafii Z, Tovilla-Hernández C. 2008. Genetic diversity enhanced by ancient introgression and secondary contact in East Pacific black mangroves. *Molecular Ecology*. 17(11): 2680-2690.

Nichols R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution*. 16(7): 358-364.

Nosil P. 2008. Speciation with gene flow could be common. *Molecular ecology*. *17*(9): 2103-2106.

Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution.* 34(8): 2101-2114.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution*. 5(5): 568-583.

Paskewitz SM, Wesson DM, Collins FH. 1994. The internal transcribed spacers of ribosomal DNA in five members of the *Anopheles gambiae* species complex. *Insect Molecular Biology*. 2(4): 247-257.

Paterson HEH. 1985. The recognition concept of species. *Transvaal Museum Monographs*. 4: 21-29.

Perdeck AC. 1958. The isolating value of specific song patterns in two sibling species of grasshoppers (*Chorthippus brunneus* Thunb. and *C. biguttulus* L.). *Behaviour.* 12(1): 1-75.

Rambaut A, Drummond A. 2003. Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. *University of Edinburgh, UK*.

Rehn JA, Hebard M. 1904. The Orthoptera of Thomas County, Georgia, and Leon County, Florida. *Proceedings of the Academy of Natural Sciences of Philadelphia*. 774-802.

Rehn JA, Hebard M. 1914. Studies in American Tettigoniidae (Orthoptera): I and II. *Transactions of the American Entomological Society (1890-)*. *40*(4): 271-344.

Rosen DE. 1979. Fishes from the uplands and intermontane basins of Guatemala: revisionary studies and comparative geography. Bulletin of the AMNH. v. 162: article 5.

Sağlam İK, Küçükyıldırım S, Çağlar SS. 2014. Diversification of montane species via elevation shifts: the case of the Kaçkar cricket *Phonochorion* (Orthoptera). *Journal of Zoological Systematics and Evolutionary Research*. 52(3): 177-189.

Saussure HD. 1859. Orthoptera nova-Americana (diagnoses præliminares).

Scobie M. 2013. A phylogenetic Analysis of the Round-headed katydids (Tettigoniidae: *Amblycorypha*). *University of North Carolina at Asheville Undergraduate Research Journal*.

Scudder SH. 1862. *Materials for a monograph of the North American Orthoptera*. HO Houghton.

Selig C, Wolf M, Müller T, Dandekar T, Schultz J. 2007. The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Research*. *36*(suppl_1): D377-D380.

Shoemaker DD, Katju V, Jaenike J. 1999. *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. *Evolution.* 53(4): 1157-1164.

Shoemaker DD, Dyer KA, Ahrens M, McAbee K, Jaenike J. 2004. Decreased diversity but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection. *Genetics*. 168(4): 2049-2058.

Soghigian J, Andreadis TG, Livdahl TP. 2017. From ground pools to treeholes: convergent evolution of habitat and phenotype in *Aedes* mosquitoes. *BMC Evolutionary Biology*: *17*(1), 262.

Song N, Cai W, Li H. 2017. Deep-level phylogeny of Cicadomorpha inferred from mitochondrial genomes sequenced by NGS. *Scientific Reports. 7*(1): 10429.

Spooner JD. 1968. Pair-forming acoustic systems of phaneropterine katydids (Orthoptera, Tettigoniidae). *Animal Behaviour.* 16(2-3): 197-212.

Spooner JD. 1995. Pair-forming phonotaxic strategies of phaneropterine katydids (Tettigoniidae: Phaneropterinae). *Journal of Orthoptera Research*. 127-129.

Stål C. 1876. Observations orthopterologiques 2. Sur Anostostoma et quelques genres voisins. Konglingar Svenska vetenskapsakademien Handlingar. 4(5).

Sukumaran J, Knowles LL. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences.* 114(7): 1607-1612.

Snyder RL, Frederick-Hudson KH, Schul J. 2009. Molecular phylogenetics of the genus *Neoconocephalus* (Orthoptera, Tettigoniidae) and the evolution of temperate life histories. *PloS one*. 4(9): e7203.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123(3): 585-595.

Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*. 21(16): 3907-3930.

Torres RA, Ganal M, Hemleben V. 1990. GC balance in the internal transcribed spacers ITS 1 and ITS 2 of nuclear ribosomal RNA genes. *Journal of Molecular Evolution*. *30*(2): 170-181.

Tuckerman JF, Gwynne DT, Morris GK. 1993. Reliable acoustic cues for female mate preference in a katydid (*Scudderia curvicauda*, Orthoptera: Tettigoniidae). *Behavioral Ecology.* 4(2): 106-113.

Ullrich B, Reinhold K, Niehuis O, Misof B. 2010. Secondary structure and phylogenetic analysis of the internal transcribed spacers 1 and 2 of bush crickets (Orthoptera: Tettigoniidae: Barbitistini). *Journal of Zoological Systematics and Evolutionary Research.* 48(3): 219-228.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research*. 40(15): e115-e115.

Vilgalys R, Hester M. 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology*. 172(8): 4238-4246.

Voelkerding KV, Dames SA, Durtschi JD. 2009. Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*. 55(4): 641-658.

Wahlberg N, Oliveira R, Scott JA. 2003. Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera: Nymphalidae): complex mtDNA variation and species delimitations. *Systematic Entomology*. 28(2): 257-274.

Walker TJ, Forrest TG, Spooner JD. 2003. The *rotundifolia* complex of the genus *Amblycorypha* (Orthoptera: Tettigoniidae): Songs reveal new species. *Annals of the Entomological Society of America*. 96(4): 433-447.

Walker TJ. 2004. The *uhleri* group of the genus *Amblycorypha* (Orthoptera: Tettigoniidae): extraordinarily complex songs and new species. *Journal of Orthoptera Research.* 13(2): 169-183.

Whelan S, Liò P, Goldman N. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. *TRENDS in Genetics.* 17(5): 262-272.

White TJ, Bruns T, Lee SJWT, Taylor JL. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: A Guide to Methods and Applications*. 1*8*(1): 315-322.

Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T. 2005. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA.* 11(11): 1616-1623.

Wolf M, Chen S, Song J, Ankenbrand M, Müller T. 2013. Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences–a proof of concept. *PloS one*. 8(6): e66726.

Wright S. 1940. Breeding structure of populations in relation to speciation. *The American Naturalist*: *74*(752): 232-248.

Yang, Z, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences.* 107(20): 9264-9269.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*. 13(5): 303-314.

Yang Z, Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution.* 31(12): 3125-3135.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. *31*(13): 3406-3415.

APPENDIX

Table A1. Collection information for specimens sequenced in this study.

Specimen ID	Genus	Species	Complex	Collection Date	County	State
MS002	*Amblycorypha*	*alexanderi*	*rotundifolia*	4-Jun-2007	Liberty	FL
MS014	*Amblycorypha*	*arenicola*	*uhleri*	19-Jul-2007	Richmond	NC
MS015	*Amblycorypha*	*arenicola*	*uhleri*	19-Jul-2007	Richmond	NC
MS019	*Amblycorypha*	*arenicola*	*uhleri*	19-Jul-2007	Richmond	NC
MS021	*Amblycorypha*	*arenicola*	*uhleri*	19-Jul-2007	Richmond	NC
MS031	*Amblycorypha*	*arenicola*	*uhleri*	19-Jul-2007	Richmond	NC
MS023	*Amblycorypha*	*bartrami*	*rotundifolia*	3-Jun-2007	Cleburne	AL
MS030	*Amblycorypha*	*bartrami*	*rotundifolia*	2-Jun-2007	Cleburne	AL
MS044	*Amblycorypha*	*bartrami*	*rotundifolia*	1-Jul-2006	Richmond	NC
MS046	*Amblycorypha*	*bartrami*	*rotundifolia*	1-Jul-2006	Richmond	NC
MS047	*Amblycorypha*	*bartrami*	*rotundifolia*	1-Jul-2006	Richmond	NC
MS020	*Amblycorypha*	*carinata*	*oblongifolia*	1-Jun-2007	Gordon	GA
MS029	*Amblycorypha*	*carinata*	*oblongifolia*	1-Jun-2007	Gordon	GA
MS036	*Amblycorypha*	*floridana*	*oblongifolia*	20-May-2011	Franklin	FL
MS054	*Amblycorypha*	*floridana*	*oblongifolia*	10-Jul-2006	Polk	FL
MS042	*Amblycorypha*	*longinicta*	*uhleri*	14-Jul-2006	Lafayette	MS
MS048	*Amblycorypha*	*longinicta*	*uhleri*	14-Jul-2006	Lafayette	MS
MS052	*Amblycorypha*	*longinicta*	*uhleri*	15-Jul-2006	Faulkner	AR
MS003	*Amblycorypha*	nr bartrami	*rotundifolia*	5-Jun-2009	Georgetown	SC
MS004	*Amblycorypha*	nr bartrami	*rotundifolia*	5-Jun-2009	Georgetown	SC
MS007	*Amblycorypha*	nr bartrami	*rotundifolia*	5-Jun-2009	Georgetown	SC
MS008	*Amblycorypha*	nr bartrami	*rotundifolia*	5-Jun-2009	Georgetown	SC
MS034	*Amblycorypha*	nr bartrami	*rotundifolia*	5-Jun-2009	Georgetown	SC
MS017	*Amblycorypha*	*oblongifolia*	*oblongifolia*	14-Jul-2006	Stone	AR
MS050	*Amblycorypha*	*oblongifolia*	*oblongifolia*	12-Jul-2006	Boone	MO
MS018	*Amblycorypha*	*parvipennis*	*rotundifolia*	7-Jul-2007	Faulkner	AR
MS027	*Amblycorypha*	*parvipennis*	*rotundifolia*	8-Jul-2007	Shannon	MO
MS035	*Amblycorypha*	*parvipennis*	*rotundifolia*	7-Jul-2007	Faulkner	AR
MS041	*Amblycorypha*	*rotundifolia*	*rotundifolia*	16-Sep-2006	Sevier	TN
MS032	*Atlanticus*	sp.	-	-	-	-
MS055	*Amblycorypha*	unknown Big Creek	*rotundifolia*	16-Sep-2006	Haywood	NC
MS013	*Amblycorypha*	unknown Conley Creek	*rotundifolia*	17-Jul-2008	Jackson	NC
MS043	*Amblycorypha*	unknown Coweeta	*rotundifolia*	3-Jul-2006	Macon	NC
MS037	*Amblycorypha*	unknown Craggy	*rotundifolia*	14-Sep-2006	Haywood	NC
MS038	*Amblycorypha*	unknown Craggy	*rotundifolia*	14-Sep-2006	Haywood	NC
MS024	*Amblycorypha*	unknown GA	*rotundifolia*	1-Jun-2007	Gordon	GA
MS025	*Amblycorypha*	unknown GA	*rotundifolia*	1-Jun-2007	Gordon	GA
MS028	*Amblycorypha*	unknown GA	*rotundifolia*	1-Jun-2007	Gordon	GA
MS053	*Amblycorypha*	unknown GA	*rotundifolia*	5-Jul-2006	Gordon	GA

Table A1 (Cont.). Collection information for specimens sequenced in this study.

Specimen ID	Genus	Species	Complex	Collection Date	County	State
MS039	*Amblycorypha*	unk. Greenbrier Picnic	*rotundifolia*	16-Sep-2006	Sevier	TN
MS049	*Amblycorypha*	unknown Joyce Kilmer	*rotundifolia*	23-Sep-2006	Graham	NC
MS040	*Amblycorypha*	unknown Max Patch	*rotundifolia*	10-Oct-2006	Madison	NC
MS051	*Amblycorypha*	unknown Max Patch	*rotundifolia*	10-Oct-2006	Madison	NC
MS005	*Amblycorypha*	unknown Rich Mountain	*rotundifolia*	7-Jul-2011	Madison	NC
MS006	*Amblycorypha*	unknown Rich Mountain	*rotundifolia*	7-Jul-2011	Madison	NC
MS045	*Amblycorypha*	unknown Rich Mountain	*rotundifolia*	10-Sep-2006	Madison	NC
MS026	*Amblycorypha*	unknown Standing Indian	*rotundifolia*	31-Jul-2008	Macon	NC
MS033	*Amblycorypha*	unknown Standing Indian	*rotundifolia*	31-Jul-2008	Macon	NC

Table A1 (Cont.). Collection information for specimens sequenced in this study.

Specimen ID	Latitude	Longitude	Collector	Determiner	Location
MS002	30.45695	84.98039	TG Forrest	TG Forrest	UNCA Insect Collection
MS014	35.05473	79.61107	TG Forrest	TG Forrest	UNCA Insect Collection
MS015	35.05473	79.61107	TG Forrest	TG Forrest	UNCA Insect Collection
MS019	35.05473	79.61107	TG Forrest	TG Forrest	UNCA Insect Collection
MS021	35.05473	79.61107	TG Forrest	TG Forrest	UNCA Insect Collection
MS031	35.05473	79.61107	TG Forrest	TG Forrest	UNCA Insect Collection
MS023	33.70266	85.59664	TG Forrest	TG Forrest	UNCA Insect Collection
MS030	33.78012	85.52666	TG Forrest	TG Forrest	UNCA Insect Collection
MS044	35.05284	79.60346	TG Forrest	TG Forrest	UNCA Insect Collection
MS046	35.05284	79.60346	TG Forrest	TG Forrest	UNCA Insect Collection
MS047	35.05284	79.60346	TG Forrest	TG Forrest	UNCA Insect Collection
MS020	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection
MS029	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection
MS036	29.72128	84.74510	TG Forrest	TG Forrest	UNCA Insect Collection
MS054	28.03390	82.02120	TG Forrest	TG Forrest	UNCA Insect Collection
MS042	34.43700	89.63936	TG Forrest	TG Forrest	UNCA Insect Collection
MS048	34.43700	89.63936	TG Forrest	TG Forrest	UNCA Insect Collection
MS052	-	-	TGF, DD	TG Forrest	UNCA Insect Collection
MS003	33.34800	79.22710	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS004	33.34800	79.22710	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS007	33.34800	79.22710	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS008	33.36090	79.22660	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS034	33.34800	79.22710	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS017	-	-	TGF, JAH	TG Forrest	UNCA Insect Collection
MS050	-	-	JAH	TG Forrest	UNCA Insect Collection
MS018	35.14181	92.45361	TG Forrest	TG Forrest	UNCA Insect Collection
MS027	37.14245	91.11211	TG Forrest	TG Forrest	UNCA Insect Collection
MS035	35.14181	92.45361	TG Forrest	TG Forrest	UNCA Insect Collection
MS041	35.73150	83.40790	TG Forrest	TG Forrest	UNCA Insect Collection
MS032	-	-	TG Forrest	TG Forrest	UNCA Insect Collection
MS055	35.75116	83.10973	TG Forrest	TG Forrest	UNCA Insect Collection
MS013	35.35937	83.33942	TG Forrest	TG Forrest	UNCA Insect Collection
MS043	35.06209	83.42120	D Cusick	TG Forrest	UNCA Insect Collection
MS037	35.58615	83.07366	TG Forrest	TG Forrest	UNCA Insect Collection
MS038	35.58615	83.07366	TG Forrest	TG Forrest	UNCA Insect Collection
MS024	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection
MS025	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection
MS028	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection
MS053	34.47749	84.74900	TG Forrest	TG Forrest	UNCA Insect Collection

Table A1 (Cont.). Collection information for specimens sequenced in this study.

Specimen ID	Latitude	Longitude	Collector	Determiner	Location
MS039	35.71191	83.38380	TGF, LD Block	TG Forrest	UNCA Insect Collection
MS049	35.39560	83.91490	TG Forrest	TG Forrest	UNCA Insect Collection
MS040	35.79680	82.96190	TG Forrest	TG Forrest	UNCA Insect Collection
MS051	35.79680	82.96190	TG Forrest	TG Forrest	UNCA Insect Collection
MS005	35.92710	82.80630	TG Forrest	TG Forrest	UNCA Insect Collection
MS006	35.92710	82.80630	TG Forrest	TG Forrest	UNCA Insect Collection
MS045	35.92875	82.77928	TG Forrest	TG Forrest	UNCA Insect Collection
MS026	35.09402	83.52210	TG Forrest	TG Forrest	UNCA Insect Collection
MS033	35.09402	83.52210	TG Forrest	TG Forrest	UNCA Insect Collection

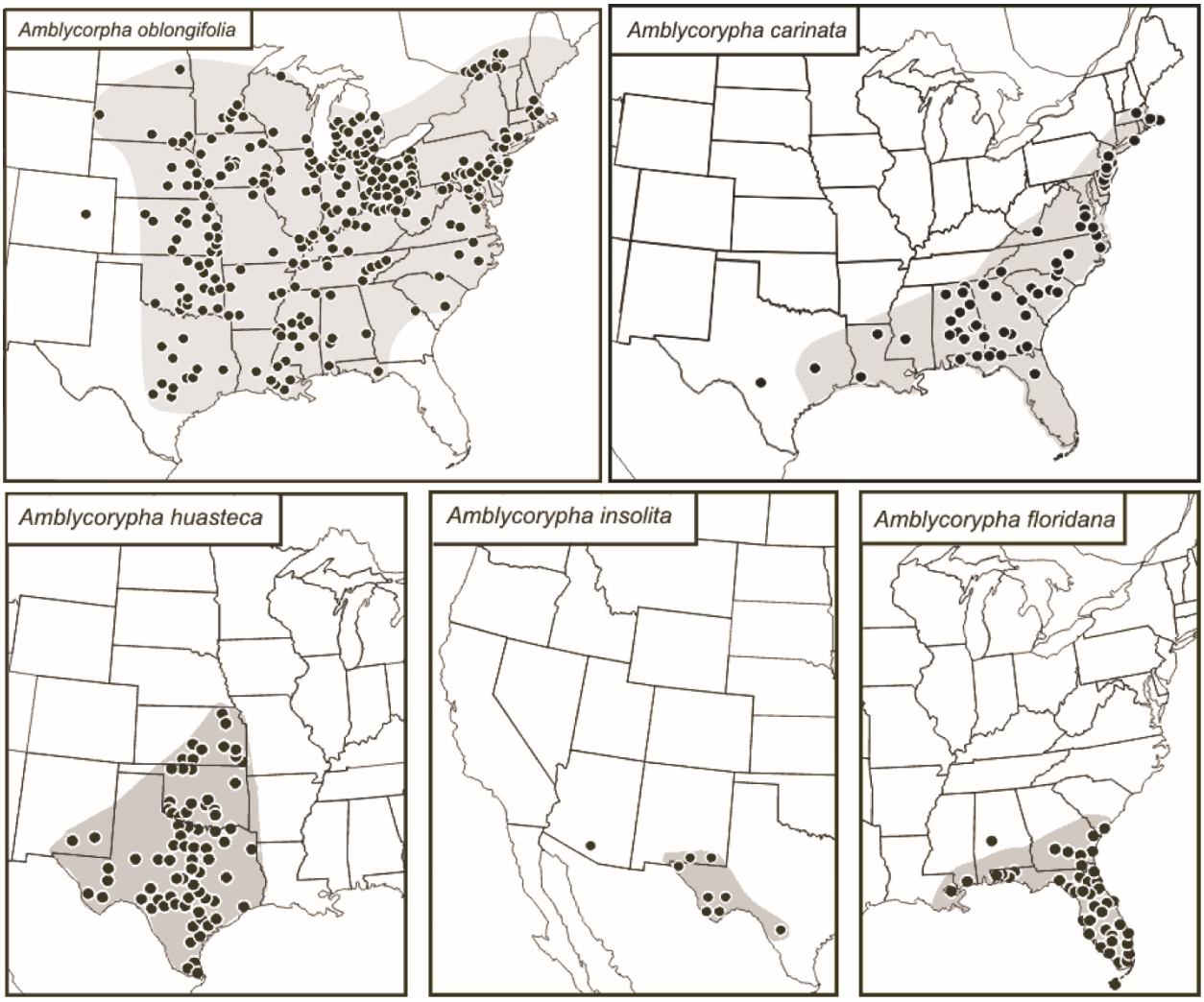

Figure B1. Distributions of *oblongifolia* complex members (adapted from http://entnemdept.ufl.edu/walker/buzz/katylist.htm#phaneropterinae *Amblycorypha* species pages)

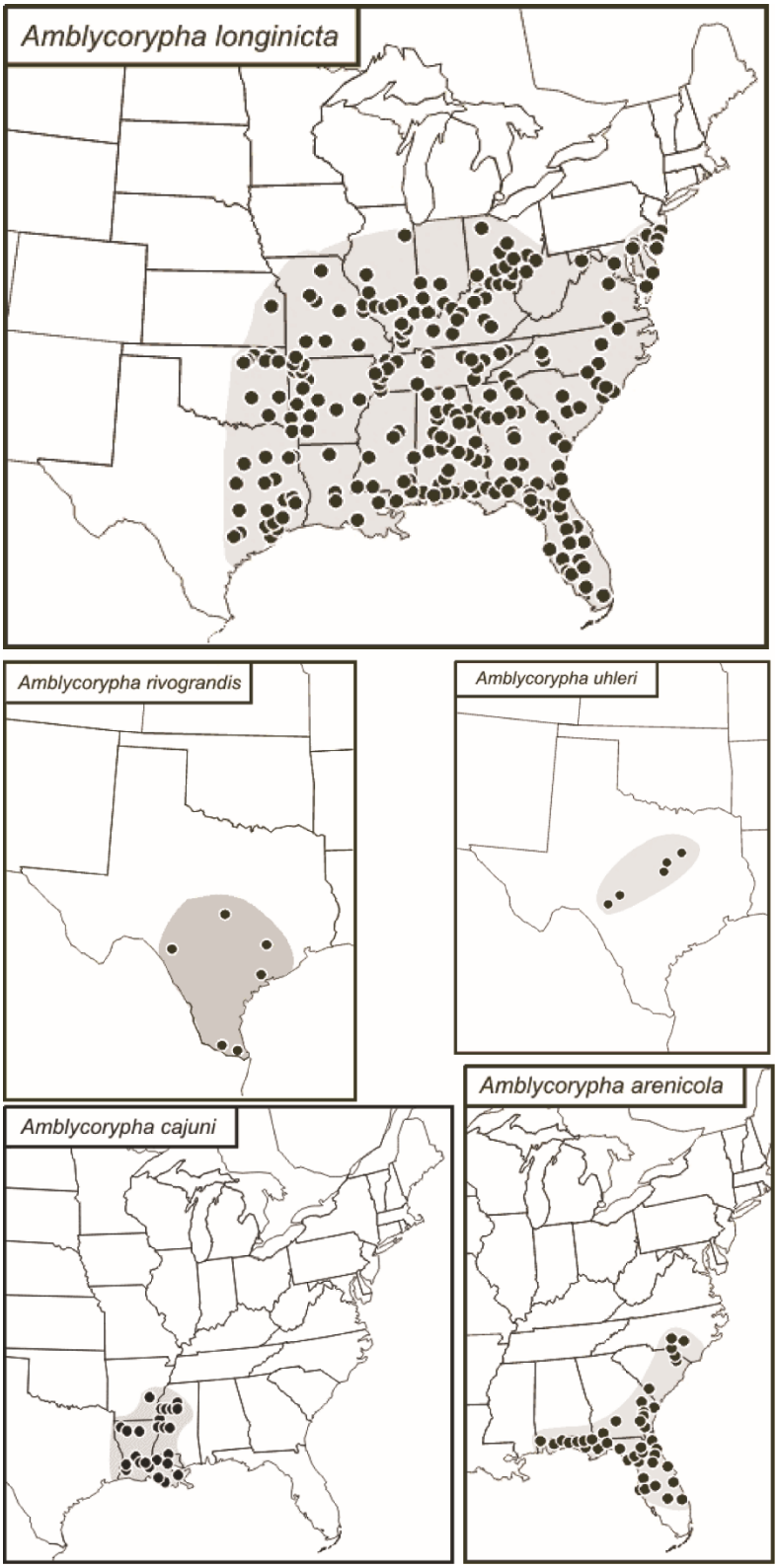

Figure B2. The *uhleri* complex geographic distributions adapted from Walker (2004).

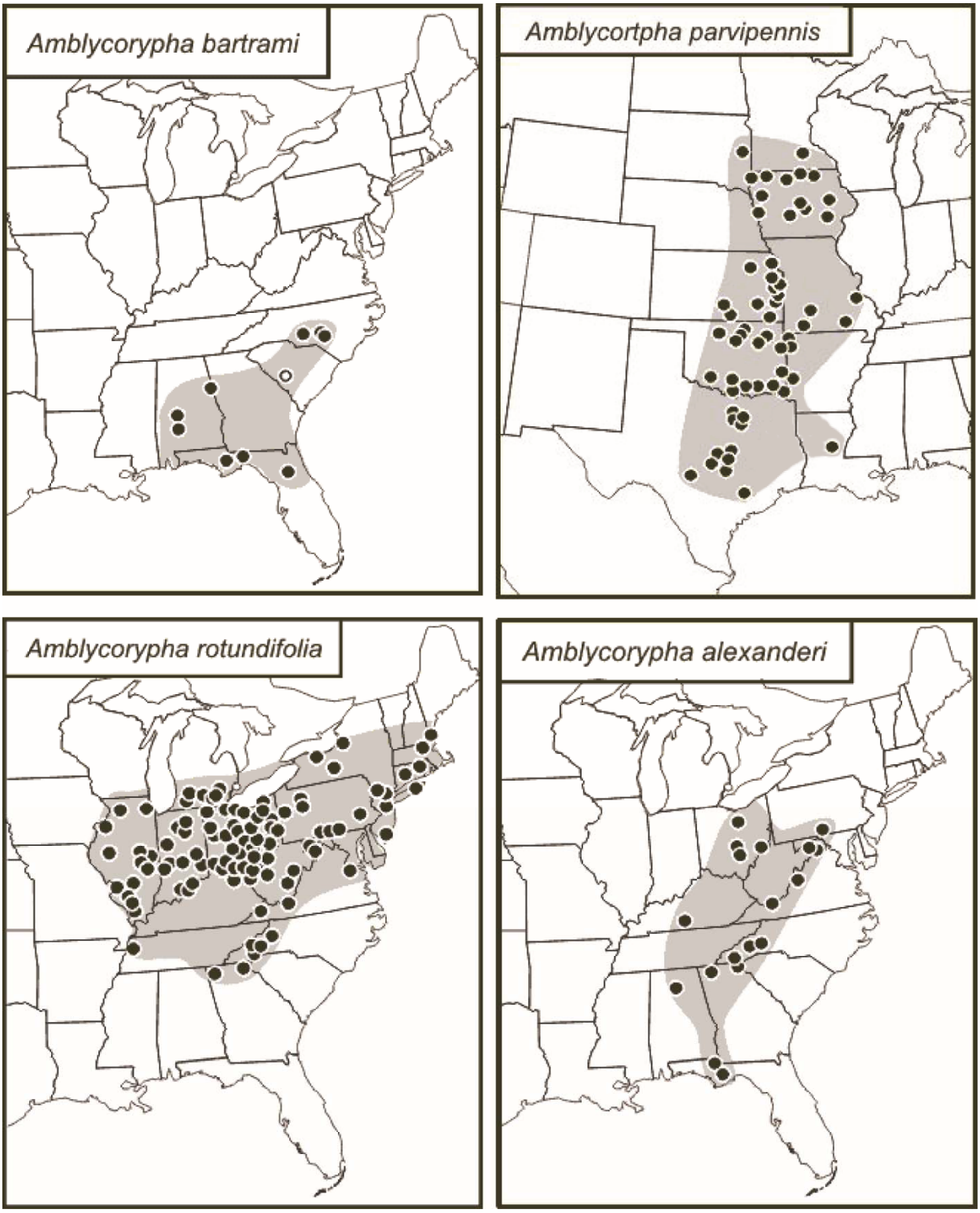

Figure B3. Geographic distributions of *rotundifolia* complex members adapted from Walker et al. (2003).

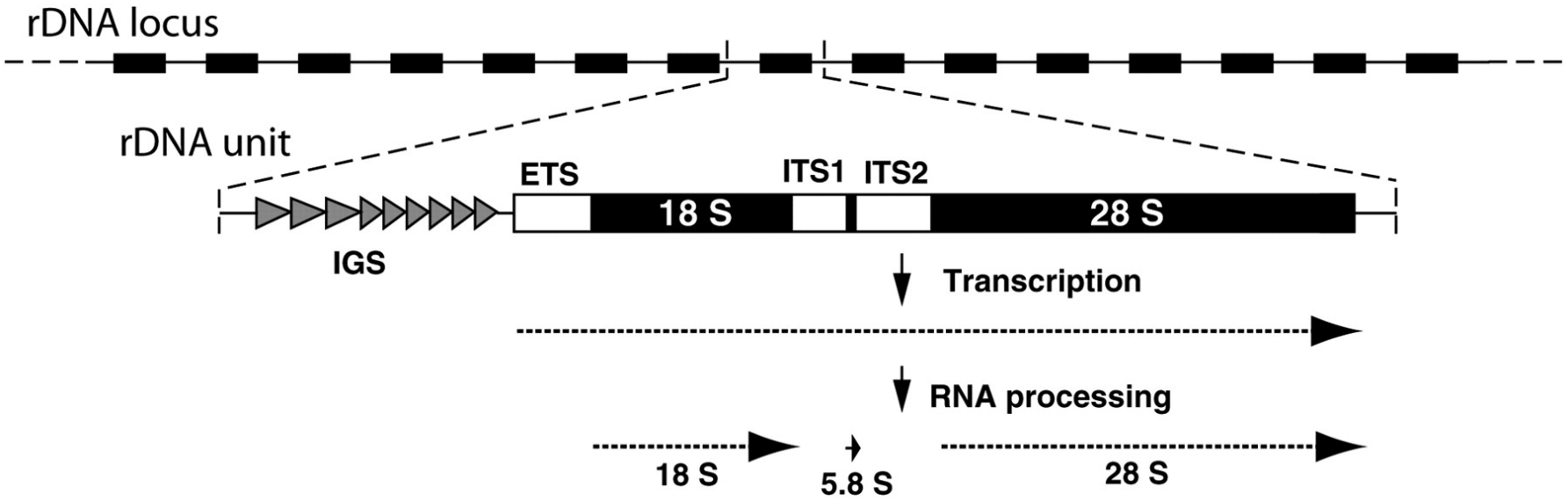

Figure B4. The tandem repeating rDNA region. Each repeat contains the external transcribed spacer (ETS), 18S, 5.8S, and 28S (part of the ribosome structure) and the internal transcribed spacers 1 and 2 (ITS). The ITS and ETS regions have functional secondary structures that allow for chaperone proteins to bind and promote maturation of the ribosomal subunits. Image Credit: Eickbush TH, Eickbush DG. 2007. Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics*. 175(2): 2477-85.