Molecular phylogenetics, phylogenomics, and phylogeography

Phylogenomics of the Leaf-Footed Bug Subfamily Coreinae (Hemiptera: Coreidae)

Michael Forthman,1,2,4 Christine W. Miller,1 and Rebecca T. Kimball3

1Entomology & Nematology Department, University of Florida, 1881 Natural Area Drive, Gainesville, FL 32611, 2California Department of Food and Agriculture, Plant Pest Diagnostics Branch, 3294 Meadowview Road, Sacramento, CA 95832, 3Department of Biology, University of Florida, 876 Newell Drive, Gainesville, FL 32611, and 4Corresponding author, e-mail: michael.forthman@cdfa.ca.gov

Subject Editor: Kevin Johnson

Received 18 February, 2020; Editorial decision 29 June, 2020

Abstract

Baits targeting invertebrate ultraconserved elements (UCEs) are becoming more common for phylogenetic studies. Recent studies have shown that invertebrate UCEs typically encode proteins—and thus, are functionally different from more conserved vertebrate UCEs—and can resolve deep divergences (e.g., superorder to family ranks). However, whether invertebrate UCE baits have the power to robustly resolve relationships at shallower phylogenetic scales has been generally limited to investigations within the Coleoptera and Hymenoptera; thus, there are many invertebrate UCE baits that remain to be tested at shallower levels (i.e., tribes and congeners). Here, we assessed the ability of a recently designed Hemiptera UCE bait set to reconstruct more recent phylogenetic relationships in the largest leaf-footed bug subfamily, the Coreinae (Hemiptera: Coreidae), using a taxon-rich sample representing 21 of the 32 coreine tribes. Many well-supported, novel relationships were congruent in maximum likelihood and summary coalescent analyses. We also found evidence for the para- and polyphyly of several tribes and genera of Coreinae, as well as the subfamilies Coreinae and Meropachyinae. Our study, along with other recent UCE studies, provides evidence that UCEs can produce robust and novel phylogenetic hypotheses at various scales in invertebrates.

Key words: phylogenomics, ultraconserved elements, Coreidae, Coreinae, target capture

Next-generation sequencing (NGS) technology has made the generation of large numbers of orthologous loci throughout the genome achievable for many nonmodel organisms. For molecular phylogenetics, one of the advantages of such data is the potential to resolve challenging nodes in the Tree of Life across various temporal scales (e.g., Faircloth et al. 2012, Lemmon et al. 2012, Li et al. 2013). Ultraconserved elements (UCEs) are one such class of loci that can be obtained using target capture approaches and NGS (see Faircloth et al. 2012). These loci have been widely used in phylogenetic estimation since Faircloth et al. (2012) introduced their utility in anchoring loci for phylogenomic analysis in vertebrates.

Although the use of UCEs in invertebrates and vertebrates is conceptually similar, invertebrate bait sets have been designed to target genomic regions with more liberal thresholds of conservation across taxa (Faircloth 2017), often requiring multiple baits for the same region to maximize capture of divergent taxa. Furthermore, recent empirical tests of UCE baits have shown that these loci are primarily protein-coding in invertebrates (Boswell and Danforth 2018, Kieran et al. 2019), making invertebrate UCEs fundamentally different from those found in vertebrates, which are believed to be primarily noncoding regulators (Bejerano et al. 2004, Sandelin et al. 2004, Woolfe et al. 2004, Pennacchio et al. 2006). Regardless, invertebrate UCE bait sets have been shown to resolve deep divergences (e.g., superorder to family ranks) in several taxonomic groups (e.g., Baca et al. 2017, Starrett et al. 2017, Kieran et al. 2019). Only a few invertebrate studies, primarily within Coleoptera and Hymenoptera, have demonstrated the utility of these primarily protein-coding UCEs at more shallow evolutionary timescales (i.e., subfamily to congeners) (e.g., Van Dam et al. 2017, Hedin et al. 2018, Boswell et al. 2019, Branstetter and Longino 2019). However, several new invertebrate UCE bait sets are available for other taxonomic groups (see Faircloth 2017), which remain to be tested in this regard. Recently, Kieran et al. (2019) empirically tested a UCE bait set for the insect order Hemiptera (Faircloth 2017), showing its utility in resolving phylogenetic relationships among the suborders and families with a small sample of taxa. The utility of UCEs in hemipteran phylogenetics has also been demonstrated for interfamilial and intersubfamilial relationships in the hemipteran superfamily Coreoidea (Forthman et al. 2019). However, the ability of UCEs to robustly resolve phylogenetic relationships among tribes and congeners within hemipteran lineages has not yet been evaluated.
Leaf-footed bugs, or Coreidae (Hemiptera: Heteroptera), are a charismatic group of phytophagous insects (Fig. 1) that also includes several pests of agricultural systems (see Gentry 1965, Nonveiller 1984, Mitchell 2000). With 2,571 extant species described in four subfamilies and 37 tribes, this is the largest family of the Coreoidea (CoreoideaSF Team 2019). The worldwide Coreinae is by far the largest coreid subfamily with 2,320 (90%) species (372 genera, 32 tribes). Some of the largest, stoutest terrestrial heteropterans are members of this subfamily (e.g., species of Pachylis Leplete & Servelle, Thasus Stål, and Petacelis Signoret; Schuh & Slater 1995, Fernandes et al. 2015), but body forms also vary from sticklike (e.g., Tylocryptus Horváth, Priomotylus Fieber) to leaflike (e.g., species of the tribe Phyllomorphini). While many species are dull in appearance, some are brightly colored and iridescent (e.g., Petalops Amyot & Servelle, Diactor Perty, Phthisiadena Brailovsky). The hind legs of males in many species are known to be sexually selected weapons that exhibit variation in size, shape, and armature (Eberhard 1998, Emlen 2008, Okada et al. 2011, Procter et al. 2012). Fighting behaviors are also variable across species that exhibit male–male competition; e.g., some species grapple end-to-end (e.g., Narnia femorata Stål; Nolen et al. 2017), while others kick, flip, and squeeze one another face-to-face (e.g., Mictis profana [Fabricius]; Tatarnic and Spence 2013). Aside from their diverse morphology, Coreinae are well known for their odious alarm pheromones (Aldrich and Blum 1978, Leal and Kadosawa 1992), paternal care in Phyllomorpha Laporte (e.g., Garcia-González et al. 2003), and gregariousness in nymphs (e.g., Aldrich and Blum 1978, Flanagan 1994, Miyatake 1995). Furthermore, species of the genus Holbymenia Leplete & Servelle superficially appear to be wasp mimics (Pereira et al. 2013). Given these captivating morphological and behavioral features, the subfamily Coreinae offers an excellent opportunity to investigate the evolution of various traits and their possible correlates. However, a well-resolved, taxon-rich phylogeny of Coreinae is lacking. Thus, a comprehensive, robust phylogeny of the group is first needed before evolutionary questions can be investigated.

Although a few molecular phylogenetic analyses have focused on Coreidae (Fang and Nie 2007; Pan et al. 2007, 2008), the most comprehensive (with respect to taxon sampling) investigation of coreine phylogeny comes from Li’s (1997) morphological phylogenetic analysis of the family Coreidae. Li (1997) analyzed 35 species (23 tribes) and 74 characters using parsimony to investigate phylogenetic relationships among subfamilies and tribes within the family Coreidae. Li (1997) recovered a paraphyletic Coreinae with respect to Meropachyinae, which supported an earlier, similar analysis on the superfamaly Coreoidea (Li 1996); this result has been corroborated by more recent molecular UCE studies (Kieran et al. 2019, Forthman et al. 2019). Li’s (1996, 1997) analyses further corroborated by more recent molecular UCE studies (Kieran et al. 2019, Forthman et al. 2019). Lastly, we explored the suitability of UCEs at shallower scales by including multiple species within several genera whose limits have been generally uncontroversial.

Materials and Methods

Taxon Sampling

In total, 124 taxa were sampled for this study, including 104 species of Coreinae from 21 tribes. For 25 of our taxa, we obtained contigs from Kieran et al. (2019) and Forthman et al. (2019). Our sampling primarily targeted freshly preserved material, with 16 samples taken from degraded museum specimens. We generated new UCE data (Supp Tables S1 and S2 [online only]) following the protocols discussed below.

DNA Extraction

For all new data, genomic DNA was extracted using a 1) Gentra Puregene Tissue, 2) Qiagen DNeasy Blood and Tissue kit (hereafter DNeasy), or 3) Qiagen DNeasy Blood and Tissue kit coupled with Qiagen QIAquick PCR purification kit (hereafter DNQIA; see Knyshov et al. 2019). Depending on the size of specimens and to use similar amounts of tissue across species, where possible, we used any part of the body (legs, abdomen, thorax, head) or the entire body from ethanol-preserved, silica-bead preserved, frozen, or dried specimens (Supp Table S1 [online only]). Freshly preserved specimens were extracted with either the Puregene or DNeasy kits (Supp Table S1 [online only]). For the Puregene Tissue Kit, we followed the manufacturer’s protocol for 5–10 mg tissue, including optional recommendations, with the following modifications: 10 µl of proteinase K was added to samples and incubated for 24–48 h; 600 µl of 100% ethanol was used for the first wash and centrifuged for 10 min to reduce the loss of DNA pellets; and 30–100 µl of molecular grade water or Puregene DNA Hydration Solution was used to resuspend isolated DNA. We also followed the manufacturer’s protocol for the DNeasy kit, but with fewer modifications: tissue was incubated in 180–190 µl Buffer ATL and 10–20 µl proteinase K for 24–48 h (200 µl total solution volume), and DNA eluted once or twice with 50 µl Buffer AE depending on the size of the specimens we extracted from.

Degraded museum specimens were extract with either the DNeasy protocol described above or a modified version of it that follows Knyshov et al. (2019) (i.e., DNQIA [Supp Table S1 [online only]]. The latter protocol is specifically designed to extract degraded DNA >100 bp in length. Briefly, the DNQIA protocol follows the DNeasy kit up to the first centrifugation, using a QIAquick spin
column. The samples were then subjected to the Qiagen QIAquick PCR purification kit, replacing AW1 and AW2 washes with PE buffer. Under the DNQIA protocol, samples were eluted in 30 µl EB buffer. We initially attempted to extract genomic DNA from six specimens using the DNeasy protocol (collected 1946–2016) and 10 others using the DNQIA protocol (collected 1935–2015). Two museum samples initially subjected to the DNeasy protocol did not yield extracts of sufficient concentration nor produce visual bands. Museum samples initially subjected to the DNeasy protocol did not produce visual bands. Two specimens using the DNeasy protocol (collected 1946–2016) and 10 others using the DNQIA protocol (collected 1935–2015). The samples were then subjected to the Qiagen QIAquick PCR purification kit, replacing AW1 and AW2 washes with PE buffer. Under the DNQIA protocol, samples were eluted in 30 µl EB buffer. We initially attempted to extract genomic DNA from six specimens using the DNeasy protocol (collected 1946–2016) and 10 others using the DNQIA protocol (collected 1935–2015). Two museum samples initially subjected to the DNeasy protocol did not yield extracts of sufficient concentration nor produce visual bands.

Focused-ultrasonicator (20–60 s).

µ normalized to 10–20 ng/l, and high molecular weight samples and quantified with a Qubit 2.0 fluorometer. Samples were then yield extracts of sufficient concentration nor produce visual bands after gel electrophoresis (see below for details on methods); these samples were successfully reextracted using the DNQIA protocol. DNA quality was assessed with 1% agarose gel electrophoresis and quantified with a Qubit 2.0 fluorometer. Samples were then normalized to 10–20 ng/l, and high molecular weight samples and quantified with a Qubit 2.0 fluorometer. Samples were then yield extracts of sufficient concentration nor produce visual bands after gel electrophoresis (see below for details on methods); these samples were successfully reextracted using the DNQIA protocol. DNA quality was assessed with 1% agarose gel electrophoresis and quantified with a Qubit 2.0 fluorometer. Samples were then normalized to 10–20 ng/l, and high molecular weight samples and quantified with a Qubit 2.0 fluorometer. Samples were then yield extracts of sufficient concentration nor produce visual bands after gel electrophoresis (see below for details on methods); these samples were successfully reextracted using the DNQIA protocol.

Library Construction, Target Enrichment, and Sequencing

We constructed libraries using the modified KAPA Hyper Prep Kit protocol of Forthman et al. (2019). Briefly, half volume reactions were used for all steps, as well as iTru universal adapter stubs and 8-bp dual indexes (Glenn et al. 2016). Library amplification involved initial denaturation at 98°C for 3 min, followed by 14 cycles of 98°C for 30 s, 60°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 5 min. Amplified libraries were inspected by gel electrophoresis, quantified with Qubit, combined into 1,000 ng pools using equimolar amounts, dried at 60°C, and resuspended in 1 µl IDTE.

For target enrichment, we used a custom myBaits kit, based on Faircloth’s (2017) probe set that was subsampled by Forthman et al. (2019). For some samples, we followed Forthman et al.’s (2019) enrichment protocol while others were subjected to a modified version of the protocol. We refer to the protocols as TE and TE-touchdown, respectively. For our TE protocol, a hybridization mixture with half volume of baits (2.75 µl) and 2.75 µl molecular-grade water was hybridized with each library pool at 65°C for 16–24 h. In the TE-touchdown protocol, baits were to hybridize with library pools at 65°C for 18 h followed by 18 h at 62°C. However, samples were prematurely removed during the TE-touchdown protocol before hybridization was complete. As a result of this, we consulted with Arbor Biosciences (Ann Arbor, MI) and followed their recommendation to add an additional 2.75 µl baits to these samples and rerun the entire hybridization protocol to completion; additional baits were added given that the initial beads were subjected to the 95°C denaturation step, which should limit their effectiveness during hybridization per Arbor Biosciences.

Dynabeads M-280 Streptavidin beads were then bound to bait-target hybrids, washed four times at 65°C (TE) or 62°C (TE-touchdown) and resuspended in 30 µl IDTE. For the postcapture PCR amplification mix, 2.5 µl each of 5 µM iTru P5/P7 primers (Glenn et al. 2016) were added. We performed 14–17 cycles of postcapture amplification following manufacturer’s protocol, except we used an annealing temperature of 65°C (TE) or 62°C (TE-touchdown) and an extension period of 45 s. Hydrophobic Sera-Mag SpeedBeads Carboxyl Magnetic Beads were used for postamplification cleanup, followed by two washes in freshly prepared 70% ethanol and resuspension in 22 µl IDTE. We quantified enriched library pools with Qubit, pooled all library pools in equimolar amounts, and sequenced on a single Illumina HiSeq3000 lane (2 × 100) at the University of Florida’s Interdisciplinary Center for Biotechnology Research (ICBR).

Phylogenetic Estimation

For each of the data sets, we concatenated single locus alignments in PHYLUCE and then selected the best-fit partition scheme and models of sequence evolution using PartitionFinder v2.1.1 (Lanfear et al. 2017). We used the rcluster algorithm with unlinked branch lengths and treated individual loci as separate data blocks. All models under the ‘raxml’ option were examined (Stamatakis 2006), and the best-fit models were selected using the corrected Akaike Information Criterion (AICc) (Harvich and Tsai 1989). Twenty partitioned maximum likelihood (ML) optimal searches were conducted in RAxML v8.2.10 (Stamatakis 2014) using random starting trees. Bootstrap

Sequence Data Processing and Alignment

Sequence data were processed following Forthman et al. (2019). Briefly, sequence reads were demultiplexed at the sequencing facility, and adapters were trimmed with illumiprocessor (Faircloth 2013, Bolger et al. 2014). Duplicate reads were filtered using PRINSEQ-lite v0.20.4 (Schmieder and Edwards 2011), and the remaining reads were corrected with QuoUM v1.1.0 (Maçar et al. 2015) and de novo assembled in Trinity (Grabherr et al. 2011). We identified and aligned UCE loci from our assembled contigs using PHYLUCe v1.5.0 (Faircloth 2016). Internal trimming of alignments was done with trimAl using the heuristic ‘automated1’ method, with all other settings at default (Capella-Gutiérrez et al. 2009). Locus alignments with at least 50% and 70% of taxa (hereafter 50p and 70p, respectively) were retained for phylogenetic inferences. We also subsampled each data set by just including the 25% most parsimony-informative loci—based on differences in raw counts—to explore the effects of this filtering strategy on phylogenetic inferences; the use of the 25% most parsimony-informative loci have been shown to improve or recover similar topological support compared to estimates based on more uninformative or informative gene trees, respectively (Hosner et al. 2016, Meiklejohn et al. 2016).

In many models of sequence evolution, guanine-cytosine (GC) content is assumed to be constant across a data set. Deviation from this assumption can cause phylogenetic estimation error, and the negative effects of GC content and heterogeneity have been demonstrated for concatenation and species tree approaches (e.g., Lockhart et al. 1994, Galtier and Gouy 1995, Phillips et al. 2004, Romiguier et al. 2013, Bossert et al. 2017; but see Duchêne et al. 2018, Sato et al. 2019). GC bias can be assessed in two ways for a given locus: 1) overall GC content across a range of aligned sites and 2) GC heterogeneity (i.e., the degree to which GC content from each taxon deviates from the overall GC content of an aligned locus) (Bossert et al. 2017). To explore whether GC content and heterogeneity affected our analyses, we assessed these variables in our concatenated and individual locus alignments. Prior to calculating GC content and heterogeneity, we removed invariant sites from our data sets using PAUP* v4.0.a165 since keeping GC invariant sites may bias comparisons among loci. We calculated GC content and GC heterogeneity for each of our data sets using AMAS v0.98 (Borowiec 2016) and a custom python script (Bossert et al. 2017), respectively. The 5% most GC heterogeneous loci were then excluded, followed by the 5% most GC-rich loci (herein referred to as ‘GC-low’ data sets); we refrained from excluding more than 10% of our loci across data sets since, such as our 50p and 70p 25% most informative data sets, were already limited in the number of loci adequate for species tree inferences (see below). Thus, we analyzed four GC-low data sets, as well as our four data sets that included the GC-rich and heterogeneous UCE loci (i.e., all loci).

Phylogenetic Estimation

For each of the data sets, we concatenated single locus alignments in PHYLUCE and then selected the best-fit partition scheme and models of sequence evolution using PartitionFinder v2.1.1 (Lanfear et al. 2017). We used the rcluster algorithm with unlinked branch lengths and treated individual loci as separate data blocks. All models under the ‘raxml’ option were examined (Stamatakis 2006), and the best-fit models were selected using the corrected Akaike Information Criterion (AICc) (Harvich and Tsai 1989). Twenty partitioned maximum likelihood (ML) optimal searches were conducted in RAxML v8.2.10 (Stamatakis 2014) using random starting trees. Bootstrap
support (BS) from 500 iterations were summarized on the best ML tree with SumTrees v4.0.0 (Sukumaran and Holder 2010).

We also estimated species trees from individual gene trees using an approach statistically consistent with the multispecies coalescent model (Degnan and Rosenberg 2006, 2009). We used MrAIC v1.4.6 (Nylander 2004) to select one out of the 56 models of sequence evolution for each locus alignment based on the AICc score in PhyML v3.1 (Guindon et al. 2010). We generated optimal gene trees by performing 20 ML searches in GARLI using results from MrAIC. The use of polymous gene trees has been shown to improve species tree estimation (Zhang et al. 2017); as such, we allowed our input gene trees to have polytomies when zero-length branches were recovered (collapse branches = 1). One hundred bootstrap replicates were also generated by reducing the termination condition parameter by half the default value (i.e., gencresholdtrtoperm = 10,000, see Zwickl 2008). Species trees were inferred from optimal gene trees in ASTRAL-III v5.6.1 (Mirarab et al. 2014, Sayyari and Mirarab 2016, Zhang et al. 2018), with clade support measured using 100 multilocus bootstrap replicates (Seo 2008).

Majority-rule consensus trees were generated from non-GC-reduced data sets in PAUP* v4.0a16 (Swofford 2003) for the following: 1) all resolved optimal trees estimated from every analysis and 2) all optimal trees with branches having <50% BS collapsed. We then computed symmetric differences (2× Robinson-Foulds [RF]) between optimal trees (excluding outgroups) within each of these two groups to assess topological variation across our analyses and to identify if conflicting nodes still existed after poorly supported branches were collapsed.

Because our results produced a polytomy among relatively deeper branches in our majority-rule consensus trees, we evaluated if the incongruence among analyses could be due to incomplete lineage sorting. Under the multispecies coalescent model, rooted three-taxon gene trees will yield a majority resolution identical to the species tree (Pamilo and Nei 1988, Rosenberg 2002, Degnan and Rosenberg 2009). The other two alternative gene tree resolutions will be equiprobable to one another and proportionally less than the majority resolution (Pamilo and Nei 1988, Rosenberg 2002, Degnan and Rosenberg 2009). To evaluate whether incongruence among our analyses was due to incomplete lineage sorting (ILS), we tested our 50p total evidence optimal gene trees for asymmetry among minority gene trees using an exact two-sided binomial test (Zwickl 2008, Richart et al. 2016, Wang et al. 2017, Forthman et al. 2019). Gene trees were pruned to include a representative of three clades at the polytomy with the highest UCE recovery (Galaeus hasticornis [Thunberg], Odontorhopala callosa Stål, Anasa tristis [De Geer]), as well as an outgroup (Halyomorpha halys [Stål]).

Results

Data Summary

For this study, we assembled 3,750–61,308 contigs across samples (mean = 13,284), with a mean length of 451 bp (Supp Table S2 [online only]). We recovered 0–55% of the targeted UCE loci (range across samples: 0–1,470 loci; mean = 1,040), with a mean length of 668 bp; of the 16 dried museum samples, 11 samples yielded 0–32 UCE loci and were subsequently dropped from the analysis. A summary of parsimony-informative sites and number of loci in each data set are provided in Supp Table S3 (online only). The most parsimony-informative UCE locus contained 1,079 informative sites, with the least informative locus having 16 informative sites. As expected, when data sets were constructed with higher loci informativeness thresholds, there was an increase in the proportion of parsimony-informative sites. There was also a decrease in the proportion of invariant sites at higher informativeness thresholds, while the proportion of parsimony-uninformative sites remained similar across all data sets. The overall GC content of our concatenated alignments—when uninformative sites were excluded—averaged 32.3% (30.6–34.0%) (Supp Table S4 [online only]), with taxa ranging from 27.5 to 37.3% across the data sets (Supp Table S5 [online only]). The GC content of each locus ranged from 16.7 to 78.2% (Supp Table S6 [online only]). Across loci, GC heterogeneity ranged from 1.35 to 175.61 (Supp Table S6 [online only]).

General Topological Congruence Across Analyses

The ML phylogram based on our largest UCE data set (50p, total evidence) showed good resolution with most nodes receiving high support (Fig. 2; other ASTRAL and ML trees can be viewed in Supp Figs. S1–S16 [online only]). Phylogenetic results were largely congruent across analyses based on all loci. However, there were still some topological conflicts among our data sets; after pruning outgroup taxa from our trees based on all loci, symmetric distances were 0–8 among optimal summary coalescent trees, 0–8 among ML trees, and 2–12 when distances were calculated between coalescent and ML trees (Supp Table S7 [online only]). We recovered similar values when using trees with poorly supported branches (BS < 50%) collapsed (Supp Table S8 [online only]).

Similar support was also recovered between our GC-low data sets and those data sets including all loci. High GC heterogeneity can adversely affect phylogenetic results by erroneously clustering taxa with similar GC content, but there was no indication that taxa in our phylogenies were clustered as a result of high or low GC content (Supp Table S5 [online only], Fig. 2, Supp Figs. S1–S16 [online only]). As a result, we focus on the all loci data sets and only highlight topological differences with GC-low data sets for nodes of interest. We also provide majority-rule consensus trees of all optimal trees and all trees when branches with BS < 50% are collapsed across analyses based on data sets including all loci in Fig. 3.

Higher-Level Relationships of the Coreinae + Meropachyinae

Most reconstructed relationships within the Coreinae (including Meropachyinae) were congruent across all analyses and highly supported (Figs. 2–4, Supp Figs. S1–S16 [online only]). Some relationships were also consistently recovered across our estimates despite lower support (e.g., Coreus Fabricius + Cletes Stål; Cloresmini + Colpurini + Micrini) (Figs. 2–4, Supp Figs. S1–S16 [online only]). Across all analyses, Coreinae and Meropachyinae were not supported as monophyletic subfamilies. The meropachyine tribes Spathophorini and Merocorini were consistently recovered within coreine clades comprised of Nematopodini + Discogastrini and Acanthocerini + Charistemini + Hypselonotini (part), respectively, with moderate to high support (Figs. 2–4, Supp Figs. S1–S16 [online only]).

The majority-rule consensus tree of all estimated trees from data sets including all loci resulted in a single polytomy near the base of Coreinae + Meropachyinae (Fig. 3), which was driven by uncertainty in the phylogenetic placement of Clade A. This clade was either recovered as the sister to all members of Clade B in summary coalescent analyses (BS = 95–100%) (Supp Figs. S1–S8 [online only]), sister to Clade B + Clade C in most ML analyses (BS = 100%) (Fig. 2, Supp Figs. S9, S10, S12–S14, and S16 [online only]), or sister to Clade C in two ML analyses (BS = 80–83%).
Given this incongruence among analyses, we tested if our estimated gene trees were consistent with the multispecies coalescent model. While most of our gene trees matched the species tree (46.64%), we detected asymmetry among our minority gene trees (33.40 and 18.26%; $P < 0.05$), suggesting incongruence was not driven by incomplete lineage sorting.

When poorly supported branches (BS < 50%) were collapsed in our estimated trees based on all loci, we recovered an additional polytomy in our majority-rule consensus tree among the Discogastrini (Supp Figs. S11 and S15 [online only]). Given this incongruence among analyses, we tested if our estimated gene trees were consistent with the multispecies coalescent model. While most of our gene trees matched the species tree (46.64%), we detected asymmetry among our minority gene trees (33.40 and 18.26%; $P < 0.05$), suggesting incongruence was not driven by incomplete lineage sorting.

When poorly supported branches (BS < 50%) were collapsed in our estimated trees based on all loci, we recovered an additional polytomy in our majority-rule consensus tree among the Discogastrini.
Table 3. Majority-rule consensus tree of all optimal trees (left) and all trees when branches with bootstrap support < 50% are collapsed (right) across maximum likelihood and summary coalescent analyses and data sets including GC-rich and heterogeneous loci (outgroups pruned for visualization). Species names are provided on the left tree, while the names of the corresponding tribes are provided on the right tree. Select tribes (including the subfamily Mepopyraeinae) that are nonmonophyletic are color-coded. Red stars at nodes indicate the location of polytomies in the majority-rule consensus trees. Numbers below branches indicate the proportion of all eight trees that recovered the corresponding clade; branches without numbers were recovered in 100% of trees.

+ Nematopodini + Spathophorini (Fig. 3). The only sampled species of Discogastrini was more often recovered as the sister to the nematopodine genera *Nematopus* Berthold + *Grammopoeius* Stål with weak to moderate support (BS = 38–78%) (Supp Figs. S2–S8, S10, S12, S14, and S16 [online only]), but several analyses recovered this tribe as sister to all Nematopodini + Spathophorini
with 100% support (Fig. 2, Supp Figs. S1, S9, S11, S13, and S15 [online only]). Furthermore, the position of the nematopodine genera *Melucha* Amyot & Serville + *Thasus* also varied across analyses, either recovered as the sister group of all other Nematopodini + Discogastrini + Spathophorini (BS = 100%) (Supp Figs. S2–S8, S10, S12, S14, and S16 [online only]), all remaining Nematopodini + Spathophorini (BS = 97–100%) (Supp Figs. S1 and S10 [online only]), or *Ouranion* Kirkaldy + *Piezogaster* Amyot & Serville + *Mozena* Amyot & Serville + Spathophorini (BS = 75–100%) (Fig. 2, Supp Figs. S9, S11, S13, and S15 [online only]).

The meropachyine tribe Merocorini was typically found to be sister to *Hypselonotus* Hahn + Chariesterini + *Piezogaster* Amyot & Serville + *Mozena* Amyot & Serville + Spathophorini (BS = 97–100%) (Supp Figs. S1 and S10 [online only]), or *Ouranion* Kirkaldy + *Piezogaster* Amyot & Serville + *Mozena* Amyot & Serville + Spathophorini (BS = 75–100%) (Fig. 2, Supp Figs. S9, S11, S13, and S15 [online only]). However, support for this relationship exhibited drastically different values: support was 82% for our 50p total evidence data set, whereas it was 27% for the 70p total evidence data set.

**Nonmonophyly of Coreine Tribes**

At the tribal level, we found support for the nonmonophyly of several tribes of Coreinae (Fig. 2–4, Supp Figs. S1–S16 [online only]). Nematopodini was consistently not monophyletic with respect to Spathophorini or Spathophorini + Discogastrini. The Coreini were polyphyletic, with *Haidara* Distant highly supported as the sister to Phyllomorphini + Gonocerini + *Coreus* and *Coreus* rendered Gonocerini paraphyletic with weak to strong support. Weak to high support for a paraphyletic Daladerini with respect to Latimbinini was observed across all analyses, except for a poorly supported monophyletic Daladerini recovered in our GC-low 50p 25% most informative gene tree summary coalescent analysis (Supp Fig. S6 [online only]). Our results also recovered a polyphyletic *Hypselonotus* with five distinct lineages throughout Clade C, as well as a polyphyletic Anisoscelini with two lineages.

While relationships within other tribes (or larger clades) generally remained unchanged across our different analyses, we did recover a few conflicting results. For example, the phylogenetic positions of genera *Dianomictis* O’Shea, *Plectropodoidea* Schouteden, and *Mygdonia* Stål varied within the Mictini, with all recovered hypotheses generally well supported in ML analyses and relatively weakly supported in coalescent analyses. One other clade, the Hypselonotini Lineage 1, also exhibited two competing hypotheses in the phylogenetic position of *Villasitocoris* Brailovsky, but all of our analyses, except our 50p total evidence coalescent analysis, support the placement of this genus as sister to all other sampled taxa in this clade.

**Genus-Level Phylogenetic Results**

Our results supported most sampled genera as clades comprised of their respective conspecifics with strong support (Figs. 2–4, Supp Figs. S1–S16 [online only]). Only a few genera were not monophyletic. *Leptoglossus* Guérin-Méneville was paraphyletic with respect to *Phthisancemia* Brailovsky, and *Paracera* Burmeister rendered *Anasa* Amyot & Serville paraphyletic, both with high support. We also found evidence for a paraphyletic *Electropoda* Bergroth with respect to *Elasmopoda* Stål with weak to high support.
**Coreinae Paraphyly and Meropachyinae Polyphyly**

Although Meropachyinae have rarely been included in phylogenetic analyses that sample Coreidae, there has been morphological and molecular evidence for the paraphyly of Coreinae with respect to this subfamily (Li 1996, 1997; Forthman et al. 2019), which we corroborate. We found paraphyly of Nematopodini with respect to the meropachyine tribe Spathophorini, as in previous UCE studies (Forthman et al. 2019, Kieran et al. 2019). Additionally, the close relationship between Merocorini (Meropachyinae), Chariesterini, and *Hypselonotus* we recovered is largely consistent with previous cladistic (Li 1996, 1997) and noncladistic (Schaef er 1965; Hepburn and Yonke 1971) studies. Both subfamilies have historically been delimited from the other coreid subfamilies by the presence of a dorsally sulcate tibia (see Forthman et al. 2019). In the taxonomic literature, the two subfamilies have been diagnosed from each other primarily by the presence (Meropachyinae) or absence (Coreinae) of an apical spine or tooth on the hind tibiae, as well as the shape of the hind femora and location of the metathoracic scent gland orifices. Our results indicate that these traits are likely homoplasic.

**Uncertain Phylogenetic Placement of Clades A, B, and C**

Our study finds robust support for a clade comprised of Dasynini, Homoeocerini, Coreini, Phyllomorphini, and Gonocerini (Clade A). The phylogenetic position of Clade A, however, remains uncertain. In our summary coalescent analyses, this clade was sister to Clade B (Daladerini, Latimbini, Cloresmini, Colpurini, and Mictini; see Fig. 3) with high support, while the ML analyses recovered two alternative topologies involving the large Clade C (*Hypselonotini, Acanthocerini, Merocorini, Chariesterini, Anisoscelini, Placoscelini, Acanthocephalini, Cheliniideini*) with weaker support.

The internal branches around the polytomy we recovered are very short relative to other branches at deep nodes (Supp Figs. S9–S16 [online only]). The short, successive branches suggest that this region of the tree might be in an anomaly zone, i.e., a region of the species tree where discordant gene trees are more common than gene trees that are concordant with the species tree due to incomplete lineage sorting (Degnan and Rosenberg 2006, Liu and Edwards 2009). However, our test of minority gene tree asymmetry suggests that our estimated gene trees are inconsistent with the multispecies coalescent model. Thus, discordance around these branches is likely due to other processes.

**Nonmonophyly of Nematopodini**

A paraphyletic Nematopodini with respect to Spathophorini has also been supported by Kieran et al. (2019) and Forthman et al. (2019). However, in some of our analyses, Discogastrini was recovered within Nematopodini rather than as the sister group of Nematopodini + Spathophorini. This was not dependent on the analytical method or type of data set used. Thus, Discogastrini may render Nematopodini (including Spathophorini) nonmonophyletic. To our knowledge, there are no previous hypotheses for a relationship between these two taxa.

Amyot and Serville (1843) included some members of Discogastrini within the Nematopodini based on the presence of enlarged, armed hind femora in males (although, in type images of several genera, the legs of Discogastrini appear slender and unarmed). Discogastrini was subsequently treated as a distinct group from the Nematopodini by Stål (1867), primarily due to the position of the abdominal spiracles. In his comparative morphological study, Schaef er (1965) included the Discogastrini, Homoeocerini, and Latimbini in his *Homoeocerini-group* (each tribe treated as a separate subgroup) based on the structure of the conjunctiva, metathoracic scent gland opening, laterotergites, and external genitalia (Nematopodini not examined). Our results are more in line with Amyot and Serville’s (1843) classification, although the shape and armature of the hind femora may not be synapomorphies for Discogastrini + Nematopodini (including Spathophorini).
Clade A and the Nonmonophyly of Coreini and Gonocerini

The taxonomic history of the Coreini has undergone drastic changes over the last decade. Many of the genera once classified in Coreini are now treated as members of Hypselonotini and other tribes, which we followed here (see CoreoideaSF Team 2019). Our results provide robust support for the exclusion of these genera from the Coreini, but we do not find evidence for a monophyletic Coreini. The relatively close relationship of Coreini and Gonocerini—which is paraphyletic in our study—is congruent with Pan et al. (2007) and Pan et al.’s (2008) Cyb phylogenies but contradicts Li’s (1997) morphological and Fang and Nie’s (2007) COI phylogenies. Schaefer (1965) placed the Coreini (which included genera now in other tribes) with several others (including Gonocerini and Dasynini) in his large Coreus-group. Our phylogenetic result provides limited support to Schaefer’s (1965) study but does not recognize the placement of many other tribes within his Coreus-group. Thus, the characters Schaefer (1965) used to diagnose and describe his Coreus-group and subgroups are likely plesiomorphic or homoplastic synapomorphies based on our molecular phylogenetic hypothesis.

The tribe Phyllomorphini had not been included in previous phylogenetic analyses. Past precladistic morphological studies led some authors to propose Phyllomorphini as a distinct subfamily due to the absence of a dorsal tibial sulcus, as well as several genitalic and abdominal traits (Schaefer 1965; Ahmad 1970, 1979). However, we find support for the inclusion of this tribe within the Coreidae.

Lastly, the sister group relationship between Dasynini and Homoeocerini recovered from our analysis is novel. Schaefer (1965) did not consider these tribes to belong to the same group in his morphological study. Li (1997) included several species of Dasynini and found this tribe to be paraphyletic, but none of the sampled species were found to be closely related to Homoeocerini.

Novel Relationships of Clade B, the Paraphyly of Daladerini, and Phylogenetic Position of Colpurini

The relationships recovered within our well-supported Clade B do not strictly agree with previous studies (Schaefer 1965, Li 1997, Fang and Nie 2007). Daladerini and Latimbini were recovered as the sister of all other tribes within Clade B. We found support for a clade comprised of Cloresmini, as well as its sister group relationship with Colpurini + Mictini that has not been previously proposed. Schaefer (1965) included Cloresmini within his Coreus-subgroup C with other tribes not recovered in this clade (although Mictini was included in a separate Coreus-subgroup within a larger Coreus-group).

Amyot and Serville (1843) described and classified Dalader in the same family-group as genera from Mictini, but it was subsequently treated as a distinct group by Stål (1873). Li’s (1997) morphological phylogenetic hypothesis suggests that the Daladerini are sister to Acanthocerini and Acanthocephalini, in support of Stål’s (1873) treatment of the tribe as separate from the Mictini. Our results also support the exclusion of the sampled daladerine genera from Mictini but predominately find new evidence for the paraphyly of this tribe with respect to Latimbinia, which has a drastically different gross morphology. Schaefer (1965) assigned Daladerini to his Coreus-subgroup B; he assigned the Latimbinia to the Homoeocerini-group but recognized that the position of Latimbinia was uncertain.

Our results for the phylogenetic placement of Colpurini and treatment as a tribe within Coreinae is contradictory with all other studies (Śtyński 1964, Kumar 1965, Schaefer 1965, Ahmad 1970, Li 1997). Past studies have characterized the Colpurini as ‘primitive’ but with many characters (primarily genitalia) suggesting an ‘intermediate’ phylogenetic position between Pseudophloeinae, Hydaraeini, and other Coreinae (see Śtyński 1964; Kumar 1965; Schaefer 1965; Ahmad 1970; Li 1996, 1997). Thus, our results are novel and suggest that genitalic features, as well as external features, should be reevaluated in light of our molecular hypothesis.

Polyphyly of Hypselonotini

With 336 species, the Hypselonotini is the most speciose tribe within the family Coreidae. This tribe has not been formally described or diagnosed even when it was first recognized by Bergroth (1913). Over the last three decades, a number of genera currently recognized within the Hypselonotini (CoreoideaSF Team 2019) have been previously treated as members of the Coreini (e.g., Brailovsky 1988, 1990, 1995, 2016; Packauskas 1994). Members of this tribe have not been included in published phylogenetic analyses and, to our knowledge, appear to have only been examined by Schaefer (1965) in his comparative morphological study (treated as part of Coreini). Our study supports the exclusion of the sampled hypselonotine genera from the Coreini but not the monophyly of this large tribe; five independent lineages were robustly supported in our analyses. Of these, only the taxonomic position of Hypselonotus within the Hypselonotini + Charisterini + Merocorini was not congruent across analyses. It is evident from our results that the taxonomic status of Hypselonotini is in further need of evaluation. Including additional genera of this tribe will provide further insights into the extent of hypselonotine polyphyly.

Polyphyly of Anisoscelini

The tribe Anisoscelini is a moderately sized group, with 183 species that exhibit a diversity of color patterns and morphology, particularly with the shape and size of the foliaceous expansions on the hind tibiae. Members of Anisoscelini were once divided among two tribes (Anisoscelini [or Anisoscelidini] and Leptoscelidini), primarily based on the presence or absence of hind tibial expansions (e.g., Stål 1867, Schaefer 1965, Packauskas 1994). Schaefer (1965, 1968) classified both of these former tribes as members of the Acanthocephala-group with Acanthocephalini and Placocerini. We find some support for Schaefer’s (1965, 1968) scheme, but we do not support his exclusion of Chelinideini and Hypselonotini from the Acanthocephala-group. The two distinct anisosceline lineages we recovered do not appear to correspond to previously proposed tribal classifications and phylogenetic hypotheses nor appear to be separated based on the presence of tibial expansions. Like Hypselonotini, careful evaluation of this tribe is needed to understand the extent of anisosceline polyphyly.

Phylogenetic Position of Chelinideini

We support a sister group relationship between Chelinideini and our Hypselonotini Lineage 5 + Anisoscelini Lineage 2 clade (Fig. 3), contrary to previous studies. Li’s (1997) phylogenetic hypothesis found a close relationship between Chelinideini and Homoeocerini. Based on a survey of genitalic morphology, Schaefer (1965) placed this tribe in a Coreus subgroup (subgroup C) that included Gonocerini, Acanthocorini (part), Cloresmini, and Coreini. In fact, Chelinideini was once classified within the Coreini, but was elevated to tribal rank by Blatchley (1926). Thus, our results on the phylogenetic position of Chelinideini are novel and require further comparative work to identify and test potential synapomorphies.

Paraphyletic Genera

The genera Electropoda and Elasmopoda are two of several genera comprising the Elasmopoda complex (Linnavauri 1978, O’Shea 1980c) and share many morphological similarities. Both genera have...
been treated as separate groups since Stål (1873), but some species have historically experienced changes in generic assignment between these two groups. Linnavuori (1978) revised the Elasmopoda complex but noted that some features used to diagnose Elasmopoda are also observed in some Plectropoda species. This may suggest that one or both genera are not monophyletic, and our study supports the paraphyly of Plectropoda with respect to Elasmopoda. This indicates that the taxonomic limits of these genera should be evaluated further.

The genus Leptoglossus is an agriculturally important group (Schaefer and Mitchell 1983, Jankevicius et al. 1993, Fernandes et al. 2015), with many of the 62 species distributed throughout the New World, while two species, L. gonagra (Fabricius) and L. occidentalis Heidemann, occur worldwide (CoreoideaSF Team 2019). Our study recovered a paraphyletic Leptoglossus, with L. gonagra recovered as sister to Phthiascemia picta (Drury), both of which were sampled from the New World. Based on our survey of the taxonomic literature and the Coreoidea Species File (CoreoideaSF Team 2019), these two genera have never been treated as or hypothesized to be the same. Allen (1969) created two divisions within Leptoglossus—Divisions A and B. While we do not support the monophyly of this genus, we do find limited support for Allen’s (1969) separate treatment of his Division A (i.e., L. gonagra) from those of Division B (i.e., L. phyllops [Linnaeus], L. clypealis Heidemann, L. concolor [Walker]).

The hypselonotine genus Anasa is a large genus comprised of 77 New World species (CoreoideaSF Team 2019), with several of economic importance (Schaefer and Mitchell 1983, Fernandes et al. 2015). We also found evidence for the paraphyly of Anasa with respect to the hypselonotine genus Paryphes. The close relationship between these two genera have not been previously shown. Stål (1867) provided a framework to separate the two genera based on the curvature of the head and the structure of the antennal segments. In this same publication, Stål (1867) transferred a single species from the hypselonotine genus Anasa—L. clypealis [Linnaeus].

Additional Observations
We observed a 1.5–2× increase in contig assembly and UCE recovery when implementing our TE-touchdown protocol. Such an increase may be due to the moderate reduction in hybridization temperature (i.e., from 65 to 62°C) and/or the sonication method associated with our target capture approach. However, this observation may have been confounded by the DNA extraction methods employed and/or by the early termination of our initial hybridization, although baits from the first hybridization should have been ineffective during the second attempt. Future UCE capture studies in invertebrates should continue exploring touchdown hybridization approaches in more controlled experiments to test for improved locus recovery.

Our sample size of successfully sequenced degraded museum material was exceptionally low. Four were recently collected (2015–2016), whereas one was much older (collected in 1946). Our recent samples recovered a relatively large number of UCEs compared to our historical sample. While our low sample size cannot afford any conclusions, we noted an increase in UCE recovery and locus lengths when the DNQIA extraction protocol was used; however, the tissues sampled among taxa may confound this observation. Although freshly preserved specimen material is ideal for UCE capture studies, such tissues may be limited for some taxonomic groups (e.g., rare tribes missing in this study). Thus, future UCE capture studies should continue exploring alternative approaches to improve UCE recovery of more historical museum specimens through rigorous analyses.

Conclusion
The results of our phylogenomic analysis suggest paraphyly and polyphyly of several genera, tribes, and subfamilies of Coreidae, indicating that the taxonomic classification of this diverse family and its largest subfamily, the Coreinae, is in critical need of evaluation and future revision. We were able to robustly re-construct relationships at shallower phylogenetic scales within the coreid subfamily Coreinae, demonstrating that invertebrate UCEs are suitable at a variety of scales (i.e., subfamily to congeners). Additional capture studies within the Hemiptera are needed to test the utility of UCEs at more recent times scales, e.g., with greater taxon sampling within genera and populations within species. Our results additionally highlight areas of the phylogeny that should be further tested with increased taxon sampling, including large and/or diverse clades (e.g., Daladerini, Nematopodini, Mictini, etc.), as well as taxa for which fresh tissue samples are not commonly available.

Supplementary Data
Supplementary data are available at *Insect Systematics and Diversity* online.

Table S1. Taxon sampling and summary information on molecular protocols.
Table S2. Summary data for sequence reads, contigs, and ultraconserved element loci generated in this study.
Table S3. Summary of informative sites and number of UCE loci for each data set.
Table S4. GC content for each UCE concatenated alignment (uninformative sites excluded from calculations).
Table S5. GC content for each taxon in the concatenated alignments (uninformative sites excluded from calculations). Abbreviations: TEv, total evidence; 25MI, 25% most informative loci.
Table S6. GC content and heterogeneity for each UCE locus alignment (uninformative sites excluded from calculations).
Table S7. Symmetric distances across all optimal trees based on data sets including GC-rich and heterogeneous loci (outgroups excluded).
Table S8. Symmetric distances across all trees after branches with less than 50% bootstrap support in optimal trees were collapsed based on data sets including GC-rich and heterogeneous loci (outgroups excluded).

Figure S1. ASTRAL-III species tree generated from all 50p gene trees (i.e., total evidence), including GC-rich and heterogeneous loci. Values at nodes represent bootstrap support.
Figure S2. ASTRAL-III species tree generated from the 25% most informative 50p gene trees, including GC-rich and heterogeneous loci. Values at nodes represent bootstrap support.
Figure S3. ASTRAL-III species tree generated from all 70p gene trees (i.e., total evidence), including GC-rich and heterogeneous loci. Values at nodes represent bootstrap support.
Figure S4. ASTRAL-III species tree generated from the 25% most informative 70p gene trees, including GC-rich and heterogeneous loci. Values at nodes represent bootstrap support.
Figure S5. ASTRAL-III species tree generated from all GC-low 50p gene trees (i.e., total evidence). Values at nodes represent bootstrap support.
Acknowledgments

We thank the village it took to make this study happen. We thank the following individuals and institutions that contributed specimens used in this study: Christiane Weirauch, Dimitri Forero, Ummat Somjee, Takahisa Miyatake, Jason Cryan, Mark Deyrup, Wei Song Hwang, Li You, Oliver Keller, John Bejerano, G. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, and Caroline Miller. Nathan Friedman and Caroline Miller assisted with molecular protocols. We thank Travis Glenn and Troy Kieran (University of Georgia) for intensive training in library preparation and target enrichment. Gavin Naylor and Shannon Corrigan (University of Florida) provided training on Covaris ultrasonication. Silas Bossert kindly provided a custom python script to calculate GC heterogeneity. Collection of Florida (USA) specimens was supported by a University of Florida Research Abroad for Doctoral Students (awarded to Zachary Emberts). Additional samples collected were supported by funding awarded to Ummat Somjee: Society of the Study of Evolution Rosemary Grant Award; Smithsonian Tropical Research Institute Short Term Fellowship; and Systematics, Evolution, and Biodiversity Endowment Award (Entomological Society of America). This study was funded by the National Science Foundation IOS-1553100 (awarded to C.W.M.).

Data Accessibility

Sequence read files are available on NCBI’s Sequence Read Archive under BioProject PRJNA546248. Alignments and gene tree files are available on Figshare under project title ‘Phylogenomics of the leaf-footed bug subfamily Coreinae (Hemiptera: Coreidae)’.

References Cited


Blatchley, W. S. 1926. Heteroptera or true bugs of Eastern North America with especial reference to the faunas of Indiana and Florida. The Nature Publishing Company, Indianapolis, IN.


