Gene 786 (2021) 145624

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Comparative mitochondrial genomics of sponge-dwelling snapping shrimps in the genus *Synalpheus*: Exploring differences between eusocial and non-eusocial species and insights into phylogenetic relationships in caridean shrimps

Caroline C. Ennis [a,1], Nariah N. Haeffner [a,1], Cameron D. Keyser [a,1], Shannon T. Leonard [a,1], Austin C. Macdonald-Shedd [a,1], Avery M. Savoie [a,1], Timothy J. Cronin [a], Werner P. Veldsman [b], Phillip Barden [c,d,*], Solomon T.C. Chak [c,e,*], J. Antonio Baeza [a,f,g,*]

[a] Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29634, USA
[b] Simon F. S. Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, SAR, China
[c] Department of Biological Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA
[d] Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA
[e] Department of Biological Sciences, SUNY College at Old Westbury, Old Westbury, NY 11568, USA
[f] Smithsonian Marine Station at Fort Pierce, 701 Seaway Drive, Fort Pierce, Florida 34949, USA
[g] Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Larrondo 1281, Coquimbo, Chile

ARTICLE INFO

Keywords:
Sociobiology
Prawn
Symbiosis
Social behavior
Mitophylogenomics

ABSTRACT

The genus *Synalpheus* is a cosmopolitan clade of marine shrimps found in most tropical regions. Species in this genus exhibit a range of social organizations, including pair-forming, communal breeding, and eusociality, the latter only known to have evolved within this genus in the marine realm. This study examines the complete mitochondrial genomes of seven species of *Synalpheus* and explores differences between eusocial and non-eusocial species considering that eusociality has been shown before to affect the strength of purifying selection in mitochondrial protein coding genes. The AT-rich mitochondrial genomes of *Synalpheus* range from 15,421 bp to 15,782 bp in length and comprise, invariably, 13 protein-coding genes (PCGs), two ribosomal RNA genes, and 22 transfer RNA genes. A 648 bp to 994 bp long intergenic space is assumed to be the D-loop. Mitochondrial gene synteny is identical among the studied shrimps. No major differences occur between eusocial and non-eusocial species in nucleotide composition and codon usage profiles of PCGs and in the secondary structure of tRNA genes. Maximum likelihood phylogenetic analysis of the complete concatenated PCG complement of 90 species supports the monophyly of the genus *Synalpheus* and its family Alpheidae. Moreover, the monophyletic status of the caridean families Alvinocaridae, Atyidae, Thoridae, Lysmatidae, Palaemonidae, and Pandalidae within caridean shrimps are fully or highly supported by the analysis. We therefore conclude that mitochondrial genomes contain sufficient phylogenetic information to resolve relationships at high taxonomic levels within the Caridea. Our analysis of mitochondrial genomes in the genus *Synalpheus* contributes to the understanding of the coevolution between genomic architecture and sociality in caridean shrimps and other marine organisms.

Abbreviations: PCG, protein coding gene.

* Corresponding authors at: Department of Biological Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA (P. Barden), Department of Biological Sciences, SUNY College at Old Westbury, Old Westbury, NY 11568, USA (S. T. C. Chak) and Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA (J. A. Baeza).
E-mail addresses: barden@njit.edu (P. Barden), chakt@oldwestbury.eduy (S.T.C. Chak), jbaezam@clemson.edu (J.A. Baeza).
[1] Equal co-authorship.

https://doi.org/10.1016/j.gene.2021.145624
Received 2 February 2021; Received in revised form 18 March 2021; Accepted 26 March 2021
Available online 31 March 2021

1. Introduction

In the order Decapoda, one of the most speciose clades in the Multicrustacea, caridean shrimps (infraorder Caridea) exhibit remarkable morphological, physiological, ecological, and behavioral diversity (de Grave and Fransen, 2011). Recent studies on Caridea have revealed impressive details of how these innovations have evolved. Examples include multiple origins of 'reciprocally altruistic' behaviors coupled with strong phenotypic convergence in fish-cleaning palaemonid shrimps (Horká et al., 2018), the colonization of semi-terrestrial environments in unison with the ability to jump when out of the water (Abele, 1970, Baeza, 2010), the adoption of a symbiotic lifestyle concomitantly with the evolution of social monogamy and a reduced or reversed sexual dimorphism in body size (females larger than males) (Baeza et al., 2016a, 2016b and references therein), alternative mating strategies evolving in parallel with dissimilar male morphotypes (Karplus and Barki, 2019, Baeza et al., 2014), and the coevolution of non-dispersing larvae and eusociality in shrimps belonging to the genus *Synalpheus* (Duffy and Macdonald, 2010, Chak et al., 2015, 2017 and references therein). Our understanding of the evolutionary ecology of caridean shrimps has improved remarkably over the past decades. Nonetheless, the biology of most of the representatives in this clade remains unknown and genomic resources are still scant, hampering the understanding of the genomic mechanisms driving evolutionary innovations in this clade.

Within the Caridea, shrimps belonging to the genus *Synalpheus* are a cosmopolitan genus of fully marine crustaceans in the intertidal zone and shallow subtidal of most tropical and subtropical regions (Hultgren et al., 2017). Most species in the genus are symbiotic with sponges and other marine invertebrates like echinoderms and cnidarians (Hultgren et al., 2014). Some species live in crowds (aggregations), others in small groups, whereas some species are socially monogamous. Furthermore, at least nine species are known to be eusocial, living in moderate to large groups (10′s to > 100′s shrimps) within the canals of sponges (Duffy, 1996, Hultgren, et al. 2017). In these eusocial species, one or a few female "queens" monopolize reproduction, while the remaining of the male and female "workers" forgo their own reproduction and assume tasks related to colony/sponge defense (Duffy et al., 2002, Tóth and Duffy, 2005, Chak et al., 2015). Eusocial behavior has evolved at least four times independently in the Caribbean Sea and at least once in the Indo-Pacific region (Fransen et al., 2006, Chak et al., 2017). Ongoing studies are using the genus *Synalpheus* as a model system to assess the impact of eusociality on multiple socioecological traits (Chak et al., 2020a, 2020b, 2021 and references therein). Insights into the molecular ecology of *Synalpheus* will boost comparative studies aimed at understanding the genomic underpinnings of complex social behavior in marine invertebrates.

The aim of this study is to describe, in detail, the complete mitochondrial genomes of shrimps belonging to the genus *Synalpheus* and compare major features between eusocial and non-eusocial species. Recently, a comparative study of this genus revealed lower rates of synonymous substitutions and relaxed purifying selection in eusocial taxa relative to non-eusocial species (Chak et al., 2021). This pattern can be explained by a suspected longer generation times in eusocial compared to non-eusocial shrimps (Hultgren et al., 2017) and a reduced effective population size in eusocial species (Chak et al., 2020a, 2020b), which is in line with patterns observed in terrestrial social insects (Chak et al., 2021). Given that mutations are more likely to become fixed in small populations (Ohta and Gillespie, 1996, Chak et al., 2021), we predict that differences in mitochondrial features (i.e., duplication, translocation, or deletion of genes) between shrimps from the genus *Synalpheus* and the crustacean ground pattern should be observed in eusocial but not in non-eusocial species. To explore the effect of sociality on mitochondrial genome features, we examined and compared (i) mitochondrial synteny, (ii) nucleotide composition and codon usage profiles of protein-coding genes (PCGs), (iii) the secondary structure of each identified tRNA gene, and (iv) the putative non-coding control region (i.e., D-Loop) of species with dissimilar social systems. Lastly, based on information provided by mitochondrial PCGs, we examined the phylogenetic position of *Synalpheus* among other species in the infraorder Caridea and explored phylogenetic relationships within the Caridea. This study contributes to a better understanding of the genomic drivers of sociality in caridean shrimps and other marine organisms.

2. Methods

2.1. Sequencing of mitochondrial genomes

We assembled a total of seven mitochondrial genomes belonging to the genus *Synalpheus*; four from non-eusocial species (*S. carpenteri, S. hoetjesi, S. kensleyi,* and *S. pandionis*) and three from eusocial species (*S. chacei, S. filidigitus,* and *S. regalis*). Details on the field collection of each of the studied species can be found in Macdonald et al. (2006). Genomic DNA (gDNA) for each species was extracted from a single alcohol-preserved specimen (one or more pereiopods) using the Qiagen DNeasy Tissue Kits (Qiagen). Extracted DNA was quantified using a Qubit 3.0 Fluorometer with the dsDNA HS assay (ThermoFisher Scientific) and visualized on 2% agarose gels. A total of ~ 1,500 ng of gDNA per specimen was provided to Novogene (Chula Vista, CA) for low-coverage shotgun whole genome sequencing (WGS). The gDNA was used to prepare a library with the TruSeq™ PCR-free Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The resulting adapter-ligated and Illumina-indexed libraries were pooled together with other libraries for multiplexed sequencing on an Illumina NovaSeq sequencer (Illumina, San Diego, CA, USA) using a pair-end (PE) 150 bp run format. Low coverage WGS recently has been confirmed to be an efficient and economical approach to assemble complete mitochondrial genomes in the genus *Synalpheus* (Chak et al. 2020, 2021).

The amount of data generated (available in FASTQ format by the sequencing facility) varied between 6.9 Gb (in *S. hoetjesi*) and 18.6 Gb (in *S. regalis*). The totality of the available data was used for mitochondrial genome assembly in the different studied species.

2.2. Mitochondrial genome assembly of Synalpheus

The mitochondrial genome of each *Synalpheus* species was *de novo*-assembled using the pipeline NOVOPlasty v. 1.2.3 (Dierckxsens et al., 2016). NOVOPlasty uses a seed-and-extend algorithm that assembles organelle genomes from WGS data, starting from a related or distant single 'seed' sequence and an optional 'bait' reference mitochondrial genome (Dierckxsens et al., 2016). For assembly, we used a previously published fragment of the *cox1* gene from each studied species (GenBank accession numbers: *S. carpenteri*: KJ595052, *S. chacei*: AF230792, *S. filidigitus*: AY344695, *S. hoetjesi*: GQ424442, *S. kensleyi*: KJ625039, *S. pandionis*: KJ595134, *S. regalis*: KJ595134) as a seed and a k-mer size ranging between 21 and 39.

2.3. Mitochondrial genome annotation and analysis

The newly assembled mitochondrial genomes were first annotated in the MITOS and MITOS2 webservers (http://mitos.bioinf.uni-leipzig.de) (Bernt et al., 2013) using the invertebrate genetic code. Annotation curation, including start and stop codons corrections, were conducted using ExPASy (https://web.expasy.org) (Artimo et al., 2012) and MEGA X (Kumar et al., 2018). Genome visualization was conducted with GenomeVx (http://wolfe.ucd.ie/GenomeVx/) (Conant and Wolfe, 2008).

Nucleotide composition and codon usage profiles of the PCGs were analyzed. MEGA X was used to estimate nucleotide composition, while codon usage was predicted using the invertebrate mitochondrial code in the Codon Usage webserver (http://www.bioinformatics.org/sms2/

codon_usage.html).

tRNA genes were identified and their secondary structure was predicted in the software MiTFi (Jühling et al., 2012) as implemented in the MITOS web server. The secondary structure of each tRNA was visualized in the webserver Forna (http://rna.tbi.univie.ac.at/forna/) (Kerpedjiev et al., 2015).

The RNAstructure webserver (http://rna.urmc.rochester.edu/RNAstructureWeb/Predict1/Predict1.html) (Reuter and Mathews, 2010) was used to predict the lowest free energy secondary structure of the putative control region, with particular attention placed on the presence of stem-loops in each species.

2.4. Comparison between eusocial and non-eusocial species

We compared mitochondrial gene synteny and homology between eusocial and non-eusocial species using the software BRIG (Alikhan et al., 2011). We also compared the nucleotide and codon usage of the PCGs (all 13 concatenated genes), as well as the nucleotide usage of the 12S and 16S rRNA using a hierarchical clustering analysis with the package pheatmap (Kolde and Kolde, 2015) in RStudio v.4.0.2 (Allaire, 2012). To additionally explore the difference in genome architecture (other than nucleotide use, codon usage in PCGs, and AT content) between eusocial and noneusocial species, we calculated heterodinucleotides (AT, AC, AG, etc.) and homodinucleotides (AA, CC, GG, and TT) frequencies with the UCSC faCount tool (Haeussler et al., 2019). Distributions of short oligonucleotides, including heterodinucleotides and homodinucleotides, are expected to cluster species according to phylogenetic relatedness (as shown in other organisms with small genomes: virus – Ahlgren et al., 2017; bacteria - Karlin et al., 1997). Thus, a clustering among studied species that departs from expected phylogenetic relatedness has the potential to reveal the effect of environmental, including social, conditions on genome architecture. A principal component analysis (PCA) was carried out using the faCount output and the PCA results were visualized with ggfortify in R (Tang et al., 2016). Lastly, nucleotide distribution for each species was calculated and visualized using the DNAvisualization web server (Lee et al., 2019).

2.5. Phylomitogenomics of caridean shrimps

We examined the phylogenetic position of the genus *Synalpheus* among other species of caridean shrimps (Decapoda: Caridea). The newly assembled mitochondrial genomes, along with the mitochondrial genomes of the 86 other caridean species available in the GenBank database, were used for the phylogenetic analysis conducted using the MitoPhAST pipeline (Tan et al., 2015). Outgroups included two species of stenopodid shrimps (*Stenopus hispidus* and *Spongicola levigatus* [infraorder Stenopodidea]) and two species of prawns (*Penaeus vannamei* and *P. monodon* [infraorder Penaeoidea]). MitoPhAST first extracts all 13 PCG nucleotide sequences from the species available in GenBank and any others provided by the user (i.e., *Synalpheus* spp.). The PCG nucleotide sequences are then translated to amino acids, before they are aligned using Clustal Omega (Sievers et al., 2011). Poorly aligned regions are removed with trimAl (Capella-Gutierrez et al., 2009) before the dataset is partitioned and the best fitting models of sequence evolution are selected with ProtTest (Abascal et al., 2005). Finally, the concatenated and partitioned PCG amino acid alignments are used to perform a maximum likelihood phylogenetic tree search in the software IQ-TREE (Nguyen et al., 2015). The robustness of the ML tree topology was ascertained by 1000 bootstrap pseudoreplicates of the tree search.

3. Results and discussion

The NOVOPlasty software pipeline automatically assembled and circularized the mitochondrial genomes of *S. carpenteri, S. chacei, S. kensleyi, S. pandionis,* and *S. regalis* (Genbank accession numbers are available in Supplementary Materials Table S1). The mitochondrial genomes of *S. filidigitus* and *S. hoetjesi* were assembled in two and four overlapping contigs, respectively, and each was circularized manually in MEGAX. Coverage of each mitochondrial genome varied between 44× and 155× (Supplementary Materials Table S1).

In the species examined, the complete mitochondrial genome varied between 15,421 bp (in *S. hoetjesi*) and 15,782 bp (in *S. filidigitus*) in length (Supplementary Materials Table S1). The studied mitochondrial genomes were compact with few intergenic spaces and overlaps among gene junctions (Fig. 1, and Supplementary Materials Table S2). In all species, the mitochondrial genome comprised 13 protein-coding genes (PCGs), two ribosomal RNA genes (*rrnS* [12S ribosomal RNA] and *rrnL* [16S ribosomal RNA]), and 22 transfer RNA (tRNA) genes. A single, long intergenic space ranging in length between 648 bp in *Synalpheus hoetjesi* and 994 bp in *Synalpheus filidigitus* was assumed to be the D-loop/Control Region (Fig. 1, Supplementary Materials Table S2). Mitochondrial gene synteny in all studied species were identical to each other and to the previously reported congeneric shrimp *S. microneptunus* (Chak et al., 2020a, 2020b). Most of the PCGs and tRNA genes were encoded on the light strand, while only four PCGs (in order from 5′ to 3′: *nad5*, *nad4*, *nad4l*, and *nad1*), two rRNA genes, and 8 tRNA genes (*trnF, trnH, trnP, trnL1, trnV, trnQ, trnC,* and *trnY*) were encoded in the heavy strand (Fig. 1, Supplementary Materials Table S2). BRIG results indicated that *S. carpenteri*, *S. kensleyi* and *S. regalis* have similar control regions to *S. microneptunus* (Fig. 2). Conversely, the control regions of *S. pandionis* and *S. hoetjesi* appear to be less similar to *S. microneptunus*, which supports the current systematic status of *Synalpheus* and suggests that the CR region might have phylogenetic informativeness in this genus. Mitochondrial genome synteny in *Synalpheus* also corresponds to the presumed Pancrustacean (Hexapoda + Crustacea) ground pattern (Tan et al., 2017, 2019 and references therein).

Mitochondrial gene order observed in *Synalpheus* spp. is different from that reported in the confamiliar taxa *Alpheus* and *Leptalpheus forceps* – the only other species belonging to the family Alpheidae with a published complete mitochondrial genome (Shen et al., 2012; Qian et al., 2011; Zhong et al., 2019; Wang et al., 2020; Scioli et al., 2020) (Fig. 3). In *Synalpheus* spp., tRNA-E is located between tRNA-S1 and tRNA-F while in *Leptalpheus forceps* and *Alpheus* spp., tRNA-E is located several genes 'downstream', between the PCG *cob* and tRNA-S2. Furthermore, in *Alpheus lobidens*, tRNA-Q is located between *nad4l* and tRNA-T, while in the other studied species, tRNA-Q is located between tRNA-I and tRNA-M. Dissimilarities in mitochondrial synteny between *A. lobidens* and cofamiliar species might be explained by gene duplication, translocation, and deletion of the original duplicated gene (see Wang et al., 2020). Furthermore, the aforementioned comparisons suggest that mitochondrial genome synteny might potentially be useful in revealing phylogenetic relationships among genera in the infraorder Caridea. Gene order dissimilarity, as explored here for the genus *Alpheus*, and its correlated phylogenetic informativeness, needs to be studied in more detail in the infraorder Caridea (Fig. 3).

In all studied species, 12 out of the 13 PCGs exhibited conventional invertebrate and Pancrustacean mitochondrial start codons (ATA, ATG, ATC, ATT) (Supplementary Materials Table S2). In two eusocial species (*S. regalis* and *S. filidigitus*), and in three non-eusocial species (*S. carpenteri, S. kensleyi*, and *S. pandionis*), *cox1* exhibited an alternative putative start codon (ACG) which has been previously reported in the eusocial species, *S. microneptunus* (Chak et al. 2020), as well as in other caridean shrimps (Miller et al., 2005; Kim et al., 2015) and decapod crustaceans (Baeza et al. 2018 and references therein). The majority of PCGs in the species examined ended with a complete and conventional stop codon (TAA or TAG) (Supplementary Materials Table S2). Exceptions included *cob* which terminated with an incomplete stop codon, T, in all studied species but *S. regalis* and *cox3* which terminated with an incomplete stop codon, T, in all studied species but *S. chacei* and *S. hoetjesi*. The other genes which terminated with an incomplete stop codon T were *cox2* in *S. carpenteri*, *nad4l* in *S. regalis*, *nad2* in *S. pandionis,* and *nad1* in *S. regalis* and *S. pandionis.* Truncated stop codons are often

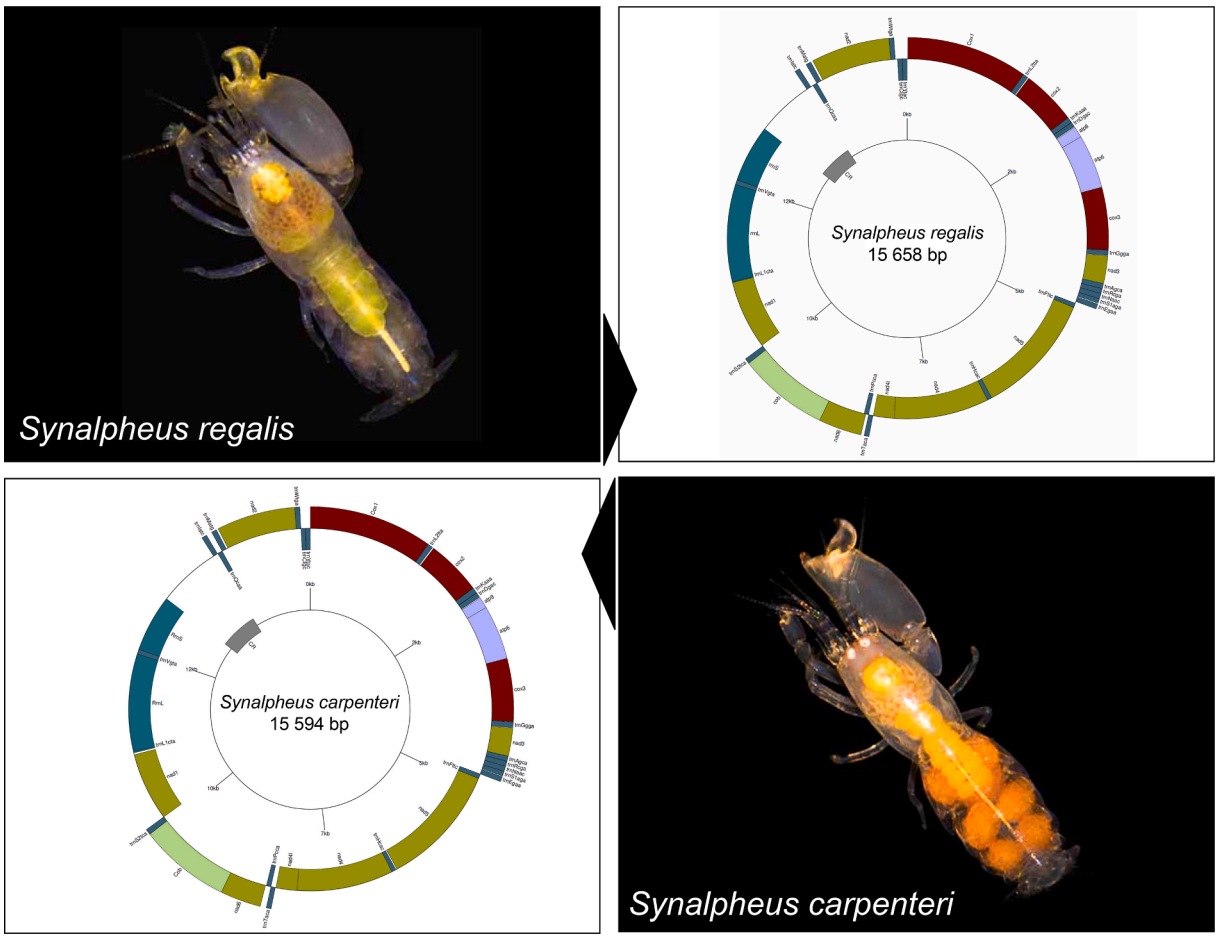

Fig. 1. Circular mitochondrial genome map of the eusocial shrimp, *Synalpheus regalis,* and the non-eusocial shrimp, *Synalpheus carpenteri*. Mitochondrial gene synteny is identical in the studied species. The annotated maps depict 13 protein-coding genes (PCGs), two ribosomal RNA genes (rrnS: 12S ribosomal RNA and rrnL: 16S ribosomal RNA), 22 transfer RNA (tRNA) genes, and the putative control region. The inner circle depicts GC content along the genome. Shrimp depictions modified from McDonald et al. (2009).

observed in mitochondrial genomes of other crustaceans and are hypothesized to be completed via post-transcriptional poly-adenylation (see Baeza, 2018).

The PCGs in the mitochondrial genome of *Synalpheus* spp. exhibited an AT bias with an overall base composition range of A = 27.17–36.00%, T = 34.00–38.69%, C = 16.49–19.52%, and G = 10.99–17.33%. No major dissimilarities in AT biases was observed between eusocial and non-eusocial species as indicated by a hierarchical clustering analysis (Fig. 4). This AT bias is within the known range reported for mitochondrial genomes in caridean shrimps, including the congeneric *S. microneptunus* (Chak et al., 2020a, 2020b). However, the eusocial *S. microneptunus* has the least AT-rich mitogenome while the non-eusocial *S. hoetjesi* has the most AT-rich genome, although these species are from different clades (Supplementary Materials, Fig. S1). Overall, the observed similarity in AT-composition supports the current systematic status of *Synalpheus*: *S. pandionis* and *S. hoetjesi* are nearest each other, and somewhat removed from the other five species in Fig. S1. Even as *S. kensleyi* and *S. microneptunus* are each other's nearest relatives in this analysis, our distribution shows that *S. microneptunus and S. carpenteri* are most similar in nucleotide distribution. Furthermore, *S. filidigitus*, like *S. microneptunus*, has a sharp decrease in GC-content at about the 14 kbp mark in its mitogenome, while the remaining organisms do not exhibit this decrease.

The PCA analysis performed on homodinucleotides measured in all studied species resulted in two principal components representing 87.3% (PC1 = 58.02%, PC2 = 29.28%) of the variability in our dataset. This analysis revealed clustering of the three eusocial species (*S. chacei*, *S. filidigitus* and *S. regalis*) included in our study (Fig. 5). The homodinucleotide distribution of the previously reported eusocial species, *S. microneptunus*, was notably separated from the cluster by the second principal component. It is important to note that the clustering effect was only observed when using homodinucleotides (AA, CC, GG, TT) and that no clustering effect was seen when using all possible dinucleotide combinations (e.g., AC, AG, AT, etc.) Dinucleotide occurrence has been shown to not only be of biochemical importance in DNA conformation, but to also have an association with species-specific properties such as DNA modification, replication and repair (Kariin and Burge, 1995). As such, dinucleotides (referred to as microsatellites when they occur in non-interspersed repetition) can be used as markers to discriminate between organisms. Moreover, DNA methylation at CpG dinucleotides has been implicated in behavioral plasticity in eusocial insects (Yan et al., 2014). This latter observation alludes to the possibility of a functional role for dinucleotides in general, which would include those dinucleotides occurring in mitochondrial sequences.

The most frequently used codons within the PCGs of *Synalpheus* spp. were AT rich and included TTA (most frequently used codon in five species and second most used codon in one species), TTT (most frequently used codon in two species and second most used codon in three species), and ATT (second most frequently used codon in one species and third most used codon in five species). Codons with the lowest frequency (excluding termination codons) included ACG, GCG, and CGG, among a few other (Fig. 4).

The mitochondrial genome of *Synalpheus* spp. encoded tRNA genes that ranged in length from 57 (tRNA-Ser1 in *S. pandionis*, *S. kensleyi*, and

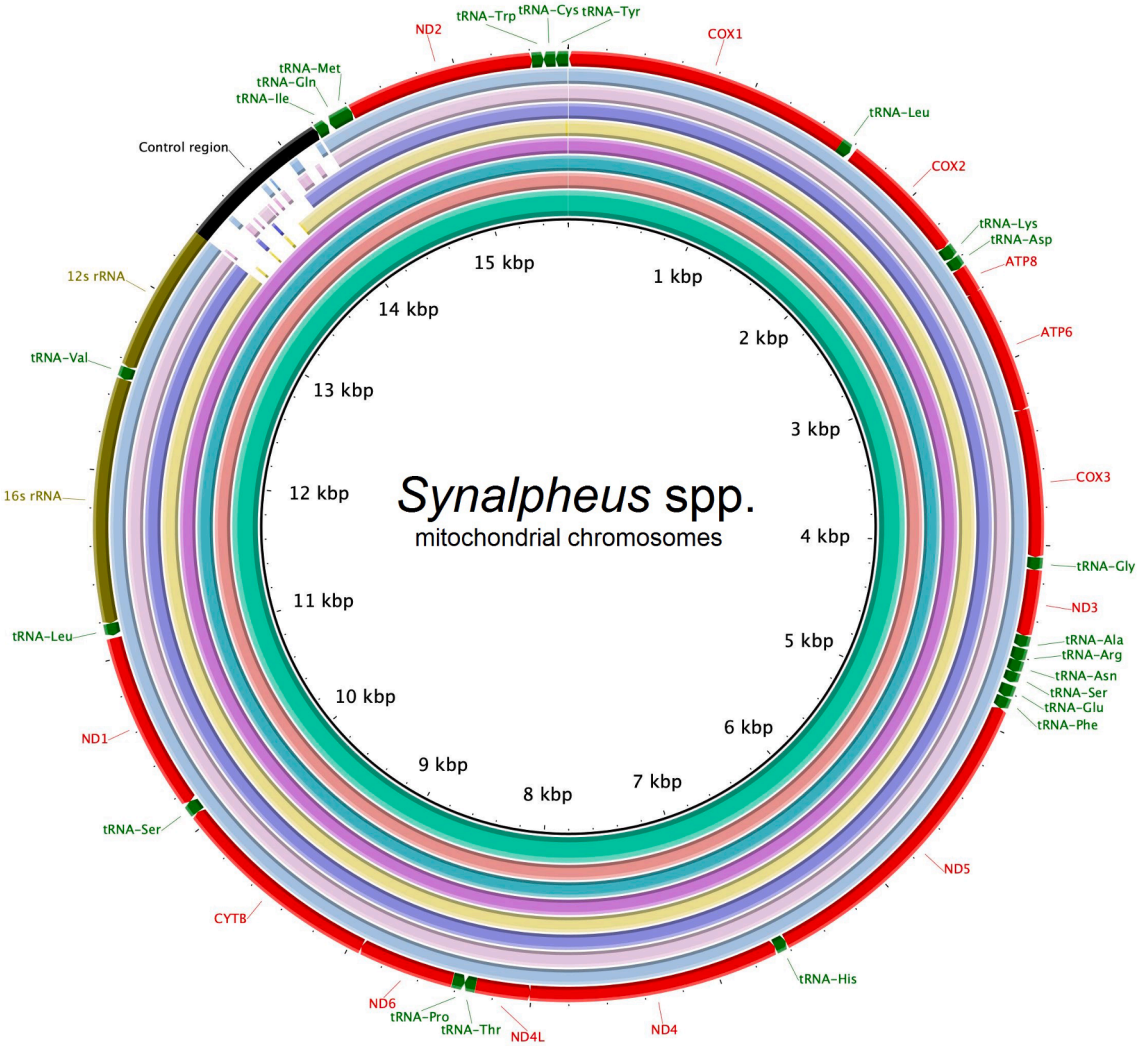

Fig. 2. Mitochondrial genome synteny and homology of shrimp species belonging to the genus *Synalpheus*. Lower homology is observed in the control region. Outer to inner circle species order is: *S. pandionis*, *S. hoetjesi, S. chacei, S. filidigitus, S. kensleyi, S. regalis*, *S. carpenteri,* and *S. microneptunus*.

S. hoetjesi) to 70 bp (tRNA-Ser2 in *S. chacei*, *S. kensleyi*, and *S. hoetjesi* and tRNA-K and in *S. regalis* and *S. filidigitus*). In the species studied here, all tRNA genes, with the exception oftRNA-Ser1, exhibited a standard 'cloverleaf' secondary structure as predicted by MIFTI (Fig. 6). In the tRNA-Ser1 gene, the stem and loop of the pseudouridine arm (T-arm) was missing, in agreement to previous reports for *S. microneptunus* (Chak et al., 2020a, 2020b). Complete (stem and loop) or partial (loop only) tRNA arm deletions are known to occur in other decapod crustaceans (Baeza, 2018 and references therein). An extra elongation factor Tu (EF-Tu) has evolved to support interactions with tRNAs lacking the T-arm during translation (i.e., in other invertebrates and vertebrates - Watanabe et al., 2014).

In the studied species, the *rrnS* gene varied between 792 (in *S. carpenteri*) and 614 bp (in *S. hoetjesi*) while the *rrnL* gene varied between 1,364 (in *S. carpenteri*) and 1,380 bp (in *S. pandionis*) in length. These genes were located close to each other and were found between tRNA-L1 and the putative D-loop/CR, but separated by tRNA-V (Fig. 1). No major differences in nucleotide usage in the *rrnS* and *rrnL* genes were observed between eusocial and non-eusocial species as indicated by a hierarchical clustering analysis (Fig. 4). As shown to occur in other crustaceans, including caridean shrimps, the two genes were highly AT biased. The overall base composition of the *rrnL* gene was A = 29.71–35.26%, T = 38.59–40.42%, C = 6.85–8.27%, and G = 17.57–23.62%. In turn, the *rrnS* gene was A = 29.38–34.61%, T = 35.96–39.23%, C = 8.17–9.65%, and G = 19.90–23.68%.

In the studied *Synalpheus* spp., the intergenic region assumed to be the D-loop/CR varied between 648 (in the noneusocial *S. hoetjesi*) and 994 bp (in the eusocial *S. filidigitus*) in length and was located between the 12S ribosomal RNA and tRNA-I (Fig. 1). The region was heavily AT rich with an overall base composition range of A = 36.35–46.59%, T = 32.57–40.89%, C = 11.97–19.84%, and G = 5.28–9.52%. No major differences in nucleotide use in this putative CR was observed between eusocial and non-eusocial species as indicated by a hierarchical clustering analysis (also, see above and Fig. 4). The secondary structure prediction analysis in RNAstructure (assuming 28 °C and other default options) resulted in a lowest free energy configuration (change in Gibbs free energy [ΔG] range = −120.7 kcal/mol in *S. kensleyi* - − 83.3 kcal/mol in *S. pandionis*) that featured many stem-loop structures interspersed along the length of the region (Suppl. Mat. Fig. S2). The CR structures were in line with previous reports for *S. microneptunus* (Chak et al., 2020a, 2020b) and other putative control regions in other decapod crustacean mitochondrial genomes (Baeza, 2018 and references therein). Overall, studies characterizing the putative D-Loop/CR in crustaceans, including shrimps, are rare, with this study being among the few which has characterized this region in caridean shrimps.

In the ML phylogenetic analysis, the final molecular data matrix comprised a total of 3,584 amino acid characters, of which 1,998 nucleotides were parsimony informative, for a total of 86 ingroup species

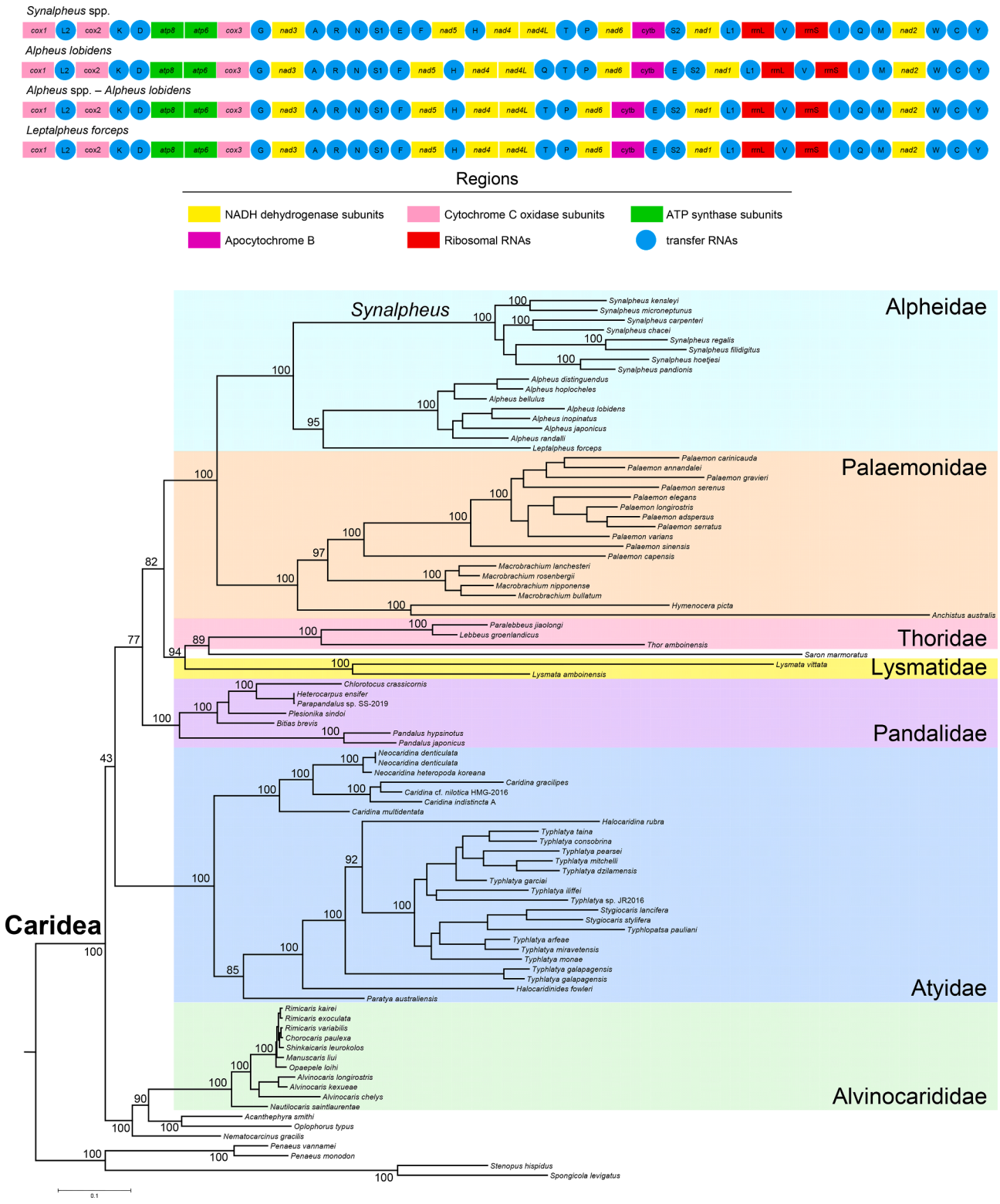

Fig. 3. Mitochondrial gene synteny (above) and phylogenetic tree obtained from ML analysis which was based on a concatenated alignment of amino acids of the 13 protein coding genes present in the mitochondrial genome of representatives of the infraorder Caridea (below). Outgroups included two species of stenopodid shrimps (*Stenopus hispidus* and *Spongicola levigatus*) and two species of prawns (*Penaeus vannamei* and *P. monodon*). Values above or below at each node represents bootstrap support. The analysis included a total of 90 terminals, with 8 species of shrimps belonging to the genus *Synalpheus*, 3584 amino acid characters, and 1998 parsimony informative sites.

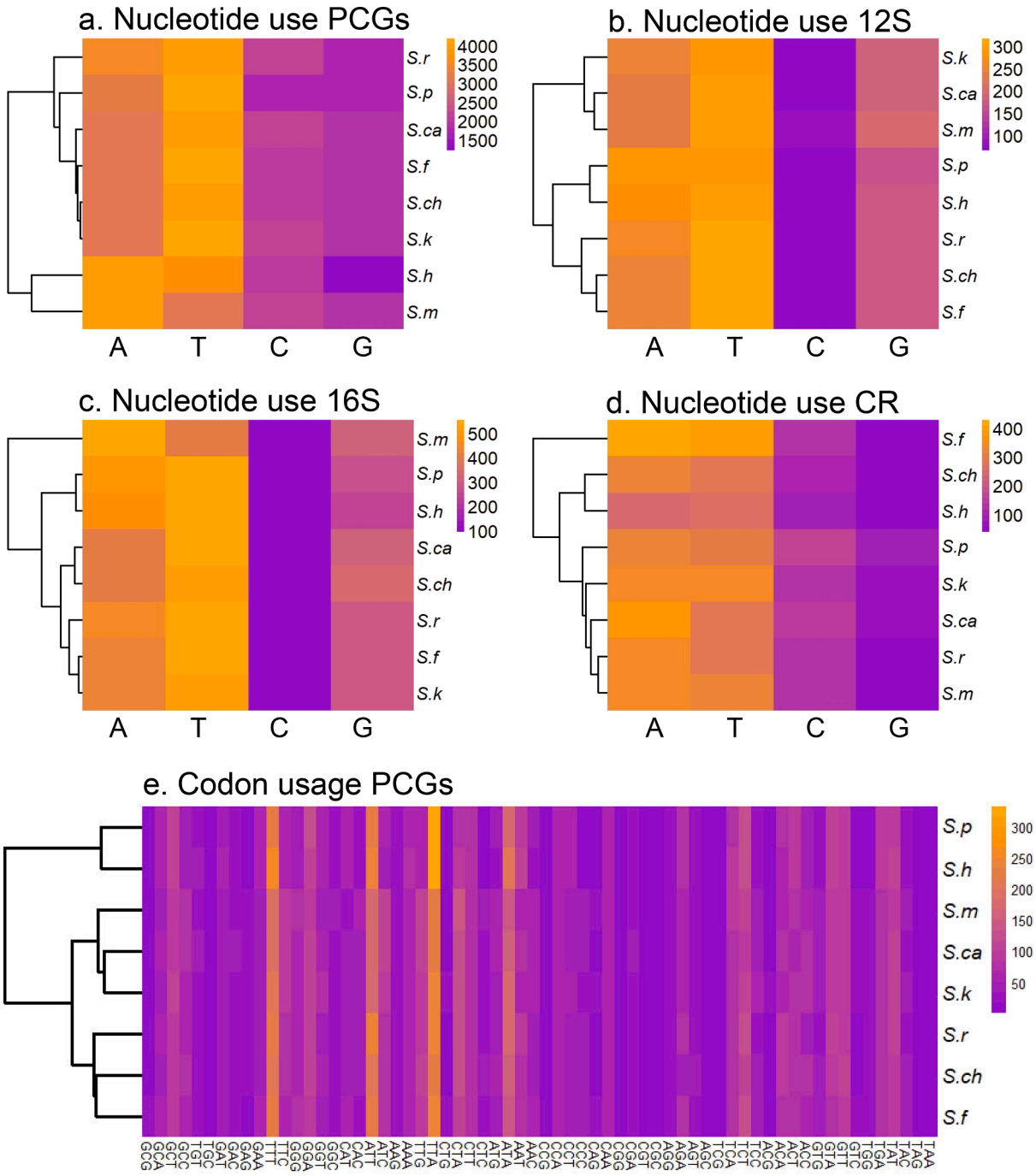

Fig. 4. Heatmap of hierarchically clustered data showing nucleotide use in protein coding genes (a), 12S rRNA DNA (b), 16S rRNA DNA (c), Control Region (d) and the codon usage in the protein coding genes (e). Purple color indicates lower usage and orange color indicates higher usage. Color keys indicate the intensity associated with normalized usage values. *S.p* = *S. pandionis*, *S.h* = *S. hoetjesi*, *S.ch* = *S. chacei*, *S.f* = *S. filidigitus*, *S.k* = *S. kensleyi*, *S.r* = *S. regalis*, and *S.ca* = *S. carpenteri*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

belonging to the infraorder Caridea and four outgroup terminals. Our analysis recovered a monophyletic infraorder Caridea with high support (bootstrap value [bv] = 100). We also recovered a monophyletic Alpheidae with high support comprising the genera *Alpheus*, *Synalpheus*, and *Leptalpheus*. These results are in agreement with previous phylogenetic studies that used either a combination of mitochondrial and nuclear gene fragments (Bracken et al., 2009) (Fig. 2) or complete mitochondrial PCGs with a limited sample of caridean species (Chak et al., 2020a, 2020b). Within Alpheidae, the ML tree also recovered *Synalpheus* as a fully supported monophyletic clade sister to another well supported clade comprised of *Leptalpheus forceps* and its sister, monophyletic genus *Alpheus* (Fig. 2). The monophyletic status of the families Alvinocaridae, Atyidae, Thoridae, Lysmatidae, Palaemonidae, and Pandalidae was fully or highly supported by the analysis. Our results suggest that mitochondrial genomes alone have enough phylogenetic information to reveal relationships at high taxonomic levels within the Caridea.

4. Conclusions

This study assembled and analyzed the mitochondrial genome of both eusocial and non-eusocial shrimp species belonging to the genus

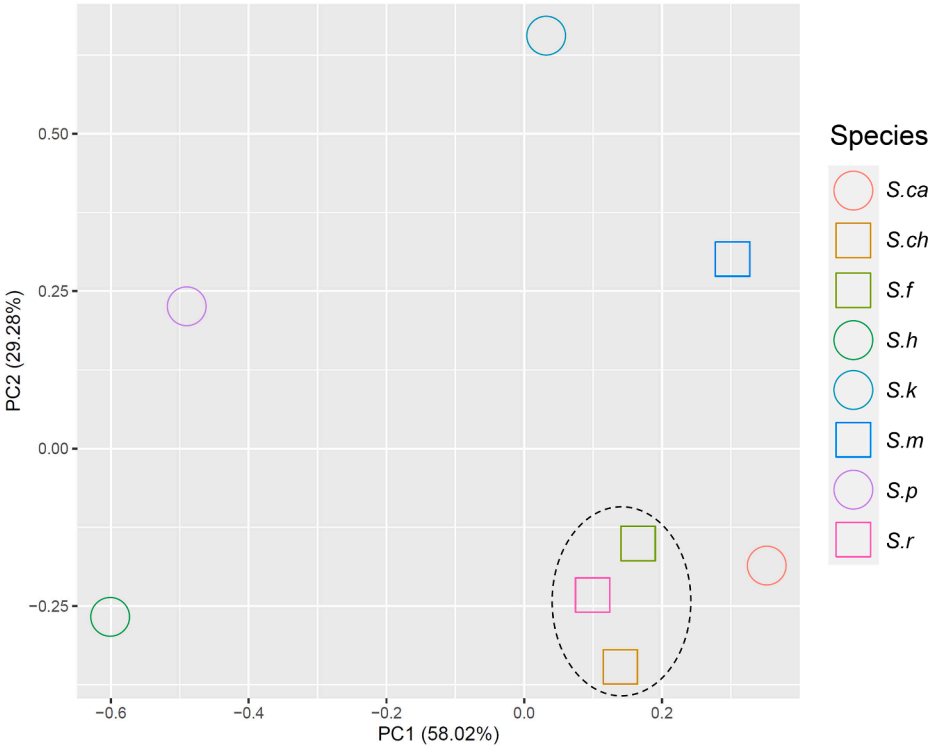

Fig. 5. Principal component analysis based on homodinucleotide (AA, TT, CC, GG) frequency in the mitochondrial genomes of shrimps belonging to the genus *Synalpheus*. Principal components describe the variance in the occurrence of these homodinucleotides among the studied species. Eusocial species, with the exception of *S. microneptunus*, form a noteworthy cluster at the bottom right of the plot. *S.p* = *S. pandionis*, *S.h* = *S. hoetjesi*, *S.ch* = *S. chacei*, *S.f* = *S. filidigitus*, *S.k* = *S. kensleyi*, *S.r* = *S. regalis*, and *S.ca* = *S. carpenteri*.

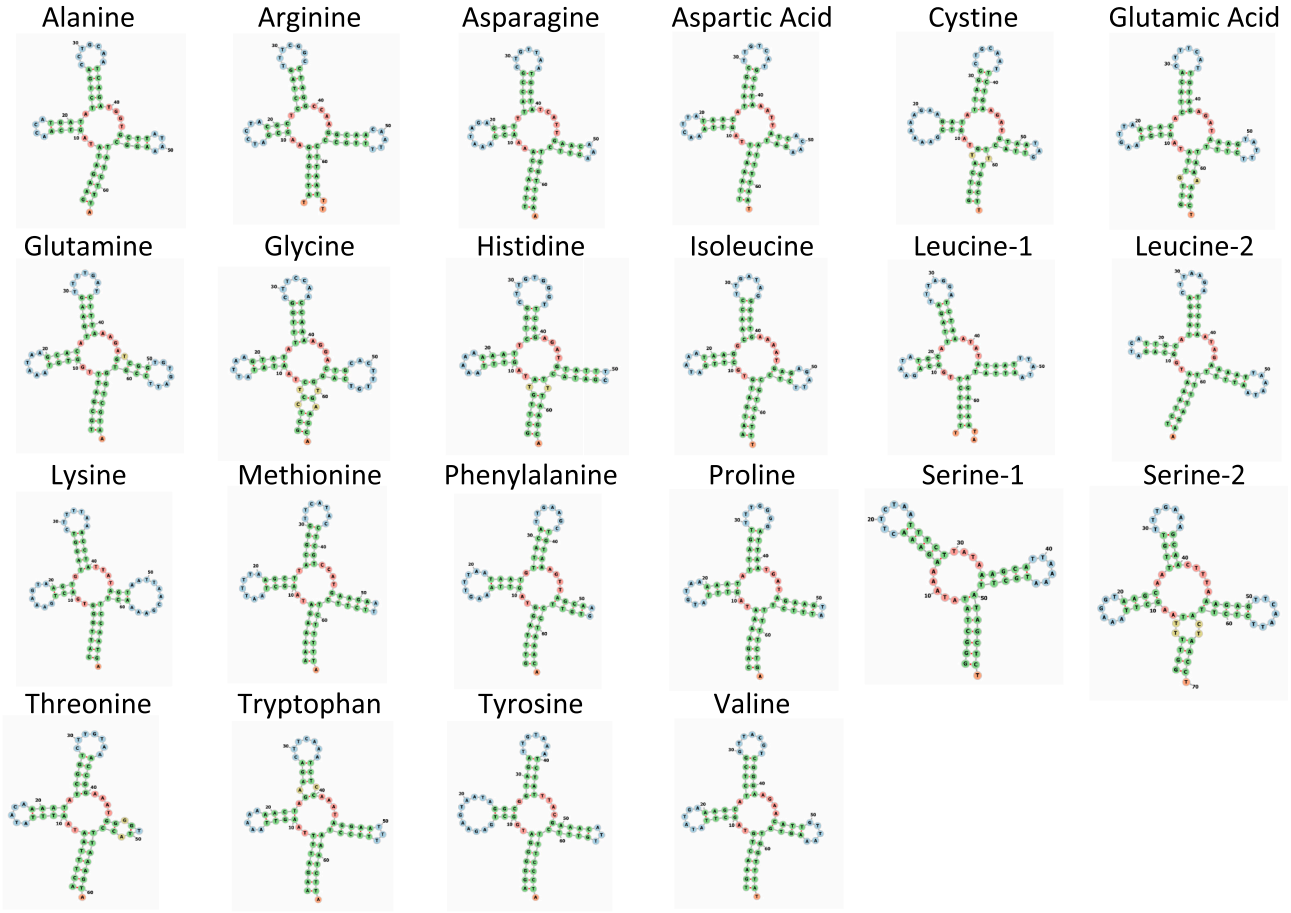

Fig. 6. Secondary structure of tRNAs in the mitochondrial genome of *Synalpheus kensleyi* visualized in the Forna web server. Secondary structure of tRNA genes was similar among the studied species.

Synalpheus. No differences between eusocial and non-eusocial species were observed in terms of mitochondrial gene synteny, nucleotide composition and codon usage profiles of protein-coding genes (PCGs), the secondary structure of each identified tRNA gene, and the putative non-coding control region (i.e., D-Loop). Only minor differences were found when the frequency of homodinucleotides was explored. The subtle dissimilarities in dinucleotide and GC content might reflect minor, but relevant effects of social complexity among the studied species. Indeed, a recent study reported lower synonymous substitution rates and higher potential signals of relaxed purifying selection in eusocial taxa relative to non-eusocial taxa (Chak et al., 2021). The complete mitochondrial genomes we report here will contribute to a better understanding of phylogenetic relationships, environmental drivers, and consequences of sociality in caridean shrimps and other marine organisms. Phylo-mitogenomic analyses provided support for the monophyly of several families within Caridea. This study suggests that entire mitochondrial genomes are useful for resolving phylogenetic relationships at higher taxonomic levels in the Caridea.

5. Ethical approval

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

6. Data availability

Data is available at GenBank (MN787593-MN787599 and SRX6711385-SRX6711392).

Funding

This project was supported by the Clemson University Creative Inquiry program.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

J.A.B. thanks Dr. Vincent P. Richards for bioinformatics support during the development of this project. S.T.T.C. thanks Gedaliah Wolosh for bioinformatics support and staff of IST Academic and Research Computing Systems at NJIT for support in the HPC system.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2021.145624.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105. https://doi.org/10.1093/bioinformatics/bti263.

Abele, L.G., 1970. Semi-terrestrial shrimp (*Merguia rhizophorae*). Nature 226, 661–662.

Ahlgren, N.A., Ren, J., Lu, Y.Y., Fuhrman, J.A., Sun, F., 2017. Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. 45, 39–53.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. BMC Genom. 12, 402. https://doi.org/10.1186/1471-2164-12-402.

Allaire, J., 2012. RStudio: Integrated development environment for R. Boston, MA 770, 394.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40 (W1), W597–W603.

Baeza, J.A., 2010. Observations on the sexual system and the natural history of the semi-terrestrial shrimp *Merguia rhizophorae* (Rathbun, 1900). Invertebr. Biol. 129, 266–276.

Baeza, J.A., 2018. The complete mitochondrial genome of the Caribbean spiny lobster *Panulirus argus*. Sci. Rep. 8 (1), 1–10.

Baeza, J.A., Simpson, L., Ambrosio, L.J., Guéron, R., Mora, N., 2016a. Monogamy in a hyper-symbiotic shrimp. PLoS One 11 (3), e0149797.

Baeza, J.A., Guéron, R., Simpson, L., Ambrosio, L.J., 2016b. Population distribution, host-switching, and chemical sensing in the symbiotic shrimp *Lysmata pederseni*: implications for its mating system in a changing reef seascape. Coral Reefs 35 (4), 1213–1224.

Baeza, J.A., Bauer, R.T., Okuno, J., Thiel, M., 2014. Molecular phylogeny of hinge-beak shrimps (Decapoda: Caridea: *Rhynchocinetes* and *Cinetorhynchus*) and allies: a formal test of familiar and generic monophyly using a multilocus phylogeny. Zool. J. Linn. Soc. 172, 426–450.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69 (2), 313–319.

Bracken, H.D., De Grave, S., Felder, D.L., 2009. Phylogeny of the infraorder Caridea based on mitochondrial and nuclear genes (Crustacea: Decapoda). Decapod Crustacean Phylogenetics 18, 274–298.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chak, S.T.C., Rubenstein, D.R., Duffy, J.E., 2015. Social control of reproduction and breeding monopolization in the eusocial snapping shrimp *Synalpheus elizabethae*. Am. Nat. 186, 660–668.

Chak, S.T.C., Duffy, J.E., Hultgren, K.M., Rubenstein, D.R., 2017. Evolutionary transitions towards eusociality in snapping shrimps. Nat. Ecol. Evol. 1 (4), 1–7.

Chak, S.T., Barden, P., Baeza, J.A., 2020a. The complete mitochondrial genome of the eusocial sponge-dwelling snapping shrimp *Synalpheus microneptunus*. Sci. Rep. 10 (1), 1–10.

Chak, S.T., Barden, P., Baeza, J.A., 2021. Eusociality shapes convergent patterns of molecular evolution across mitochondrial genomes of sponge-dwelling snapping shrimps. In press, Molecular Biology and Evolution.

Chak S.T.C., Harris, S.E., Hultgren, K.M., Duffy, J.E., & Rubenstein, D.R. 2020. Demographic inference provides insights into the extirpation and ecological dominance of eusocial snapping shrimps. bioRxiv:2020.2009.2007.283994.

Conant, G.C., Wolfe, K.H., 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24 (6), 861–862.

Dierckxsens, N., Mardulyn, P., Smits, G., 2016. NOVOPlasty:de novoassembly of organelle genomes from whole genome data. 2016. Nucleic Acids Res. 45 (4), e18 https://doi.org/10.1093/nar/gkw955.

De Grave, S., Fransen, C.H.J.M., 2011. Carideorum catalogus: the recent species of the dendrobranchiate, stenopodidean, procarididean and caridean shrimps (Crustacea: Decapoda). NCB Naturalis, Leiden, p. 195.

Duffy, J.E., 1996. Eusociality in a coral-reef shrimp. Nature 381, 512–514.

Duffy, J.E., Macdonald, K.S., 2010. Kin structure, ecology and the evolution of social organization in shrimp: a comparative analysis. Proc. Royal Soc. B: Biol. Sci. 277, 575–584.

Duffy, J.E., Morrison, C.L., Macdonald, K.S., 2002. Colony defense and behavioral differentiation in the eusocial shrimp *Synalpheus regalis*. Behav. Ecol. Sociobiol. 51, 488–495.

Fransen, C., Didderen, K., de Voogd, N., 2006. Observations on sponge-dwelling colonies of *Synalpheus* (Decapoda, Alpheidae) of Sulawesi, Indonesia. Crustaceana 79, 961–975.

Horká, I., De Grave, S., Fransen, C.H., Petrusek, A., Ďuriš, Z., 2018. Multiple origins and strong phenotypic convergence in fish-cleaning palaemonid shrimp lineages. Mol. Phylogenet. Evol. 124, 71–81.

Hultgren, K.M., Hurt, C., Anker, A., 2014. Phylogenetic relationships within the snapping shrimp genus *Synalpheus* (Decapoda: Alpheidae). Mol. Phylogenet. Evol. 77, 116–125.

Hultgren, K. M., Duffy, J. E. & Rubenstein, D. R. 2017. Sociality in shrimps. in Comparative Social Evolution (eds D.R. Rubenstein & P. Abbot) 224-249 (Cambridge University Press, 2017).

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40, 2833–2845.

Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C. M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., Gibson, D., 2019. The UCSC genome browser database: 2019 update. Nucleic Acids Res. 47 (D1), D853–D858.

Kariin, S., Burge, C., 1995. Dinucleotide relative abundance extremes: a genomic signature. Trends Genet. 11, 283–290.

Karplus, I., Barki, A., 2019. Male morphotypes and alternative mating tactics in freshwater prawns of the genus *Macrobrachium*: A review. Rev. Aquacul. 11, 925–940.

Kim, S.-J., Pak, S.J., Ju, S.-J., 2015. Mitochondrial genome of the hydrothermal vent shrimp *Nautilocaris saintlaurentae* (Crustacea: Caridea: Alvinocarididae). Mitochondrial DNA 26, 127–128.

Kerpedjiev, P., Hammer, S., Hofacker, I.L., 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics 31, 3377–3379. https://doi.org/10.1093/bioinformatics/btv372.

Kolde, R., Kolde, M.R., 2015. Package 'pheatmap'. R Package 1 (7), 790.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35, 1547-1549, doi:10.1093/molbev/msy096.

Lee, B.D., Timony, M.A., Ruiz, P., 2019. DNAvisualization.org: a serverless web tool for DNA sequence visualization. Nucleic Acids Res. 47, W20–W25.

Macdonald, K.S., Rios, R., Duffy, J.E., 2006. Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. Divers. Distrib. 12 (2), 165–178. https://doi.org/10.1111/ddi.2006.12.issue-210.1111/j.1366-9516.2005.00213.x.

Miller, A.D., Murphy, N.P., Burridge, C.P., Austin, C.M., 2005. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). Mar. Biotechnol. 7 (4), 339–349. https://doi.org/10.1007/s10126-004-4077-8.

Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., Minh, B., 2015. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Ohta, T., Gillespie, J.H., 1996. Development of neutral and nearly neutral theories. Theor. Popul Biol. 49, 128–142.

Qian, G., Zhao, Q., Wang, A.N., Zhu, L.I.N., Zhou, K., Sun, H., 2011. Two new decapod (Crustacea, Malacostraca) complete mitochondrial genomes: bearings on the phylogenetic relationships within the Decapoda. Zool. J. Linn. Soc. 162, 471–481.

Reuter, J.S., Mathews, D.H., 2010. RNAstructure: Software for RNA secondary structure prediction and analysis. BMC Bioinf. 11, 129. https://doi.org/10.1186/1471-2105-11-129.

Scioli, J.A., Plouviez, S., Felder, D.L., 2020. The complete mitochondrial genome of the symbiotic infaunal snapping shrimp *Leptalpheus forceps* (Decapoda, Alpheidae). Mitochondrial DNA Part B 5, 629–630.

Shen, X., Li, X., Sha, Z., Yan, B., Xu, Q., 2012. Complete mitochondrial genome of the Japanese snapping shrimp *Alpheus japonicus* (Crustacea: Decapoda: Caridea): Gene rearrangement and phylogeny within Caridea. Sci. China Life Sci. 55 (7), 591–598. https://doi.org/10.1007/s11427-012-4348-1.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7 (1), 539. https://doi.org/10.1038/msb:2011.75.

Karlin, S., Mrazek, J., Campbell, A.M., 1997. Compositional biases of bacterial genomes and evolutionary implications. J. Bacteriol. 179, 3899–3913.

Tan, M.H., Gan, H.M., Schultz, M.B., Austin, C.M., 2015. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. Mol. Phylogenet. Evol. 85, 180–188.

Tan, M. H., Gan, H. M., Lee, Y. P., Poore, G. C. B. & Austin, C. M. 2017. Digging deeper: new gene order rearrangements and distinct patterns of codons usage in mitochondrial genomes among shrimps from the Axiidea, Gebiidea and Caridea (Crustacea: Decapoda). PeerJ 5, e2982, doi:10.7717/peerj.2982.

Tan, M.H., Gan, H.M., Lee, Y.P., Bracken-Grissom, H., Chan, T.Y., Miller, A.D., Austin, C. M., 2019. Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. Sci. Rep. 9, 10756. https://doi.org/10.1038/s41598-019-47145-0.

Tang, Y., Horikoshi, M., Li, W., 2016. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. R J. 8 (2), 474. https://doi.org/10.32614/RJ-2016-060.

Tóth, E., Duffy, J.E., 2005. Coordinated group response to nest intruders in social shrimp. Biol. Lett. 1, 49–52.

Wang, Q., Wang, Z., Tang, D., Xu, X., Tao, Y., Ji, C., Wang, Z., 2020. Characterization and comparison of the mitochondrial genomes from two Alpheidae species and insights into the phylogeny of Caridea. Genomics 112, 65–70.

Watanabe, Y.I., Suematsu, T., Ohtsuki, T., 2014. Losing the stem-loop structure from metazoan mitochondrial tRNAs and co-evolution of interacting factors. Front. Genet. 5, 109. https://doi.org/10.3389/fgene.2014.00109.

Yan, H., Simola, D.F., Bonasio, R., Liebig, J., Berger, S.L., Reinberg, D., 2014. Eusocial insects as emerging models for behavioural epigenetics. Nat. Rev. Genet. 15, 677–688.

Zhong, S., Zhao, Y., Zhang, Q., 2019. The complete mitochondrial genome of *Alpheus hoplocheles* (Decapoda: Alpheidae). Mitochond DNA Part B 4, 189–190.