2011

Multi-Locus Phylogeny Of Sponge-Dwelling Snapping Shrimp (Caridea: Alpheidae: Synalpheus) Supports Morphology-Based Species Concepts

KM Hultgren

JE Duffy

JOURNAL OF CRUSTACEAN BIOLOGY, 31(2): 352-360, 2011

Journal of
Crustacean Biology
A quarterly of The Crustacean Society for publication of research on the biology of Crustacea & marine arthropods

MULTI-LOCUS PHYLOGENY OF SPONGE-DWELLING SNAPPING SHRIMP (CARIDEA: ALPHEIDAE: *SYNALPHEUS*) SUPPORTS MORPHOLOGY-BASED SPECIES CONCEPTS

Kristin M. Hultgren and J. Emmett Duffy

(KMH, hultgrenk@si.edu) National Museum of Natural History, Smithsonian Institution, MRC 163, PO Box 37012, Washington, DC 20013-7012;
(JED, jeduffy@vims.edu) Virginia Institute of Marine Science, P.O. Box 1346, Gloucester Pt., Virginia 23062-1346, U.S.A.

ABSTRACT

Alpheid snapping shrimp are one of the most diverse groups of coral-reef fauna, and sponge-dwelling shrimp in the genus *Synalpheus* (*gambarelloides* species group) have in particular become a model system for studying the evolution of social biology and host use in marine invertebrates. Despite recent advances in understanding the evolution and systematics of *Synalpheus*, the taxonomy and phylogenetic relationships within this group remain challenging. More than 20 new species in the *S. gambarelloides* species group have been described over the past two decades, primarily within several cryptic species complexes, which has doubled the known diversity of this group in the West Atlantic. Here we construct a new phylogenetic tree describing relationships between 40 different species from the *S. gambarelloides*-group (119 individuals from across the Caribbean), using a combined dataset consisting of two mitochondrial loci (16S and COI), one nuclear protein-coding gene (elongation-factor 2), and 33 morphological characters. Putative conspecific specimens of *Synalpheus* from multiple locations across the Caribbean were always monophyletic (with one exception), providing strong support for the validity of species concepts based on morphology. Our study also provides further evidence for the monophyly of the *S. gambarelloides*-group in the Caribbean, resolves the molecular relationships within many recently described species complexes, and provides a new phylogenetic framework for future evolutionary studies of this group.

KEY WORDS: Alpheidae, Caridea, phylogeny, snapping shrimp, *Synalpheus gambarelloides*

DOI: 10.1651/10-3382.1

INTRODUCTION

The snapping shrimp genus *Synalpheus* is one of the most taxonomically and ecologically diverse group of coral-reef fauna, with over 150 species worldwide, and many species are symbionts with a diverse range of larger host taxa (Coutière, 1909; Banner and Banner, 1975; Bruce, 1976; Rios and Duffy, 2007). Species in the *Synalpheus gambarelloides*-group (Nardo, 1847) are by far the most well-studied group of these snapping shrimp. The *S. gambarelloides*-group is largely limited to the western North Atlantic (with some exceptions), is thought to be monophyletic based on prior molecular phylogenies (Morrison et al., 2004), and its members share a few key morphological synapomorphies, most notably a brush of setae on the dactyl of the minor chela (Coutière, 1909; Rios and Duffy, 2007). Members of this species group also live exclusively in the canals of marine sponges (Dardeau, 1984; Duffy, 1992; Macdonald et al., 2006; Rios and Duffy, 2007). Although accounts of sponge-dwelling have been noted in several other synalpheid and alpheid species (Banner and Banner, 1975, 1982), this symbiotic habit is most well documented in members of the *S. gambarelloides*-group (Duffy, 1996c; Macdonald et al., 2006). Most notably, several species within this group are eusocial—living in large, reproductively skewed colonies consisting of tens to hundreds of individuals with one or a very few breeding females (Duffy, 1996a, 2007). As such, this group has become a model system for studying not only the evolutionary ecology of host use (Duffy, 1992, 1996b; Macdonald et al., 2006; Hultgren and Duffy, 2010), but also the ecology and evolution of social biology (Duffy et al., 2000, 2002; Duffy, 2007; Duffy and Macdonald, 2010).

Our understanding of the evolutionary ecology of host use and social systems in these shrimp has been transformed by prior work on the phylogenetics of this group from the West Atlantic (Duffy et al., 2000; Morrison et al., 2004). Previous phylogenetic work has demonstrated that eusociality evolved independently at least three times in the West Atlantic *Synalpheus* (Duffy et al., 2000) and that the high diversity of the *S. gambarelloides*-group may have originated from a rapid species radiation ~ 5-7 MYA (Morrison et al., 2004), concurrent with radiations of several other Caribbean reef-associated species. This period of ancient rapid cladogenesis is evident in branching patterns in phylogenies of *Synalpheus*, which are characterized by short basal branches and poorly supported nodes at the base of the tree, and much better support for more terminal species-level nodes (Morrison et al., 2004). Alternately, this pattern could have resulted from the genes used to reconstruct relationships; both previously published phylogenies of *Synalpheus* relied only on mitochondrial molecular loci (16s rRNA and cytochrome oxidase I), which tend to evolve rapidly and may be less useful in reconstructing more basal relationships. Nuclear loci

typically evolve more slowly than mitochondrial loci (Palumbi et al., 2001; Sanderson and Shaffer, 2002), and serve as independent molecular loci from the nuclear genome. As such, nuclear loci have been useful in reconstructing older (family-level) crustacean relationships (Porter et al., 2005; Ahyong et al., 2007; Tsang et al., 2008; Bracken et al., 2009), and may be useful in reconstructing the more basal relationships of the *S. gambarelloides*-group.

Synalpheus, like many other groups of snapping shrimp in Alpheidae, are characterized by high levels of cryptic speciation (Morrison et al., 2004; Rios and Duffy, 2007; Anker and De Grave, 2008; Macdonald et al., 2009; Mathews and Anker, 2009), similar to many other coral reef-associated organisms (Knowlton, 1993; Duda and Kohn, 2005; Puebla et al., 2008). In the West Atlantic alone, systematic surveys of sponge-dwelling *Synalpheus* have doubled the number of described species in the last 14 years from 20 to 43 (Duffy, 1996d, 1998; Rios and Duffy, 1999, 2007; Macdonald and Duffy, 2006; Anker and Toth, 2008; Macdonald et al., 2009; Hultgren et al., 2010). Twenty of these species (46% of West Atlantic diversity) were formally described subsequent to the most recent phylogeny of *Synalpheus* (Morrison et al., 2004). Many of these new species belong to closely related species complexes – groups in which one or two nominal morphological species were recently split into multiple species, often based on subtle morphological differences. These species complexes and the morphological characters that define them have been extensively described elsewhere (DOI: 10.1651/10-3382.1 on-line Appendix 1; Macdonald and Duffy, 1998; Rios and Duffy, 2007; Anker and Toth, 2008; Macdonald et al., 2009; Hultgren et al., 2010), but the distinctness and monophyly of many of these species have not yet been corroborated with molecular data. An updated molecular phylogeny with increased taxon sampling, including species replicates from multiple locations across the Caribbean, would aid in assessing the validity of species concepts based on morphology, in addition to elucidating the phylogenetic placement of these new species.

In this study, we combine molecular sequence data from a nuclear locus (elongation-factor 2) with data from two mitochondrial loci (cytochrome oxidase I and 16S rRNA) and 33 morphological characters to construct an updated phylogenetic tree of 40 Caribbean species of the *S. gambarelloides*-group ($\sim$ 80% of described species in the *S. gambarelloides*-group and $\sim$ 25% of described world-wide synalpheid diversity). Our aims were to: 1) utilize sequence data from an independent nuclear locus to resolve more basal relationships within the group, 2) include in the phylogenetic analysis a substantial number of species, many recently described, that were not studied in previous phylogenetic analyses, and 3) better resolve molecular relationships among species in several putative species complexes of morphologically similar taxa. The taxa used in this study include 9 species from the *S. gambarelloides*-group that have not been included in previous phylogenetic analyses, representing a 30% increase in taxon sampling from previous studies, and include multiple geographically distinct populations for 19 of the 40 species of *Synalpheus* used in this study. Our study resolves molecular relationships in four cryptic species complexes, and provides an updated phylogenetic framework for this diverse shrimp group.

Materials and Methods

Taxon Sampling

For this study, we focused on representatives of the *S. gambarelloides*-group in the West Atlantic, in which most members of the group are endemic. We sampled 70 new individual specimens of *Synalpheus* from the *S. gambarelloides*-group and 3 species of *Synalpheus* from outside of the *S. gambarelloides*-group from the West Atlantic, and also included sequences from 45+ specimens of *Synalpheus* from a previous study (Morrison et al., 2004). Although our sample overlaps broadly with those 32 species sampled in Morrison et al. (2004), we include 9 additional species of *Synalpheus* that have recently been described (Rios and Duffy, 2007; Anker and Toth, 2008; Macdonald et al., 2009; Hultgren et al., 2010) or are in the process of being described (Macdonald, Hultgren, Duffy, unpublished data) [DOI: 10.1651/10-3382.1 on-line Appendix 1]. New shrimp vouchers used in this study were collected between 2005 and 2009 from Belize (Carrie Bow Cay; 16°48′N, 88°05′W), Caribbean Panama (Bocas del Toro; 9°20′N, 82°15′W), Jamaica (Discovery Bay; 18°28′N, 77°27′W), Curaçao (12°12′N, 69°4′W), and Barbados (13°13′N, 59°38′W). In general, shrimp were collected from their host sponges and identified while alive; whole specimens were preserved in 95% ethanol (pre-2009 collections); for 2009 collections, a sample (legs or eggs) of each specimen was preserved in RNAlater (Applied Biosystems), and the rest of the specimen was preserved in 95% ethanol for morphological identification.

Morphological Characters

We used a dataset of 33 morphological characters (DOI: 10.1651/10-3382.1 on-line Appendices 2 and 3) for phylogenetic analysis, based on a previously published character set (Morrison et al., 2004). As we were primarily interested in relationships within the *S. gambarelloides*-group, we omitted several characters from this previous character set that primarily delineated these species from other *Synalpheus* and were uninformative for resolving relationships within the *S. gambarelloides*-group. Morphological characters for all new species to the study were scored by both investigators (KMH and JED).

DNA Extraction, Amplification, Sequencing, and Alignment

For all new specimens (collected 2007-2009) preserved in RNAlater, we extracted total RNA using either the SV Total RNA Isolation System (Promega, Madison WI) using a modified protocol that preserved some genomic DNA (Regier, 2008). For specimens preserved in 95% EtOH, we directly extracted genomic DNA using the tissue protocol on a Biosprint 96 workstation (Qiagen).

We sequenced three loci for specimens in this study: the mitochondrial cytochrome oxidase I gene (COI; $\sim$ 600 bp of the 3′ end); the mitochondrial large-subunit ribosomal gene (16S, $\sim$ 510 bp); and the nuclear gene elongation factor 2 (EF2, $\sim$ 700 bp). For COI, we used *Synalpheus*-specific primers COI-G4 (5′ CACCCAGAAGTYTATATTC-TAAT 3′) and COI-G (5′ TGTTGGGGGAAGAATGTAAT 3′) from Morrison et al. (2004). For 16S, we used 16ar/br primers (Palumbi et al., 1991), or in rare cases *Synalpheus*-specific 16S primers (16S_3F: 3′ TAAAGGGCTGCGGTAATTTG′, 16S_3R: 3′ CGAACAGGCCTTCCC TTTA 5′). For both of these loci, we directly amplified genomic DNA using previously-described PCR conditions optimized for *Synalpheus* (Morrison et al., 2004) and purified PCR products using a shrimp alkaline phosphate exonuclease protocol (USB Corporation).

We amplified EF2 using a RT-PCR protocol and primers originally optimized for the related genus *Alpheus* (Hurt et al., 2009). Using total RNA extractions, we synthesized cDNA using MuLV reverse transcriptase (Applied Biosystems), RNase inhibitor (Applied Biosystems), and an exterior reverse EF2 primer (EF2–1587R: 5′-AYR ATG TGY TCT CCR GAY TC-3′). This cDNA was used as a template in a PCR reaction that

utilized an exterior forward EF2 primer (EF2–723F: 5′-MMA AGY TST GGG GTG ARA AC-3′), and PCR/thermocycler conditions from Regier (2008). We then used a second set of internal EF2 primers (EF2–739F: 5′-GAG RGC YTT CAA CAC CTA YA-3′, EF2-1499r: 5′-ART CGG AGG GGT TCT TGG-3′) in a nested PCR protocol (Regier, 2008; Hurt et al., 2009). PCR products were run on a 1% low-melt agarose gel, and products of the correct size were gel-excised and extracted using the GELase protocol (Epicentre Biotechnologies). As studies in other systems (including the related genus *Alpheus*) indicated that intra-individual variation in this nuclear locus was low (Regier, 2008; Hurt et al., 2009), we did not perform cloning and directly amplified EF2 PCR products. All purified PCR products (EXO-SAP or gel extractions) were sequenced on an ABI Prism 3730XI sequencer; forward (5′-3′) and reverse (3′-5′) sequences from all loci were visually checked and trimmed using the program SEQUENCHER v4.8 (Gene Codes corporation).

For each individual locus, we performed alignments using the default parameters on the program Muscle v3.8 (Edgar, 2004), and manually checked alignments using MacClade 4.08 (Maddison and Maddison, 2005). We used MacClade to calculate codon positions for the two protein-coding genes (COI and EF2) and to check for stop codons, which can indicate the presence of pseudogenes (Buhay, 2009); no stop codons were detected in either set of coding sequences. We also compared individual gene trees to check for unusual taxon placements that could indicate pseudogenes. We calculated the appropriate model of nucleotide substitution for each of the three loci using the program MrModeltest v2.3 (Nylander, 2004), and used the Akaike Information Criterion (AIC) to select the model of molecular evolution that best fit the data.

Taxon Sampling and Tree Construction

In general, RNA extraction and the RT-PCR method used to amplify EF2 worked poorly for older templates, e.g., preserved for $\gg$ 6 years in EtOH; because we lacked EF2 sequence data for specimens included in prior analyses (Morrison et al., 2004), portions of our molecular dataset were missing (DOI: 10.1651/10-3382.1 on-line Appendix 1). Thus, in addition to single-locus trees, we constructed two different combined-locus trees. First, for all taxa that were successfully sequenced for all loci, we constructed Bayesian trees using a combined dataset of 1849 molecular and 33 morphological characters (16S, COI, EF2, and morphology) that maximized character sampling (''complete-character'' dataset, $n = 37$ individuals, 34 species). However, simulation studies suggest that utilization of taxa with incomplete sequence data (taxa with $\geq$ 50% of the character matrix sampled; DOI: 10.1651/10-3382.1 on-line Appendix 1) can increase the accuracy of the final tree if the number of sampled characters is high ($\gg$ 200 characters) (Wiens, 2005, 2006). Thus, we also constructed a combined-data tree that maximized taxon sampling (''complete-taxa'' dataset, $n = 119$ individuals, 48 species), and included all individuals with at least two of the molecular loci sequenced ($\geq$ 1100 molecular characters) and the morphological character set. *Alpheus estuariensis* Christoffersen, 1984 was used as the out-group.

We constructed all trees using model-based approaches (Bayesian and maximum likelihood methods) because of their ability to incorporate information about the model of evolution of individual loci and correct for multiple substitutions (Huelsenbeck et al., 2002; Leache and Reeder, 2002). We ran partitioned Bayesian analyses on MrBayes v3.12 (Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003) and used maximum likelihood (ML) methods to run a combined-data tree using the program GARLI-part v0.97 (Zwickl, 2006, 2010). For all analyses, we treated all gap data, e.g., for the non-coding 16S locus, as missing data because there are few well-investigated methods for implementing gap information into model-based approaches to phylogenetic inference (Phillips et al., 2000; Simmons et al., 2007). For Bayesian analyses, we used information about the general model of evolution for each locus calculated by MrModeltest (shape of rate distributions), but allowed MrBayes to estimate more specific model parameters (proportion invariable sites, base frequencies, nucleotide substitution rates) individually for each locus. We ran Markov Chain Monte Carlo (MCMC) searches with two runs (Nruns = 2) and four chains for 5×10^6 generations (for single-locus trees), 1×10^7 generations (complete-character tree), or 3×10^7 generations (complete-taxa tree); at this point runs had converged to a stationary distribution (standard deviation of split frequencies $\ll 0.01$). We sampled the chain every 100 generations, and discarded the first 25-30% of the samples as the burn-in (standard deviation of split frequencies after this burn-in sample < 0.01); support for nodes on trees was estimated using Bayesian posterior probabilities (bpp). For the ML analysis, we used model parameters calculated by MrModeltest to set the model of evolution for each genetic dataset in GARLI-part, used the Mkv model (Lewis, 2001) for the morphological dataset, and estimated support using 1000 bootstrap replicates (complete-character tree) or 500 bootstrap replicates (complete-taxa tree).

Results

For 16S, we obtained sequence data from 114 individuals representing 45 different species (569 bp of aligned sequence data). For COI, we obtained sequence data from 112 individuals (47 different species, 580 bp of aligned sequence data). For EF2, we obtained sequence data from 47 different individuals (38 different species, 700 bp of aligned sequence data). We obtained 33 morphological characters from all 49 species. We used best-fit nucleotide substitution models calculated from MrModeltest (16S and COI: GTR + I + G, $N_{st} = 6$ rates = invgamma; EF2, SYM + I + G, $N_{st} = 6$, rates = invgamma).

We initially ran trees for each molecular locus separately; while these trees successfully resolved distinct species and species complexes (Table 1), they did not generally recover deeper clades with high Bayesian posterior probabilities, and we present only combined-data trees in this analysis. For both combined-data trees, the ML analysis did not resolve basal branches with high bootstrap (bs) confidence values, but demonstrated strong support for most major species complexes; thus we present the Bayesian topology for both combined-data trees, but include ML bs values on all nodes resolved by both methods (Figs. 1-2). The complete-character Bayesian tree showed strong support for a monophyletic *S. gambarelloides*-group (bpp = 1), but the ML tree showed fairly weak support for this clade (bs = 68), albeit higher support than ML trees from prior studies (Morrison et al., 2004). The Bayesian tree resolved the group into several distinct clades (Fig. 1). The *Synalpheus longicarpus* (Herrick, 1891) species complex (bpp = 1, bs = 100) was the most basal clade in the group, followed by a clade consisting of the *Synalpheus rathbunae* Coutière, 1909 species complex (bpp = 1, bs =100) and a small group of species related to *Synalpheus sanctithomae* Coutière, 1909 (bpp = 1, bs = 98). The *Synalpheus paraneptunus* Coutière, 1909 species complex, which here we define to include *S. kensleyi* (Rios and Duffy, 2007), was also supported (bpp = 1, bs = 90). Finally, there was moderate support (bpp = 1, bs = 75) for a large complex of species related to *Synalpheus brooksi* Coutière, 1909 .

The complete-taxa tree (Fig. 2), consisting of all taxa with 2/3 loci sequenced, showed considerably less resolution at deeper branches, though major species complexes were resolved with similarly high Bayesian posterior probabilities (bpp = 1, bs = 85-100; Table 1). In most cases, there was high support (bpp $\geq$ 0.98) for the monophyly of individual species from different geographic locations. For one species in the *S. paraneptunus* species complex, the undescribed species *S. ''paraneptunus-4,''* there was somewhat weaker support (bpp = 0.85, bs =57) possibly due to the difficulty of classifying the voucher specimen from Florida (parsFL01). There was no support

Table 1. Statistics for Bayesian (BY) or maximum likelihood (ML) trees constructed from different taxon and character sets. Clade confidence values for species groups or species complexes are given as posterior probabilities (Bayesian analyses) or bootstrap confidence values (ML analysis). Trees 1-4 are based on single character datasets (16S, COI, EF2, or morphology; not figured); trees 5-10 are based on combined datasets.

Tree statistics	16S tree	COI tree	EF2 tree	Morph. tree	Morrison et al., 2004 (Fig. 2b)	Morrison et al., 2004 (Fig. 2c)	"Complete character" tree (Fig. 1)		"Complete taxa" tree (Fig. 2)	
Method	BY	BY	BY	BY	ML	BY	ML	BY	BY	ML
# Individuals (total)	114	114	47	119	50	50	37	37	119	119
# Species (total)	46	48	38	49	41	41	34	34	48	48
# Species (*gambarelloides*)	39	39	34	41	32	32	31	31	41	41
Clade confidence values										
"*gambarelloides*" group	–	–	1	–	–	0.87	68	1	1	85
S. longicarpus complex	1	0.72	–	–	81	1	100	1	1	99
S. rathbunae complex	1	1	–	0.93	96	1	100	1	1	100
S. paraneptunus complex	1	0.86	–	–	87	1	90	1	1	87
S. brooksi complex	–	0.98	–	–	52	0.99	75	1	1	88

for the monophyly of species (from a large range of different hosts and locations) identified as *S. brooksi*.

Discussion

Our new combined-data analysis of West Atlantic sponge-dwelling species of *Synalpheus* includes the majority of currently known described and undescribed species from this region, and demonstrates strong support for the monophyly of these species from multiple Caribbean locations, despite often subtle morphological differences separating them. In addition to resolving relationships within many recently described cryptic species complexes, our analyses also helped elucidate deeper phylogenetic relationships among these groups, although the exact branching patterns remain elusive.

The most well-supported and well-resolved tree was constructed from species from which we had all four character sets (16S, COI, EF2, and morphology) using Bayesian methods; the complete-character tree had the fewest clades rooted at polytomies, and showed strong support for the monophyly of both individual species complexes and for the *S. gambarelloides*-group. Although the ML analysis resolved individual species complexes with high confidence, it did not resolve basal branching patterns with high (bs $>$ 70) confidence values. In the complete-taxa Bayesian tree, inclusion of additional taxa with some molecular character data missing ($\sim$ 33% missing data, primarily EF2 sequences) resulted in a tree with strong support for the *S. gambarelloides*-group and for all major clades and species complexes, albeit with relatively less resolution at the deeper nodes of the tree. Most importantly, this tree showed strong support for the monophyly of nearly all 48 species of *Synalpheus* recognized from our study, most notably 19 species we were able to sample from multiple geographic regions. The *S. gambarelloides*-group has historically been difficult to identify and diagnose because of the subtlety of morphological features distinguishing closely related species, geographic variability in other morphological characters, and the tendency of closely related species to co-occur in the same geographic locations and sometimes in the same sponge hosts (Coutière, 1909; Chace, 1972; Dardeau, 1984; Macdonald et al., 2006; Rios and Duffy, 2007). Our data lend strong support to the validity of these morphology-based species concepts, suggesting that the tendency of the pioneer alpheid systematist Henri Coutière (Coutière, 1909) to "split" species of *Synalpheus* into multiple species and subspecies was generally on target. These data also support our decision to combine sequence data from multiple individuals (or use sequence data from one individual) as "exemplars" for that species in the complete-character tree (Fig. 1).

In addition to validating the distinctiveness of Caribbean species of *Synalpheus*, our tree also resolves relationships among several new species described in the last two decades, often within closely related species complexes. Since Dardeau's extensive (1984) monograph on Caribbean *Synalpheus,* the number of species in the Western Atlantic has more than doubled from 19 to 44 (with descriptions of five additional putative species currently in progress). The *S. brooksi*-complex is perhaps the most indicative of this trend; seven new species that would formally have keyed out to *S. brooksi* have been described since 1984 (when only *Synalpheus bousfieldi* Chace, 1972 and *S. brooksi* were described), representing a $>$ 400% increase in species diversity within this complex. Perhaps surprisingly, *S. brooksi* and *S. bousfieldi* were the only two taxa from our study that were not distinctly resolved in our trees. Although there was support for a clade of *S. bousfieldi* in the complete-taxa tree (Fig. 2, bpp = 0.99, bs = 0.75), there was no support for a monophyletic *S. brooksi*, and these two morphologically very similar species were never resolved in single-locus trees or in combined-locus trees based only on molecular data (data not shown). Both species live in a number of different sponge hosts and have widespread, partially overlapping species ranges in the Caribbean; they differ in some rather subtle morphological characters (Macdonald and Duffy, 2006) and exhibit clear host differences where they co-occur (Macdonald et al., 2006). Our study suggests that increased taxon sampling from multiple sponge hosts, as well as utilization of additional rapidly evolving loci, are needed to adequately assess the genetic distinctions between these species, and within *S. brooksi* in particular.

The *S. paraneptunus*-complex has also traditionally been a taxonomically difficult group to diagnose; although members of the *S. paraneptunus* complex can be distinguished clearly from the rest of the *S. gambarelloides*-group by the a distinctive character – specifically, sparse

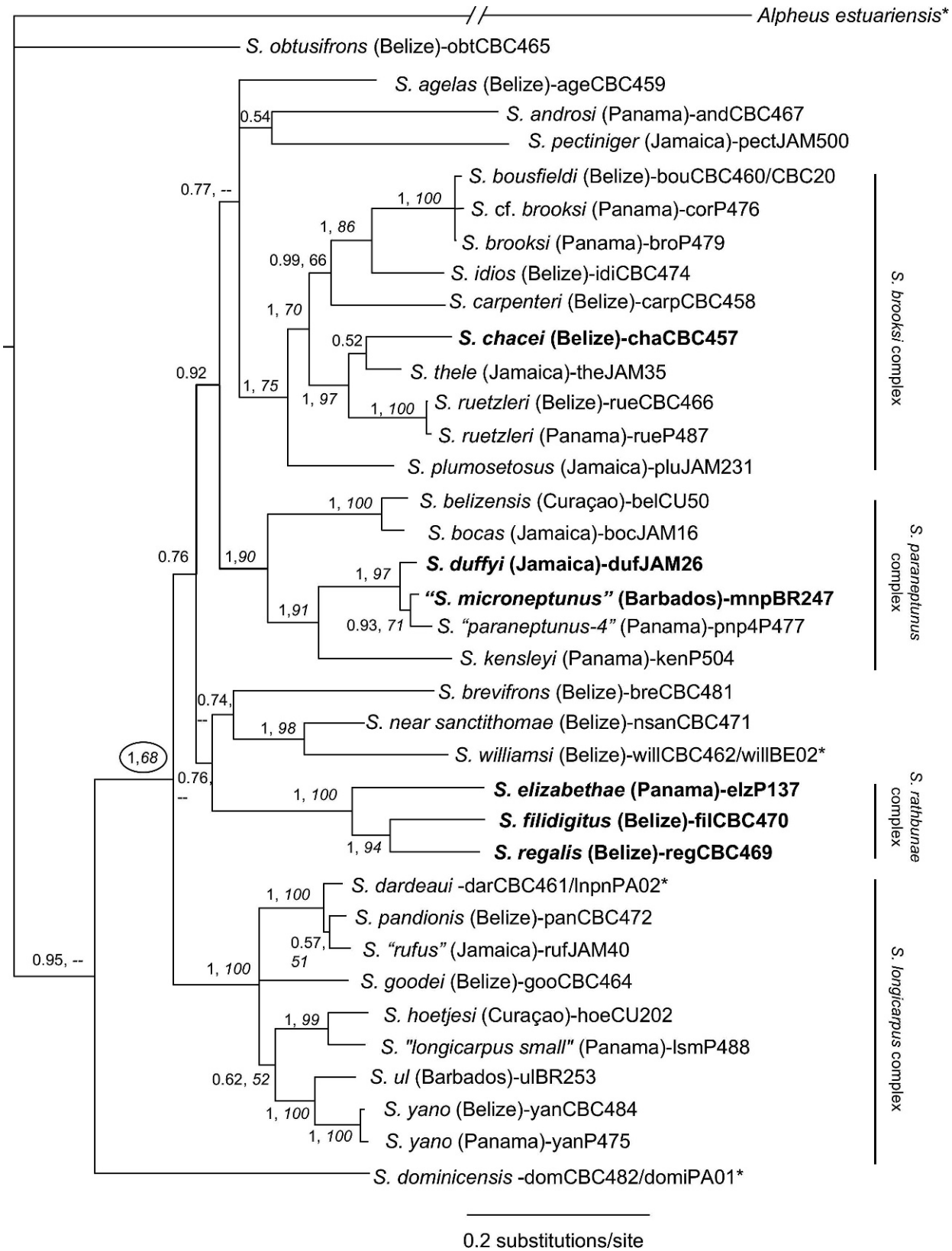

Fig. 1. Phylogenetic tree of West Atlantic sponge-dwelling *Synalpheus*, shown as a Bayesian consensus tree using the ''complete-character'' dataset (16S, COI, EF2, and morphology). Taxon names include species, collection location, and voucher number; species in bold are eusocial; a ''*'' indicates that sequences originated from earlier studies (see on-line Appendix 1). Numbers above or below each node indicate Bayesian posterior probability values or bootstrap support from ML analysis (in italics); for ML bootstrap values, ''--'' indicates that the clade was present but bootstrap confidence is < 50. The circled node confidence value indicates support for the *S. gambarelloides*-group; vertical lines indicate species complexes. *Alpheus estuariensis* is the out-group.

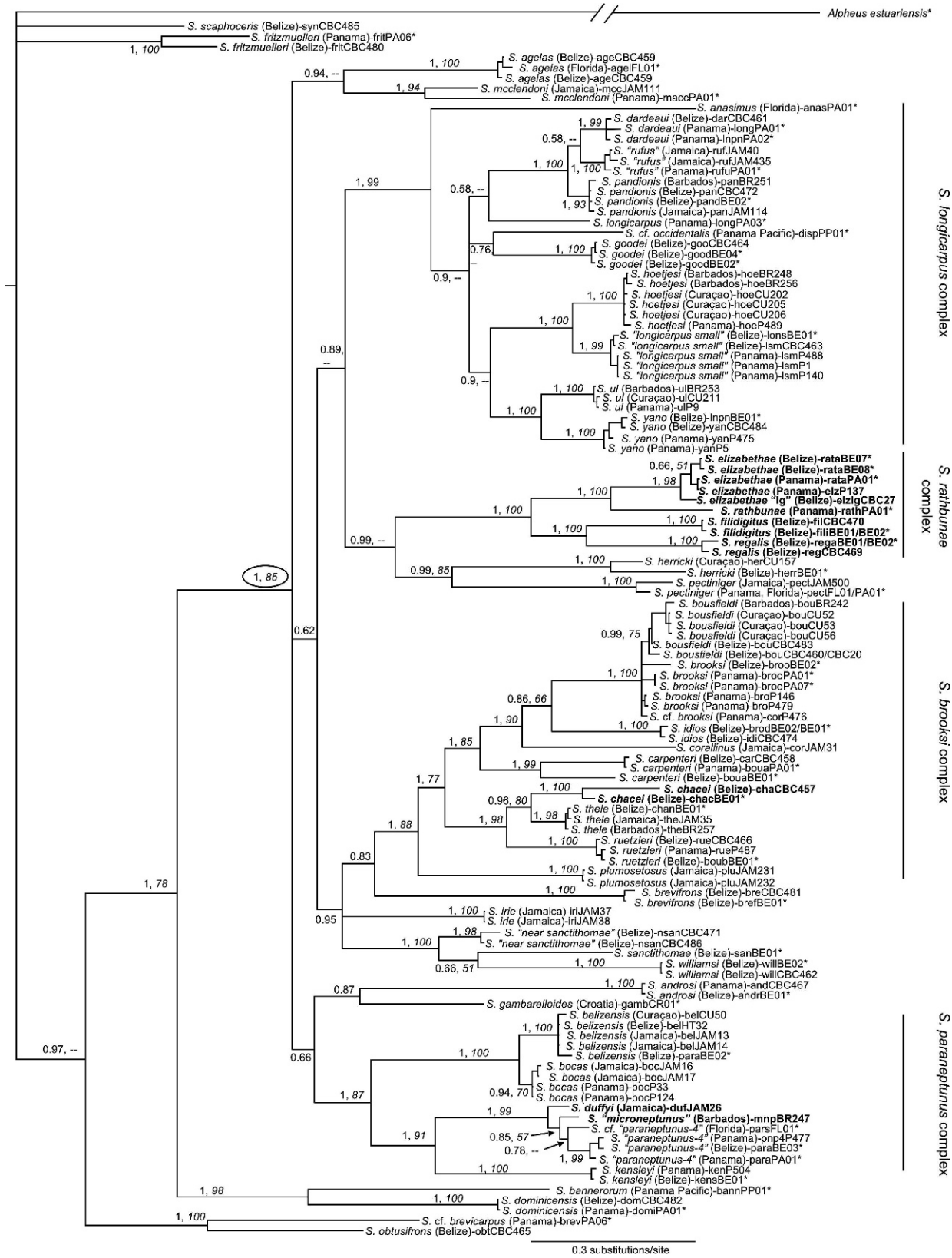

Fig. 2. Bayesian phylogenetic tree of West Atlantic sponge-dwelling *Synalpheus*, using the ''complete-taxa'' dataset (all individuals with at least two of the molecular loci sequenced and the morphological character set). Out-group, vertical lines, and taxon names as in Fig. 1; numbers above or below each node (clade support values) as in Fig. 1; values below the level of species are omitted.

unorganized rows of setae on the minor chela – some morphological similarities with the *Synalpheus coutièrei* Banner, 1953 (formerly ''*biunguiculatus*'')-species group have led some to question whether *S. paraneptunus* is clearly allied to the *S. gambarelloides*-group (Dardeau, 1984; Rios and Duffy, 2007). Furthermore, populations identified as *S. paraneptunus* showed strong variation in several key morphological characters (Dardeau, 1984) and in social structure, ranging from pair-dwelling to living in large, presumably eusocial colonies (Duffy, 2007). Anker and Tóth (2008) completed a preliminary treatment of this species complex, limiting *S. paraneptunus* to the holotype series found in Colombia and describing an additional five species (two of them eusocial) based on the paratype series from Dominica (*Synalpheus riosi,* Anker and Tóth, 2008) and on material from Panama and Belize (*Synalpheus duffyi* Anker and Tóth, 2008, *Synalpheus belizensis* Anker and Tóth, 2008, *Synalpheus bocas* Anker and Tóth, 2008, and *Synalpheus brevidactylus* Anker and Tóth, 2008). However, careful morphological examination and sequencing of additional specimens of *S. paraneptunus* from Belize, Panama, Jamaica, Barbados, and Florida suggests that there are at least two more undescribed species in the complex (provisionally called *S. ''microneptunus''* and *S. ''paraneptunus-4''*), and these data have allowed us to assign the *S. paraneptunus* vouchers from Morrison (2004) to several different species in the complex. In addition to confirming that the *S. paraneptunus* complex is part of the *S. gambarelloides*-group, our tree delineates the close molecular relationships between five of the eight species in this complex, and confirms earlier morphological (Rios and Duffy, 2007) and molecular (Morrison et al., 2004) work suggesting that the species *S. kensleyi* is also closely related to the *S. paraneptunus* complex. Finally, our trees also resolve relationships within several new described and undescribed species in the *S. longicarpus* complex (11 taxa, 6 described in the last 2 decades).

Our analysis also gives strong support for the monophyly of the *S. gambarelloides*-group within the genus *Synalpheus.* Both Bayesian combined-data trees showed support for monophyly of this group (bpp = 1), and substantially stronger support than previous analyses based on combined parsimony with molecular and morphological data (65% bootstrap support) or Bayesian analyses of molecular data (bpp = 0.87; Morrison et al., 2004). Stronger support for the monophyly of this group in our study may have been aided by our inclusion of the nuclear-protein-coding gene EF2; this was the only gene that supported a monophyletic clade of *S. gambarelloides* in single-locus trees (Table 1), suggesting it may have utility in resolving older phylogenetic relationships within *Synalpheus* and Alpheidae in general (Hurt et al., 2009). Strong support for monophyly remained even after removal of ~ 20 morphological characters from the original morphological character set used by Morrison et al. (2004), many of which coded for synapomorphies that distinguished the clade of *S. gambarelloides*. However, unequivocal support for the monophyly of the *S. gambarelloides*-group, as well as the other five subgeneric groups within *Synalpheus* proposed by Coutière (1909), only three of which have been retained by later workers (Banner and Banner, 1975; Dardeau, 1984), awaits a worldwide molecular analysis of the genus *Synalpheus* (Anker and De Grave, 2008).

Perhaps the most important outcome of this work is that it provides a phylogenetic framework for studying other aspects of the social biology of *Synalpheus*, host use, behavior, and speciation in a rigorous comparative context. *Synalpheus* provides a model system for studying the evolution of eusociality in the sea, and previous work has demonstrated that eusociality evolved independently at least three times in this group (Duffy et al., 2000; Duffy, 2007; Duffy and Macdonald, 2010). Interestingly, both combined-data trees in our analysis suggest at least one instance of eusociality being secondarily lost in *Synalpheus;* specifically, the eusocial species *S. duffyi* and *S. ''microneptunus''* appear to be the closest relatives to the undescribed, pair-living species *S. ''paraneptunus-4*.'' The known geographic ranges of these three species do not currently overlap, and they can be clearly distinguished by differences in morphology and social system, despite strong similarities in molecular sequence data. Additional work on the social system and sex ratio of these closely related species is necessary to confirm this finding.

Finally, our new phylogenetic data hint at geographic trends that can help us understand the evolution and diversification of this group. For example, theoretical and empirical work suggests that most speciation is allopatric and that sister species should rarely have overlapping geographic ranges (Barraclough and Vogler, 2000; Fitzpatrick et al., 2009; but see Bolnick and Fitzpatrick, 2007), whereas our complete-taxa tree (Fig. 2) is inconsistent with this prediction. Specifically, this tree indicates that many of the most closely related species in the *S. gambarelloides*-group (in the current phylogeny) currently co-occur in at least part of their range. For example, the sister species *S. bocas* and *S. belizensis* co-occur in Jamaica, where they live in the same sponge species and often co-inhabit the same individual sponges (Macdonald et al., 2009). In other cases, sister species co-occur, but live in different sponge species; for example, *Synalpheus yano* (Rios and Duffy, 2007) and *Synalpheus ul* (Rios and Duffy, 2007) are both common at the same sites in Panama, but typically live in different sponge species (Hultgren and Duffy, unpublished data). Along with several recent site-specific monographs of the distribution of *Synalpheus* and host use throughout the Caribbean (Macdonald et al., 2006, 2009; Rios and Duffy, 2007; Hultgren et al., 2010), these phylogenetic data provide a crucial framework for examining speciation of this hyperdiverse invertebrate group.

Acknowledgements

Funding for this project was provided by a Marine Science Network postdoctoral fellowship from the Smithsonian (to KMH), and by funds from the Smithsonian Institution's Caribbean Coral Reef Ecosystem (CCRE) fund. Collection work at the Smithsonian's Carrie Bow Cay research station in Belize was greatly facilitated by Mike Carpenter and the staff at the Pelican Bay resort; we are also grateful to the staff at the Smithsonian Marine Station in Bocas del Toro, Panama for facilitating collection work there. Tripp Macdonald provided invaluable assistance collecting and identifying specimens in the field and laboratory. Finally, we would like to thank the staff at the Smithsonian's Laboratory of

Analytical Biology (LAB) and Dr. Carla Hurt (University of Miami) for assistance with genetic work.

References

Ahyong, S. T., J. C. Y. Lai, D. Sharkey, D. J. Colgan, and P. K. L. Ng. 2007. Phylogenetics of the brachyuran crabs (Crustacea: Decapoda): The status of Podotremata based on small subunit nuclear ribosomal RNA. Molecular Phylogenetics and Evolution 45: 576-586.

Anker, A., and S. De Grave. 2008. *Zuzalpheus* Rios And Duffy, 2007: A junior synonym of *Synalpheus* Bate, 1888 (Decapoda: Alpheidae). Journal of Crustacean Biology 28: 735-740.

———, and E. Toth. 2008. A preliminary revision of the *Synalpheus paraneptunus* Coutiere, 1909 species complex (Crustacea: Decapoda: Alpheidae). Zootaxa 1915: 1-28.

Banner, A. H. 1953. The Crangonidae, or snapping shrimp, of Hawaii. Pacific Science 7: 3-144.

Banner, D. M., and A. Banner. 1975. The alpheid shrimp of Australia, Part 2: The genus *Synalpheus*. Records of the Australian Museum 29: 267-389.

———, and ———. 1982. The alpheid shrimp of Australia, Part III: The remaining alpheids, principally the genus *Alpheus*, and the family Ogyrididae. Records of the Australian Museum Supplement 34: 1-357.

Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. American Naturalist 155: 419-434.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annual Review of Ecology Evolution and Systematics 38: 459-487.

Bracken, H. D., S. De Grave, and D. L. Felder 2009. Phylogeny of the Infraorder Caridea based on mitochondrial and nuclear genes (Crustacea: Decapoda), pp. 555-580. In, J. Martin, D. L. Felder and K. A. Crandall (eds.), Decapod Crustacean Phylogenetics (Crustacean Issues). CRC Press, Boca Raton.

Bruce, A. J. 1976. Shrimps and prawns of coral reefs, with special reference to commensalism, pp. 37-94. In, O. A. Jones and R. Endean (eds.), Biology and Geology of Coral Reefs. Volume III, Biology 2. Academic Press, New York.

Buhay, J. E. 2009. ''COI-like'' sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29: 96-110.

Chace, F. A. 1972. The shrimps of the Smithsonian-Bredin Caribbean expeditions with a summary of the West Indian shallow-water species (Crustacea: Decapoda:Natantia). Smithsonian Contributions to Zoology 98: 1-179.

Christoffersen, M. L. 1984. The western Atlantic snapping shrimps related to *Alpheus heterochaelis* Say (Crustacea, Caridea), with the description of a new species. Papéis Avulsos de Zoologia, Sâo Paulo 35(19): 189-208.

Coutière, H. 1909. The American species of snapping shrimps of the genus *Synalpheus*. Proceedings of the United States National Museum 36: 1-93.

Dardeau, M. R. 1984. *Synalpheus* shrimps (Crustacea: Decapoda: Alpheidae). I. The *Gambarelloides* group, with a description of a new species. Memoires of the Hourglass Cruises 7: 1-125.

Duda, T. F., and A. J. Kohn. 2005. Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. Molecular Phylogenetics and Evolution 34: 257-272.

Duffy, J. E. 1992. Host use patterns and demography in a guild of tropical sponge-dwelling shrimps. Marine Ecology-Progress Series 90: 127-138.

———. 1996a. Eusociality in a coral-reef shrimp. Nature 381: 512-514.

———. 1996b. Resource-associated population subdivision in a symbiotic coral-reef shrimp. Evolution 50: 360-373.

———. 1996c. Species boundaries, specialization, and the radiation of sponge-dwelling alpheid shrimp. Biological Journal of the Linnean Society 58: 307-324.

———. 1996d. *Synalpheus regalis*, new species, a sponge-dwelling shrimp from the Belize Barrier Reef, with comments on host specificity in *Synalpheus*. Journal of Crustacean Biology 16: 564-573.

———. 1998. On the frequency of eusociality in snapping shrimps (Decapoda: Alpheidae), with description of a second eusocial species. Bulletin of Marine Science 63: 387-400.

———. 2007. Ecology and evolution of eusociality in sponge-dwelling shrimp, pp. 387-409. In, J. E. Duffy and M. Thiel (eds.), Evolutionary ecology of social and sexual systems: crustaceans as model organisms. Oxford University Press, Oxford.

———, and K. S. Macdonald. 2010. Kin structure, ecology and the evolution of social organization in shrimp: a comparative analysis. Proceedings of The Royal Society B-Biological Sciences 277: 575-584.

———, C. L. Morrison, and K. S. Macdonald. 2002. Colony defense and behavioral differentiation in the eusocial shrimp *Synalpheus regalis*. Behavioral Ecology and Sociobiology 51: 488-495.

———, ———, and R. Rios. 2000. Multiple origins of eusociality among sponge-dwelling shrimps (*Synalpheus*). Evolution 54: 503-516.

Edgar, G. J. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2009. Pattern, process and geographic modes of speciation. Journal of Evolutionary Biology 22: 2342-2347.

Herrick, F. H. 1891. *Alpheus*: a study in the development of Crustacea. Memoires of the National Academy of Sciences 5: 370-463.

Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology 51: 673-688.

———, F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294: 310-314.

Hultgren, K. M., and J. E. Duffy. 2010. Sponge host characteristics shape the community structure of their shrimp associates. Marine Ecology-Progress Series 407: 1-12.

———, K. S. Macdonald, and J. E. Duffy. 2010. Sponge-dwelling snapping shrimps of Curaçao, with descriptions of three new species. Zootaxa 2372: 221-262.

Hurt, C., A. Anker, and N. Knowlton. 2009. Identifying the true geminates: a multilocus test of simultaneous divergence across the Isthmus of Panama. Evolution 63: 514-530.

Knowlton, N. 1993. Sibling species in the sea. Annual Review of Ecology and Systematics 24: 189-216.

Leache, A. D., and T. W. Reeder. 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): A comparison of parsimony, likelihood, and Bayesian approaches. Systematic Biology 51: 44-68.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50: 913.

Macdonald, K. S., and J. E. Duffy. 2006. Two new species of sponge-dwelling snapping shrimp from the Belizean Barrier Reef, with a synopsis of the *Synalpheus brooksi* species complex. American Museum Novitates 3543: 1-22.

———, K. M. Hultgren, and J. E. Duffy. 2009. The sponge-dwelling snapping shrimps (Crustacea, Decapoda, Alpheidae, *Synalpheus*) of Discovery Bay, Jamaica, with descriptions of four new species. Zootaxa 2199: 1-57.

———, R. Rios, and J. E. Duffy. 2006. Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. Diversity and Distributions 12: 165-178.

Maddison, D., and W. Maddison. 2005. MacClade 4: Analysis of phylogeny and character evolution. Version 4.08. Sinauer, Sunderland, Massachusetts.

Mathews, L. M., and A. Anker. 2009. Molecular phylogeny reveals extensive ancient and ongoing radiations in a snapping shrimp species complex (Crustacea, Alpheidae, *Alpheus armillatus*). Molecular Phylogenetics and Evolution 50: 268-281.

Morrison, C. L., R. Rios, and J. E. Duffy. 2004. Phylogenetic evidence for an ancient rapid radiation of Caribbean sponge-dwelling snapping shrimps (*Synalpheus*). Molecular Phylogenetics and Evolution 30: 563-581.

Nardo, G. D. 1847. Sinonimia moderna delle specie registrate nell' opera intitolata: Descrizone de' Crostacei, de' Testacei e de' Pesci che abitano le Lagune e Golfo Veneto, rappresentati in figure, a chiaro-scuro ed a colori dall' Abate Chiereghini Ven. Clodiense applicata per commissione governativa, Venezia.

Nylander, J. A. A. 2004. MrModeltest v2. Program distributed by the author. URL http://www.abc.se/~nylander/ [accessed on 7-1-2010]

Palumbi, S. R., F. Cipriano, and M. P. Hare. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55: 859-868.

———, A. Martin, S. Romano, W. O. Mcmillan, L. Stice, and G. Grabowski. 1991. The simple fool's guide to PCR. A collection of PCR protocols. University of Hawaii, Honolulu.

Phillips, A., D. Janies, and W. C. Wheeler. 2000. Multiple sequence alignment in phylogenetic analysis. Molecular Phylogenetics and Evolution 16: 317-330.

Porter, M. L., M. Perez-Losada, and K. A. Crandall. 2005. Model-based multi-locus estimation of decapod phylogeny and divergence times. Molecular Phylogenetics and Evolution 37: 355-369.

Puebla, O., E. Bermingham, and F. Guichard. 2008. Population genetic analyses of Hypoplectrus coral reef fishes provide evidence that local processes are operating during the early stages of marine adaptive radiations. Molecular Ecology 17: 1405-1415.

Regier, J. C. 2008. Protocols, concepts, and reagents for preparing DNA sequencing templates. URL www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf [accessed on 6-1-2009]

Rios, R., and J. E. Duffy. 1999. Description of *Synalpheus williamsi*, a new species of sponge-dwelling shrimp (Crustacea: Decapoda: Alpheidae), with remarks on its first larval stage. Proceedings of the Biological Society of Washington 112: 541-552.

———, and ———. 2007. A review of the sponge-dwelling snapping shrimp from Carrie Bow Cay, Belize, with description of *Zuzalpheus*, new genus, and six new species (Crustacea: Decapoda: Alpheidae). Zootaxa 1602: 3-89.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 20: 407-415.

Sanderson, M. J., and H. B. Shaffer. 2002. Troubleshooting molecular phylogenetic analyses. Annual Review of Ecology and Systematics 33: 49-72.

Simmons, M. P., K. Muller, and A. P. Norton. 2007. The relative performance of indel-coding methods in simulations. Molecular Phylogenetics and Evolution 44: 724-740.

Tsang, L. M., K. Y. Ma, S. T. Ahyong, and T. Y. Chan. 2008. Phylogeny of Decapoda using two nuclear protein-coding genes: origin and evolution of the Reptantia. Molecular Phylogenetics and Evolution.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology 54: 731-742.

———. 2006. Missing data and the design of phylogenetic analyses. Journal of Biomedical Informatics 39: 34-42.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation. University of Texas, Austin, Texas.

———. 2010. GARLI-part v0.97. URL http://www.nescent.org/wg_garli/ [accessed on 11-2-2010 2010]

RECEIVED: 26 July 2010.
ACCEPTED: 11 November 2010.

Supplemental Appendixes on-line

(see DOI: 10.1651/10-3382.1)

Appendix 1. Taxa sampled for molecular analyses. Species complex refers to the morphological species complex to which each taxon belongs; superscripts indicate references that describe species complexes (A = Anker and Toth, 2008; B = Macdonald and Duffy, 2006; C = Macdonald et al., 2009; D = Rios and Duffy, 2007; E = Hultgren et al., 2010). ''CODE'' refers to the unique specimen identifier in GenBank; an asterisk (*) indicates sequences came from previous studies. In ''Location'' column, Panama refers to Caribbean Panama unless otherwise specified.

Appendix 2. Morphological characters used in *Synalpheus* phylogenetic analyses. Character numbers associated with different states correspond to the morphological character matrix (Appendix 3).

Appendix 3. Matrix of morphological data for taxa used in the study.