Novel molecular resources for single-larva barcoding of enigmatic crustacean y-larvae

Niklas Dreyer[1], Ferran Palero[2], Mark Grygier[3], Alexandra Savchenko[4], Gregory Kolbasov[5], Ryuji Machida[6], Benny Chan[6], and Jørgen Olesen[7]

[1]Affiliation not available
[2]5Institut Cavanilles de Biodiversitat i Biologia, Evolutiva (ICBIBE)
[3]National Taiwan Ocean University Center of Excellence for the Oceans
[4]Lomonosov Moscow State University
[5]Moskovskij gosudarstvennyj universitet imeni M V Lomonosova Biologiceskij fakul'tet
[6]Academia Sinica
[7]Natural History Museum of Denmark

December 6, 2022

Abstract

The enigmatic "y-larvae" (Pancrustacea: Facetotecta) still have an incompletely understood lifecycle, and their adult forms remain unknown despite their discovery more than 100 years ago and their documented global occurrence from shallow waters to the deep-sea. Only two of the 17 formally described species, all based on larval stages, have been investigated using an integrative taxonomic approach that, besides providing descriptions of the morphology of the naupliar and cyprid stages, also made use of exuvial voucher material and DNA barcodes. To improve our knowledge about the systematics and phylogenetics of y-larvae, we developed a novel protocol that maximizes the amount of morphological, ecological, and molecular data that can be harvested from single individuals of these tiny larvae. This revolves around single larva barcoding, and includes daily imaging of y-nauplii reared in culture dishes, mounting of their last naupliar exuviae on a slide as a reference voucher, live imaging of the y-cyprid instar that follows, and fixation, DNA extraction, amplification, and sequencing of the y-cyprid specimen. By developing and testing a suite of new primers for both nuclear and mitochondrial protein-coding and ribosomal genes, we estimated the most comprehensive phylogeny of Facetotecta to date. We expect that our novel procedure will help to unravel the complex systematics of y-larvae and show how these fascinating larval forms have evolved. Moreover, we posit that our protocols should work on larval specimens of a diverse array of molting marine invertebrate taxa.

1. INTRODUCTION

Increased availability of diverse phylogenomic resources has transformed pancrustacean (Hexapoda and 'Crustacea') systematics, yet numerous clades remain recalcitrant to analysis due to a critical lack of molecular resources (Regier et al., 2010; von Reumont et al., 2012; Oakley et al., 2013; Schwentner et al., 2017, 2018; Lozano-Fernández et al., 2019; Bernot et al., 2022). This particularly applies to the so-called 'dark' taxa (Hartop et al., 2022), of which the enigmatic crustacean y-larvae (Pancrustacea: Thecostraca: Facetotecta) constitute one of the most remarkable examples.

Y-larvae are fascinating invertebrates that are almost exclusively known from planktonic larval stages (Glenner et al., 2008; Dreyer et al., 2022). They develop through a series of dispersive stages (y-nauplii: Itô, 1986; Kolbasov et al., 2021; Olesen et al., 2022) and a single putative attachment stage (y-cyprid: Kolbasov et al., 2022). A subsequent stage, called the ypsigon, which lacks segmented appendages, a gut, and compound eyes, was recently induced through *in vivo*exposure of y-cyprids to crustacean molting hormone (Glenner

et al., 2008; Pérez-Losada et al., 2009; Dreyer et al., 2022). As the ypsigon exites the y-cyprid, it is tentatively regarded as an early instar/juvenile of a hypothetical endoparasitic adult stage (Pérez-Losada et al., 2009; Dreyer et al., 2022). The y-larva life cycle thus envisaged is reminiscent of that of parasitic barnacles (Cirripedia: Rhizocephala; Dreyer et al., in press).

Few studies have cast light on the systematics and evolution of facetotectans, despite their discovery more than a century ago and their global distribution at depths of 0-6000m (Hansen, 1899; Grygier, 1987; Dreyer et al., 2022; Kolbasov et al., 2022). Morphological studies based on wild-caught specimens have proven unsuccessful in delimiting species within geographically widespread "types" such as type IV (Hansen, 1899; Schram, 1972), "Pacific Type I" (Itô, 1986), and type VIII with its three subtypes VIII-a, -b, and -c (Itô, 1987), which are each practically identical wherever found. Formal cladistic analyses using morphological characters have not resolved facetotectan phylogeny (Pérez-Losada et al., 2009; Kolbasov et al., 2022), and the need for DNA sequences of live-imaged and properly vouchered material is inescapable (Olesen et al., 2022). Phylogenetic analyses show that y-larvae form a distinct and monophyletic group (Grygier, 1987; Chan et al., 2021; Dreyer et al., 2022), but conflicting datasets currently place Facetotecta as sister either to Ascothoracida (parasitic "gall barnacles": Petrunina et al., 2014; Dreyer et al., 2022) or a monophyletic group composed of Ascothoracida and Cirripedia (the stalked, acorn, burrowing, and parasitic barnacles; see Pérez-Losada et al., 2009). These studies are all problematic in that they have mostly used unvouchered "ghost" sequences of y-larvae or only a few quite conservative molecular markers (Pérez-Losada et al., 2009; Petrunina et al., 2014; Dreyer et al., 2022). No facetotectan species description has been supplemented by both nuclear and mitochondrial sequence data, and a complete or near-complete series of naupliar and cyprid instars has been described for only three species to date (viz., *Hansenocari furcifera* Itô, 1989, *H. itoi* Kolbasov and Høeg, 2003, and*H. demodex* Olesen, Dreyer, Palero and Grygier in Olesen et al., 2022). The incompletely known life cycle of y-larvae, their small size, and a dearth of molecular data have hampered an accurate assessment of their evolution and systematics.

The current systematic resolution of y-larvae is unsatisfactory and problematic for several reasons. Considering that their local and global diversity may be significantly larger than what is described (Hansen 1899; Glenner et al., 2008; Dreyer et al., in press), authentic sequences of vouchered and named species are essential for biodiversity inventories. Currently only 17 species are formally described, in the single genus *Hansenocaris* Itô, 1985 (Olesen et al., 2022; Olesen and Grygier, 2022). Solving phylogenetic relationships is, moreover, crucial for robust macroevolutionary modeling of life histories and ecomorphological traits, especially when compelling life-history evidence suggests that y-larvae attain maturity as endoparasites in unknown hosts (Glenner et al., 2008; Pérez-Losada et al., 2009; Dreyer et al., 2022).

Destructive DNA-extraction methods often result in complete digestion of the specimen, which limits adequate morphological species delimitation and museum storage of vouchers. Here, we present novel molecular resources and optimized protocols for successful DNA extraction, voucher exuvium retainment, and PCR-amplification of single-specimen y-larvae below 500 µm in size. This is intended to supplement the single-specimen rearing and live-imaging protocol previously developed by us (Olesen et al., 2022). Through extensive laboratory experimentation, we developed an optimal DNA-extraction and PCR-amplification protocol for Facetotecta and here compare it to two other methods that can also yield single-specimen nucleotide sequences. To this end, and as part of a larger campaign investigating the phylogeny and evolution of Facetotecta (Olesen et al., 2022; Dreyer et al., submitted), we designed novel oligonucleotide primers to amplify nuclear and mitochondrial genes and tested 28 primer pairs in more than 2000 PCR reactions. We thereby show that up to 6700 aligned nucleotide resolution can be achieved for a single specimen. We demonstrate the utility of our protocol by estimating a preliminary phylogeny of Facetotecta that expands upon previous phylogenetic analyses. This estimate includes more markers (n=6) and specimens (n=74) than previous efforts, using material obtained from Pacific, Northeast Atlantic, Antarctic (NCBI data), and Arctic waters. Finally, we discuss the importance of our protocols for further advancing the knowledge of y-larva systematics and evolution.

2. MATERIALS & METHODS

2.1 Collections, rearing, imaging & digital sorting

About 11,000 y-larvae were collected with 75-µm-mesh plankton nets from surface waters (0-5m) at three sites: the University of the Ryukyus Sesoko Laboratory on Sesoko Island, Okinawa, Japan (JA) and Gongguan Harbor on Green Island and WangHaiXiang Fishing Harbor in Keelung, both in Taiwan (TA) (Olesen et al., 2022; Fig. 1; Table S1); 35 of these were sequenced and included in this work. Supplementary material came from Tioman Island (Malaysia; MAL), Piscinas do Pesquerio in Ponta Delgada, Azores, Portugal (AZ, n=10), and waters off the White Sea Marine Biological Station, Russia (RU; n=14) between 2017 and 2021. Larvae were sorted from plankton samples with glass Pasteur pipettes under an Olympus stereomicroscope and either (1) photographed and video-recorded live with a Nikon ECLIPSE 80i compound microscope (JA) or either an Olympus IX70 inverted compound microscope or a Zeiss AX10 light microscope (TA), both of the latter being equipped with Nomarsky (DIC) optics and a Canon EOS 5D Mark IV digital camera, or (2) bulk-fixed for subsequent molecular work (TA, RU, AZ). All images were sorted digitally to identify groups of morphologically similar specimens (i.e., morphotypes; Olesen et al., 2022), so that at least 2-3 specimens per group were used for Sanger sequencing.

2.2 Primer design, DNA extraction, PCR, and sequencing

Reference nucleotide sequences (Ascothoracida and Cirripedia) of selected loci were downloaded from NCBI GenBank and aligned with in-house (see below) or NCBI Facetotecta sequences with MAFFT version 7.450 (–auto –maxiterate 1000; Katoh and Standley 2013). The alignments were imported, visualized, and used to design primers in Geneious Prime version 2022.0.2. To enhance specificity and amplification success, we strived to design longer (>20bp) primers with GC-contents between 30-60%, a GC-clamp of 2 (3' terminating in at least one G/C to promote optimal binding), a melting temperature (T_m) between 55°C-68°C, and no runs of >4 bases (homopolymeric regions, e.g., TGGGGG). We tested a total of 28 primer pairs (23 new) and performed 2056 PCR reactions using these. Tables 1, 2, and S2 provide an overview of the primers, their sequences, and their amplification success with single specimens.

Our goal was to develop and test the efficacy of fast "filter-free" DNA-extraction methods that amplify single-specimen y-larvae while retaining their "exuviae" (here understood as any cuticular remains originating from an acutal molt of after tissue digestion during DNA extraction) as vouchers. We therefore tested and compared DNA-extraction using the GeneReleaser® kit (BioVentures, TN, USA), which has previously proven to be an efficient protocol for retrieving DNA from zooplankton (Schizas et al., 1997; Böttger-Schnack and Machida, 2011; Watanabe et al., 2016; Olesen et al., 2022) and a simplified DNeasy method using a subset of the reagents in the Blood and Tissue kit (QIAGEN, CA, USA). To this end, and as a part of a larger phylogenetic and barcoding campaign, we extracted DNA of a total 421 single y-larva specimens, of which nine were from the Azores, 326 from Sesoko Island (Japan), 4 from Malaysia, 105 from various locations in Taiwan, and 17 from the White Sea (Russia).

DNA extracts were kept in a freeze Eppendorf© vial rack during preparation of the PCR assay. We pipetted 12.2 µL filtered ddH_2O, 5 µL "hot start" HOT FirePol® polymerase master mix (Solis BioDyne, Tartu, Estonia), 0.4 µL of each oligonucleotide primer (usually ordered as "desalted"), and 3 µL gDNA template to each vial. For multiplex PCR reactions, which use two or more primer pairs in combination, we subtracted the equivalent amount of added primer from the amount of ddH_2O. To increase primer annealing sensitivity, specificity to complex templates (e.g., >60% G-C content), and yield we generally used a Touchdown PCR profile under the following conditions: initial denaturation for 15min at 95°C, then 10 cycles of denaturation at 95°C for 1min, annealing at T_m+10°C decreasing by 1°C/cycle for 30 seconds, extention at 72°C for 1kb/60secs, then 30 cycles of denaturation at 95°C for 1min, annealing at T_m-2°C for 30 seconds, extension at 72°C for 1kb/60secs, final extension at 72°C for 10min, and finally 20min at 4°C to halt the reaction. Primer details including sequences and annealing temperatures is listed in Supplementary Table 1.

All PCR reactions were conducted in a DNA Engine Thermal Cycler (Bio-Rad, Richmond, CA, United States) and PCR products were visualized in agarose gel with varying concentrations depending on fragment size. DNA sequencing was performed by Genomics BioSci & Tech Ltd. (New Taipei City, Taiwan) and

Macrogen Europe BV (Amsterdam, The Netherlands).

We sequenced genomic DNA templates of three ribosomal loci (mitochondrial 16S and nuclear 18S, 28S) and two protein-coding genes (mitochondrial COX1 and nuclear Histone-3) of 74 specimens and added new sequences to NCBI GenBank (Tables 1, 2, S2). When we sequenced PCR amplicons, we defined "successful amplification" as those amplicons that yielded clear gel electrophoresis bands and clean chromatograms that aligned with our in-house database. Less than 10% of amplicons were not sequenced but nonetheless yielded clear bands, and these were computed as successful amplicons.

2.3. Protocol 1. Complete information of live-imaged specimens with preserved voucher exuviae

2.3.1 Culturing and morphological data. Plankton-collected y-larvae were sorted into petri dishes and reared further in mini-cultures of up to five morphologically distinct types or forms as outlined in Olesen et al., (2022). These were then followed alive daily in culture by video under a light microscope (LM). Live specimens were transferred between petri dishes and the light microscope in a glass Pasteur pipette. The final naupliar exuviae of the developing specimens were mounted in glycerin jelly on glass slides (GL; Grygier et al., 2019), or (although only attempted for a few specimens) on SEM-stubs for high-resolution morphological investigations (Grygier et al., 2019).

2.3.2 DNA extraction. Specimens were extracted using both the GeneReleaser and DNeasy kits. For GeneReleaser, we follow a slightly modified protocol from that outlined in Schizas et al., (1997). 1 µL of 10xPCR buffer and 9 µL ddH_2O were mixed in a 250 µL Eppenorph tube, to which a single y-larva was added. Protein denaturization was initiated by incubating the tubes at 94-95°C for 2 mins, and the specimens were subsequently transferred to ice. We then added 1 µL protein kinase K (QIAGEN, Chatsworth, CA, USA) and vortexed the samples thoroughly before incubating them at 55°C for 15 min and 70°C for 10 min. Subsequently, we added 10 µL of GeneReleaser (BioVentures, Inc, Murfreesboro TN, USA) and finally incubated the samples under the following thermal program: 65°C for 15 sec, 8°C for 15 sec, 65°C for 45 sec, 97°C for 90 sec, 8°C for 30 sec, 65°C for 90 sec, 97°C for 30 sec, 65°C for 30 sec, 80°C for 3 min, and lastly 4°C for 10 min to halt the reaction. The tubes were then centrifuged for 1 min, after which 15 µL supernatant and 10 µL AE-buffer (QIAGEN, Chatsworth, CA, USA) were transferred to fresh Eppendorph tubes. To locate and retrieve voucher "exuviae", the samples were spun down and inspected under a dissection microscope. Any exuviae were recovered and mounted as described below.

For the simplified DNeasy method, we transferred individual y-larval specimens from their vials into small, individual petri dishes. After gently vortexing the dish by hand, the specimens were often easy locatable at the center of the dish, after which they were transferred to a 250 µL Eppendorf tube in a 0.5-1.0 µL droplet of ethanol by using a 2.5 µL pipette fitted with a sterile filter tip. The vial(s), each containing a single y-larva specimen, were then incubated for 2 min at 36°C to evaporate the remaining ethanol. 40 µL AE buffer and 4uL Protease K (Qiagen, DNeasy kit) were subsequently added to each vial, which were then incubated for 1h at 56°C followed by enzyme inactivation at 72°C for 12 min. The vials were then spun down and placed in a rack. The "exuviae", now located at the bottom of their vial, were removed in a 1µL droplet by using a 2.5µL pipette fitted with a filter tip and either placed directly in pre-heated glycerin jelly on a glass slide or in a droplet in a petri dish with a small note written on the lid above the droplet indicating the specimen's ID code for later mounting. We usually performed PCR of "legacy" markers (12S, 16S, 18S, and 28S rDNA, COX1, and H3) before mounting the exuviae to avoid potential contamination prior to PCR. Using the pipette tip or an eyelash taped to a stick, y-naupliar exuviae (those of the last y-nauplius or an earlier stage) were rotated so the dorsal side faced upward, and y-cyprids were placed on their lateral side. We found intact exuviae in vials that had been stored at -18°C for up to 3 years, suggesting that continued freezing does not impact the quality of prospective vouchers.

We also tested the non-destructive protocol provided by Cornils (2015), which uses Protease K and ATL-buffer instead of AE buffer for DNA extraction. Their protocol consistently worked for a range of copepod species, but for y-larvae yielded faint bands, fragmented exuviae, and noisy chromatograms (i.e., low quality DNA template amplicons).

2.3.3 PCR. PCR was done as described in the Methods and Materials section. The vials were sequenced with their PCR-identifier, complete specimen-designation, and gene listed sequentially (e.g., 1-JA-2019-001-Hansenocaris-demodex_COX1, 2-JA-2019-100-Hansenocaris-demodex_COX1, etc).

2.4 Protocol 2: Assorted information of specimens with or without voucher exuviae

Culturing and morphological data. This method is clearly distinguished from Protocol 1 by retaining notably fewer data layers than Protocol 1. For example, while the WS-specimens are represented by all data layers, except for live images of the sequenced specimens, the AZ-material is exclusively represented by exuvial vouchers photographed in LM. Although admittedly artificial, we decided to lump all cases with varying degrees of morphological information, but no live imaging, into one category. Protocol 2 was applied to the three live-imaged specimens of *H. demodex* from Green Island (TA) in Olesen et al., (2022), the specimens of *H. itoi* from the White Sea (WS; for which all data layers except live-images of single specimens exist), and the Azores specimens (AZ; for which only y-naupliar exuviae are available).

2.4.1 DNA extraction. We tested three extraction procedures: that used exclusively for Protocol 1, the DNeasy kit (as described in detail above and in the Supplementary Material) following the manufacturer's specifications, and the GeneReleaser kit (as described in detail above).

2.5 Protocol 3: Anecdotal information on non-imaged, non-vouchered specimens.

2.5.1 Culturing and morphological data. This protocol is clearly distinguished from any other by the lack of any image data for the specimens. Information about single/bulk-sequenced specimens is, therefore, anecdotal at best. We did not apply this method, but it was used previously by others to generate the published y-larva sequences in NCBI GenBank of "Facetotecta spp." and *H. itoi* (Perez-Losada et al., 2002, 2009; Gallego et al., 2015). These authors used different DNA extraction kits but successfully amplified long and short ribosomal and protein-coding mitochondrial (16S and COX1, respectively) and nuclear (18S, 28S, and H3, respectively) markers.

2.6 Phylogeny estimation and barcoding of individual y-larvae

Chromatograms of new forward and reverse sequences were trimmed and assembled in Geneious Prime. Protein-coding genes were translated (invertebrate mitochondrial code for COX1 and universal code for H3) to proteins, aligned, and manually checked for stop codons. Using MAFFT (–auto –maxiterate 1000), nucleotides of each locus and specimen were then aligned with outgroup sequences from Ascothoracida and previously sequenced material of Facetotecta (Perez-Losada et al., 2009). Individual locus alignments were first inspected by eye, then subsequently curated by assessing ambiguities with the less stringent options in GBLOCKS version 0.91b (-t=p -b1=24 -b2=39 -b3=4 -b4=10 -b5=n -b6=y; Castresana 2000).

The nucleotide alignments were then concatenated in Geneious Prime (6584bp and 3415bp before/after GBLOCKS), partitioned (each codon position in protein-coding genes), and used to estimate the best-fit substitution model with ModelFinder (Kalyaanamoorthy et al., 2017) by allowing partitions to be merged if this increased model fit (-m MFP+MERGE). Maximum Likelihood (ML) phylogenetic trees were built with IQ-TREE multicore version 2.0.3 (Minh et al., 2020) and node support was assessed with 10,000 ultrafast bootstrap (UFBoot) replicates and 10,000 replicates of the SH-like approximate likelihood ratio test (SH-aLRT; –alrt 10,000). To alleviate issues inherent to the UFBoot approximation method we specified the "-bnni" option. We allowed each partition to have its own substitution rate (-spp) and performed a more thorough nearest-neighbour interchange search (-allnni). We also estimated a phylogeny with Bayesian inference in Mr. Bayes 3.2 (Ronquist et al., 2012) using the concatenated dataset. We estimated posterior probabilities through six independent runs with four Markov chains for 10,000,000 generations and with sampling every 1000 generations. We used the 'invgamma' model as the lset rate and computed the potential scale reduction factor with the 'sump' command. We estimated a consensus phylogenetic tree with the first 25% of the iterations discarded as burn-in.

The unrooted tree estimates were visualized in Geneious Prime, rooted in Ascothoracida, and further annotated in Corel Draw(c) with morphological information from live and exuvial specimens. These illustrations

were adjusted individually for brightness, contrast, and background color.

3 RESULTS

3.1 Loci selection, DNA extraction, and primer efficacy

We tested 28 primer pair combinations by performing 2056 individual PCR reactions and provided 74 new sequences of Facetotecta that have been deposited in NCBI GenBank with the accession numbers XXXX-YYYY (Table S1).

We also tested the efficacy and amplification success of the two DNA-extraction protocols (Fig. 2A-C; Tables 1, S2). We performed 815 PCR reactions with the GeneReleaser kit and 1241 reactions with the simplified DNeasy kit (Table S2). The minimum and maximum amplification rates varied between 0% (DNeasy method, 16S 1471F/1427R, COI_Palerointernal, both n=6; Table S2) and ˜98% (Dneasy method, 18Sface2F/2R, n=109; Table S2). Both protocols produced nearly identical overall amplification rates (˜68%; Fig 2A, C; Table S2). The GeneReleaser protocol yielded at least ˜42% and at most ˜90% amplification success across loci (Fig 2; Table S2). Extraction and PCR amplification with the simplified DNeasy method yielded a slightly higher minimum amplification success (˜47%) and a slightly lower maximum (˜88%) (Fig 2A; Table S2); because different primer combinations were tested with the two protocols, these results may be biased. Combining the data from the two extraction methods, we computed an overall amplification success rate of 68% across all primer combinations (2096 successfully amplified fragments; Tables 1, S2; Fig 2A). Average (mean) successful amplification rates were ˜60% for COX1 (296 out of 495 reactions; Fig. 2; Tables 1, S2), ˜87% for 18S (515 of 589 reactions; Fig. 2C; Tables 1, S2), ˜85% for 28S (361 of 426 reactions; Fig. 2C; Tables 1, S2), ˜52% for 12S (73 of 139 reactions; Fig. 2C; Tables 1, S2), and 56% for 16S (168 of 300 reactions; Fig. 2C; Tables 1, S2). The amplification rate difference between the two extraction methods was notably higher for 16S and H3 than any other locus (Fig 2B, C; Table 1, S2), with 12S, COX1 and 18S yielding more similar amplification success rates between the two methods (Fig 2B, C; Tables 1, S2). For all primer combinations, 18S and 28S expectedly had the highest amplification rates, but we did not observe this pattern for H3 (˜40% with GeneReleaser and ˜75% with the simplified DNeasy method; Tables 1, S2).

Based on the highest observed primer efficacy (i.e., amplification success), as well as the overlap with publicly available primers, the fragment lengths obtained, and the quality of the sequences (e.g., the presence of clear and distinct chromatogram peaks), we recommend sequencing the following five "legacy" loci in future investigations of facetotecta systematics. COX1: F1new and R1 with the internal "Leray"-segment mlCOIInF/jgHCO2198 (˜93% amplification success; ˜750bp; Leray et al., 2013; Fig. 3; Tables 2, S2). 18S: initially with Face1F/2R (˜85% success; ˜1800bp; Fig. 3; Tables 2, S2); subsequently, depending on the amplification rate of the latter, Face1F/1R (˜83% success; ˜300 bp; Fig. 2; Tables 2, S2) and Face 2F/2R (˜98% success; ˜300bp; Fig. 3; Tables 2, S2). 28S: 84F (˜74% success; ˜880bp; Fig. 3; Tables 2, S2) with the internal primer Face3F/3R (˜94% success; ˜350bp; Fig. 3; Tables 2, S2). 12S: 3F (˜83% success; ˜300bp; Fig. 3; Tables 2, S2). 16S: SF/SR (˜60% success; 700bp; Tsang et al., 2009; Fig. 3; Tables 2, S2) with 137F/548R (˜50% success; ˜400bp; Fig. 3; Tables 2, S2).

3.2 *Morphotype assignment and phylogeny estimation*

Based on LM images we recovered two named species and six undescribed morphotypes (*Hansenocaris demodex* from Green I., Taiwan and Sesoko I., Okinawa; *H. itoi* from the White Sea, Russia; Type A* from Green I., Sesoko I., and the Azores; Type C from Sesoko I., Type D* from WangHaiXiang harbor, Taiwan and Sesoko I.; Type I* from the Azores; and Type AE* from Taiwan), with morphotype designations following Dreyer et al., submitted). To this were added four specimens that could not be morphotyped prior to sequencing due to lack either of images or any voucher material (Facetotecta sp. 1 from Antarctica and Facetotecta sp. 5 from Sesoko Island; Fig. 1).

Our alignment contained 81 specimens (of which seven are non-imaged and voucherless "Facetotecta sp." sequences), 6730 bps, 1680 distinct patterns, 1443 parsimony-informative sites, 384 singleton sites and 4903 constant sites (Table 3). Following the best substitution-model fit inferred by ModelFinder, we merged the

partitions as follows: TNe+R2 (18S rDNA; BIC-score 14930.890, LnL: -24693.263), TIM2+F+R2 (H3, all codons; BIC-score 6798.637, LnL: -24693.133), TN+F+I+G4 (28SrDNA; BIC-score 12621.319; LnL N/A), TPM2+F+I (12SrDNA; 3435.374, LnL N/A) and GTR+F+G4 (16S rDNA, COX1, all codons; BIC-score11902.261, LnL: -24705.039) (Table 3). We recovered two clades (I and II), although the monophyly of the second one was not well-supported (Clade I: 98.6% Ultrafast Bootstrap (UFBoot), 93% SH-aLRT support; Clade II: 66.6% UFBoot, 81% SH-aLRT; Fig. 1).

In Clade I the planktotrophic morphotype A* is paraphyletic (97.5% UFBoot/96% SH-aLRT). The placement of three Type A* specimens from the Pacific (TA-2018-096, JA-2019-126) and the Azores (AZ-2019-5) within a highly supported, monophyletic subclade comprising pigmented and yolky lecithotrophs (95.8% UFBoot/71% SH-aLRT support). The relationship between Types C, D*, and *H. demodex* remains unresolved in our data set (Fig. 1; 69.2% UFBoot/32% SH-aLRT). In our dataset, all the Clade II specimens and morphotypes are planktotrophic, although the feeding type of Facetotecta sp. 5 and Antarctic species 1 and 2 is unknown. All multi-specimen morphotypes in the clade are monophyletic with high support, including an unnamed morphotype from the Azores ("Unnamed y-nauplius"; AZ-2019-1, 2, 4, 6, 7, 9; 90.3% UFBoot/99% SH-aLRT), Type AE* (100 UFBoot/100% SH-aLRT) and *H. itoi* (99.9% UFBoot/100% SH-aLRT).

4 DISCUSSION

The enigmatic facetotectans (commonly referred to as "y-larvae") are a severely understudied crustacean subclass, and the notable dearth of molecular resources and published morphological voucher data concerning them has impeded reconstruction of the group's evolutionary history of (Perez-Losada et al., 2002, 2009; Kolbasov et al., 2022). To overcome the challenges inherent in inferring systematic relationships based on unvouchered, unphotographed specimens, as exemplified by Perez-Losada et al., (2002, 2009) and Gallego et al., (2015), we developed a cheap, fast, and voucher-preserving protocol capitalizing on the widely-used DNeasy extraction kit. Our aim was to develop a rearing and extraction protocol that maximizes morphological and molecular information for single y-larval specimens spanning what is emerging as a complex and wide phylogenetic range (Olesen et al., 2022; Kolbasov et al., 2022; Dreyer et al., submitted; data herein). By relying on a suite of light microscopic images of live larvae (Olesen et al., 2022) and by developing primers targeting both nuclear and mitochondrial genomic markers, we demonstrate a way to infer phylogenetic relationships based on larval instars of a crustacean taxon with a complicated and partly unknown lifecycle. In the following we discuss three central topics related to the development of these new molecular resources: (1) The efficacy and usefulness of the new primers and the two new DNA extraction methods, (2) how three overall protocols and their varying yields of morphological resolution may affect systematic accounts and future taxonomic classification, and finally (3) interesting features of the preliminary phylogeny, for example the molecular and biogeographical extent of morphotypes, and how to use such information in building a taxonomy that, for now, must rely entirely on larval stages. We believe that our protocols should also work on other molting invertebrate larvae.

4.1. Efficacy of DNA extraction protocols and primers

We first tested two different DNA extraction methods using a series of novel and published nuclear and mitochondrial primer pairs. Although primer amplification success varied between 0% and ˜98% (Table S2), the average amplification success of our "legacy" primers (Fig 3; Tables 3, S1) was ˜80%. The lowest amplification rates overall were seen for legacy primers 16S (average ˜62%) and H3 (average ˜60%), which is likely caused by divergent priming sites or too low copy numbers for individual y-larvae, respectively. Importantly, maximum amplification rates were not restricted to nuclear loci as amplification with the COX1 "Folmer" and "Leray" primers yielded ˜93% success. In combination with nuclear ribosomal markers that demonstrated high amplification rates, this will likely positively influence phylogeny estimation and robust species delimitation in future systematic efforts.

DNA extraction relied on either the GeneReleaser kit, which had previously been used to retrieve high-quality DNA templates from minute zooplankton specimen (Schizas et al., 1997; Bottger-Schnack and Machida, 2011; Watanabe et al., 2016), including y-larvae (Olesen et al., 2022), or a simplified yet highly effective modifi-

cation of the QIAGEN DNeasy Blood and Tissue kit. There are important advantages and disadvantages related to both protocols. On the one hand, GeneReleaser yields higher amplification rates for certain loci. On the other hand, the GeneReleaser kit only yields ˜15-20 µL DNA extract and its thermal incubation program passes through high temperatures (>90°C). This may have a severe impact on successful retrieval voucher "exuviae" because (1) in initial trials with various crustacean larvae (data not shown), incubation at such temperatures almost always resulted in the disintegration of the "exuviae", and (2) the "exuviae" may easily become trapped in the GeneReleaser powder grains when the tubes are spun down after incubation. Less than 5% of the vials contained "exuviae" after transferring the GeneReleaser supernatant to new vials. Moreover, the GeneRelaser-kit is more expensive (US$125 for a kit on 20 October 2022), considering that it does not come with buffers and protease K included, which thus have to be acquired elsewhere. By contrast, the simplified DNeasy kit contains all relevant products for fast and efficient DNA extraction and produces bright gel-bands in the expected size ranges, clean chromatograms, and high rates of both overall and locus-specific amplification (Figs. 2, 3; Table S2). For these reasons, we recommend using our extraction method with the simplified DNeasy kit and with the legacy primers. We are currently designing a series of primers for nuclear protein-coding genes in order to expand this approach to higher-level phylogenetics.

4.2. Comparing protocols and paving the road for comprehensive single-specimen systematics

Besides providing the first multi-locus, vouchered phylogeny estimate for Facetotecta, our preliminary phylogeny of this group (Fig. 1) demonstrates the advantages and disadvantages of the three protocols that we outlined above. Protocol 1 outperformed all previous attempts to study the molecular phylogenetics and evolution of Facetotecta. First of all, it retained the information obtained from living specimens (y-nauplii, y-cyprid) throughout their larval development and also retained voucher exuviae of single specimens after their DNA extraction. Secondly (Fig. 1) it allowed for consistent amplification of long (>1500bp) and short (<400bp) DNA templates of protein-coding and ribosomal loci of both the mitochondrial and nuclear genomes (Table 1, 2, 3, S2). Protocol 1 thus ensured maximal gathering of morphological and molecular information from individual specimens and allowed the scoring both y-naupliar and y-cyprid characters. Given the remarkably high diversity at local scales such Sesoko Island (Glenner et al., 2008; Olesen et al., 2022; Dreyer et al., in press), this protocol will surely become an important asset to understanding Facetotecta diversity. It has already proven invaluable for species descriptions when dealing with sympatric distributions of multiple species (Olesen et al., 2022). The last-stage nauplius (LSN) is an unambiguously homologous and easily recognizable stage, of which there is never more than one for a given taxon (Olesen et al., 2022). The exuviae of the LSN (shed during metamorphosis to the y-cyprid), can serve as either a primary or complementary voucher specimen whether it is mounted on a glass slide or fixed and preserved in a liquid storage medium.

For example, the integrative description of *Hansenocaris demodex*was based on a series of nearly identical specimens that were first reared and live-imaged and subsequently amplified and sequenced individually with a short 18S fragment (Olesen et al., 2022). To this was added crucial details of the y-cyprid as observed by both LM and SEM, the latter admittedly not relying on sequenced material, but on specimens confidently assigned to the species due to similarities of their preceding nauplii (in particular the LSN) to those of sequenced individuals. In future work, it is not unlikely that several data layers, including live images, LM images of voucher exuviae (of y-nauplii), SEM images (of y-cyprids and some LSN voucher exuviae, either naturally molted or remaining after DNA extraction), and DNA sequence data, will be harvested from a single specimen (Grygier et al., 2019). We recommend mounting naupliar exuviae in glycerin jelly as it solidifies at room temperature, thus easing maintenance in museum collections and tropical research facilities. Some LSN exuvial specimens may also be imaged by confocal laser-scanning microscopy (CLSM), although we have not attempted this yet for DNA-extracted specimens (for information on mounting exuviae for SEM, see Grygier et al., 2019; Olesen et al., 2022; Kolbasov et al., 2022).

Protocol 2 lumps together a series of culturing and sequencing strategies, all of which retain a less informative and less complete assortment of visual information than Protocol 1. It is typically applied to y-larvae with planktotrophic nauplii, such as y-naupliar Types A* and AE* and the planktotrophic y-nauplii from the

Azores and the White Sea, as these cannot yet be reared in the lab. The morphological information is sufficient for screening morphotype affinities established by molecular data but, in our view, does not provide a comprehensive basis for species descriptions. Namely, different instars in the life cycle (e.g., y-nauplii and -cyprids) cannot easily be linked together, especially at localities Sesoko Island with many sympatric morphotypes. We also demonstrate that protocol 2 works on other Thecostraca larvae, e.g., an ascothoracid larva identifed as *Baccalaureus* sp. nests as a sister species to *B. maldivensis* (Fig. 1). Thus, we anticipate that procols 1 and 2 should work on other marine invertebrate larvae that molts and leaves exuviae.

Protocol 3, which involves the use of non-cultured y-larvae and has been used for all specimens with sequences currently deposited in GenBank, except for those of Olesen et al., (2022). It does not produce any images, and thus no vouchered data, and is therefore impractical for all levels of y-larva phylogenetics. At best, it retains anecdotal information about the specimens obtained prior to sequencing. Despite these limitations, the present work has allowed us to determining the identity of a least some of the y-larval material ("Facetotecta sp. 1-6") used by Pérez-Losada et al., (2009), as some matches to the vouchered material sequenced herein from the same locality (Sesoko Island) are apparent. For example, their "Facetotecta spp. 1, 2, and 6" cluster within our Type A* and "Facetotecta sp. 4" is molecularly identical to Type D*. On the other hand, "Facetotecta sp. 5" and the Gallego et al's (2012, 2015) two sequences from Antarctica are still 'ghost' sequences and remains unlinked to known, imaged Facetotecta. Although these specimens are unvouchered and lack any kind of published morphological information, they have played a critical role at higher taxonomic levels, helping to place Facetotecta systematically within Thecostraca and Pancrustacea (Pérez-Losada et al., 2002, 2009; Petrunina et al., 2014).

4.3. Implications for classification, taxonomy and phylogeography

Our sequence data matrix recovered the same two clades found earlier (Pérez-Losada et al., 2009; Olesen et al., 2022), but we provide enhanced phylogenetic resolution by sequencing more markers for multiple new MOTUs, some of which may be putative species awaiting taxonomic description (Fig. 1). Importantly, most morphotypes are supported by molecular and morphological data in concert, testifying to the usefulness of combining these two data layers for phylogeny estimation and evaluation. Given the molecular diversity in the phylogeny and the morphological disparity of the larval forms, these two clades may represent high-level taxonomic units (e.g., families), but more study is required.

It is an inescapable fact that Facetotecta is in critical need of a novel higher-level classificatory scheme, as all species are currently placed in the single genus *Hansenocaris*. These new species are both morphologically and phylogenetically so diverse that it is highly unsatisfactory to retain them within the same genus. We expect that single-specimen culturing, live-imaging, amplification, and sequencing will be the key step towards a new classification of Facetotecta (Olesen et al., 2022). For planktotrophic specimens, which we are still unable to culture *in vitro* , this poses a significant challenge that may postpone detailed taxonomic descriptions. It is exceedingly difficult to morphologically separate Types A* and AE*, for example, into what are likely multiple "species" (or MOTUs) that together occupy a geographical range that extends from Japan and Taiwan to the Azores. It is highly unlikely that each Type constitutes one, big breeding population. Wide biogeographical ranges of morphotypes could occur in some planktotrophic y-nauplii, however, because of their long developmental times and putative long-distance dispersal capabilities. For example, completion of the larval development of the Arctic facetotectan *Hansenocaris itoi* requires three months (April through June: Kolbasov et al., 2021) and this species may have a wide distribution. Hansen (1899) identified y-nauplii that are virtually identical to *H. itoi* in the Baltic Sea (Kiel Bay), a place separated from the type locality in the White Sea by thousands of kilometers of open ocean. The lack of sequence data for the Baltic population, if it still exists, illustrates the importance of Protocol 1, in which molecular sequence data are linked to live image data.

In conclusion, the easy-to-follow and inexpensive Protocol 1 described herein allows the extraction of maximal molecular and morphological information from single y-larval specimens, although we expect that this method should work for any invertebrate taxon with free-living larvae. It should be used in future efforts to, for example, model life-history evolution in the Facetotecta or other groups. With a growing inventory of sequence

data from cultured, imaged, and vouchered y-larva specimens/species from geographically diverse places, the evolution of Facetotecta, as well as a comprehensive new taxonomic scheme for them, may finally be within reach.

ACKNOWLEDGEMENTS

The authors extend their gratitude to Dr. Danny Eibye-Jacobsen (University of Copenhagen, Denmark) for assisting collections in Taiwan and Japan, sometimes under trying weather conditions. Members of the Coastal Ecology lab at Academia Sinica are also thanked for assistance with sampling in Taiwan. This work was supported by a double degree graduate grant from the Natural History Museum of Denmark and the Taiwan International Graduate Program to ND and a Villum Experiment grant to JO.

REFERENCES

1. Bernot, J. P., Avdeyev, P., Zamyatin, A., Dreyer, N., Alexeev, N., Pérez-Losada, M., & Crandall, K. A. (2022). Chromosome-level genome assembly, annotation, and phylogenomics of the gooseneck barnacle*Pollicipes pollicipes* . *GigaScience* , *11*
2. Böttger-Schnack, R., & Machida, R. J. (2011). Comparison of morphological and molecular traits for species identification and taxonomic grouping of oncaeid copepods. *Hydrobiologia* , *666* (1), 111-125
3. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* , *17* (4), 540-552.
4. Chan, B. K. K., Dreyer, N., Gale, A. S., Glenner, H., Ewers-Saucedo, C., Pérez-Losada, M., Crandall, K. A., & Høeg, J. T. (2021). The evolutionary diversity of barnacles, with an updated classification of fossil and living forms. *Zoological Journal of the Linnean Society* , *193* (3), 789-846
5. Cornils, A. (2015). Non-destructive DNA extraction for small pelagic copepods to perform integrative taxonomy. *Journal of Plankton Research* , *37* (1), 6-10.
6. Dreyer, N., Tsai, P. C., Olesen, J., Kolbasov, G. A., Høeg, J. T., & Chan, B. K. K. (2022). Independent and adaptive evolution of phenotypic novelties driven by coral symbiosis in barnacle larvae.*Evolution* , *76* (1), 139-157
7. Dreyer, N., Zardus, J. D., Høeg, J. T., Olesen, J., Yu, M. C., & Chan, B. K. (2020). How whale and dolphin barnacles attach to their hosts and the paradox of remarkably versatile attachment structures in cypris larvae. *Organisms Diversity & Evolution* , *20* (2), 233-249
8. Gallego, R., Heimeier, D., Lavery, S., & Sewell, M. A. (2015). The meroplankton communities from the coastal Ross Sea: A latitudinal study. *Hydrobiologia* , *761* (1), 195–209. https://doi.org/10.1007/s10750-015-2487-7
9. Gallego, R., Heimeier, D., Lavery, S., & Sewell, M. A. (2015). The meroplankton communities from the coastal Ross Sea: a latitudinal study. *Hydrobiologia* , *761* (1), 195-209.
10. Glenner, H., Høeg, J. T., Grygier, M. J., & Fujita, Y. (2008). Induced metamorphosis in crustacean y-larvae: towards a solution to a 100-year-old riddle. *BMC biology* , *6* (1), 1-6.
11. Grygier, M. J. (1987). New records, external and internal anatomy, and systematic position of Hansen's y-larvae (Crustacea: Maxillopoda: Facetotecta). *Sarsia* , *72* (3-4), 261-278.
12. Grygier, M. J., Høeg, J. T., Dreyer, N., & Olesen, J. (2019). A new internal structure of nauplius larvae: a "ghostly" support sling for cypris y left within the exuviae of nauplius y after metamorphosis (Crustacea: Thecostraca: Facetotecta). *Journal of Morphology* ,*280* (8), 1222-1231
13. Hansen, H. J., 1899. Die Cladoceren und Cirripedien der Plankton-Expedition. *Ergeb. Atlantic Ocean Plankton-Exped.Humboldt-Stiftung 2* : 1–58.
14. Hartop, E., Srivathsan, A., Ronquist, F., & Meier, R. (2022). Towards large-scale integrative taxonomy (LIT): Resolving the data conundrum for dark taxa. *Systematic Biology* , syac033. https://doi.org/10.1093/sysbio/syac033
15. Itô, T. (1985). Contributions to the knowledge of cypris y (Crustacea: Maxillopoda) with reference to a new genus and three new species from Japan. In: *Special Publication of the Mukaishima Marine Biological Station 1985* , 113–122
16. Itô, T. (1986a). Three types of "nauplius y" (Maxillopoda: Facetotecta) from the North Pacific. *Publi-*

cations of the Seto Marine Biological Laboratory , *31* (1/2), 63–73.
17. Itô, T. (1987b). Three forms of nauplius y type VIII larvae (Crustacea: Facetotecta) from the North Pacific. *Publications of the Seto Marine Biological Laboratory* , *32* ,141-150.
18. Itô, T. (1989). A new species of *Hansenocaris* (Crustacea: Facetotecta) from Tanabe Bay, Japan. *Publications of the Seto Marine Biological Laboratory* , *34* , 55-72.
19. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* , *14* (6), 587-589
20. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* , *30* (4), 772-780
21. Kolbasov , G. A., Petrunina, A. S., Dreyer, N., Chan, B. K. K., & Høeg, J. T. (2022). A synthesis of the external morphology and cypridiform larvae of Facetotecta (crustacea [sic]: Thecostraca) and the limits of *Hansenocaris* . *Ecology & Evolution* , 12, e9488, https://doi.org/10.1002/ece3.9488
22. Kolbasov, G. A., & Høeg, J. T. (2003). Facetotectan larvae from the White Sea with the description of a new species (Crustacea: Thecostraca). *Sarsia* , *88* (1), 1–15. https://doi.org/10.1080/00364820308471
23. Kolbasov, G. A., Petrunina, A. S., Dreyer, N., Prudkovsky, A. A., Olesen, J., Chan, B. K. K., & Høeg, J. T. (2021). Naupliar development of Facetotecta (Crustacea: Thecostraca) and the nature of the first nauplius instar in the Crustacea - pro et contra the Hexanauplia concept. *Organisms Diversity & Evolution* ,*21* (1), 209-230
24. Lozano-Fernández, J., Giacomelli, M., Fleming, J. F., Chen, A., Vinther, J., Thomsen, P. F., ... & Olesen, J. (2019). Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remipede sampling. *Genome Biology and Evolution* ,*11* (8), 2055-2070
25. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era.*Molecular biology and evolution* , *37* (5), 1530-1534
26. Oakley, T. H., Wolfe, J. M., Lindgren, A. R., & Zaharoff, A. K. (2013). Phylotranscriptomics to bring the understudied into the fold: monophyletic Ostracoda, fossil placement, and pancrustacean phylogeny.*Molecular biology and evolution* , *30* (1), 215-233.
27. Olesen, J., & Grygier, M. J. (2022). Two new species of lecithotrophic nauplius y with remarkable labra from Okinawa, Japan, and a family-group name for y-larvae (Crustacea: Thecostraca: Facetotecta: Hansenocarididae fam. nov.). *Species Diversity* , *27* (2), 301-317
28. Olesen, J., Dreyer, N., Palero, F., Eibye-Jacobsen, D., Fujita, Y., Chan, B. K. K., & Grygier, M. J. (2022). Integrative taxonomy of crustacean y-larvae (Thecostraca: Facetotecta) using laboratory-rearing and molecular analyses of single specimens, with the description of a new vermiform species. *Zoological Journal of the Linnean Society* .
29. Pérez-Losada, M., Høeg, J. T., & Crandall, K. A. (2009). Remarkable convergent evolution in specialized parasitic Thecostraca (Crustacea).*BMC biology* , *7* (1), 1-12
30. Pérez-Losada, M., Høeg, J. T., Kolbasov, G. A., & Crandall, K. A. (2002). Reanalysis of the relationships among the Cirripedia and the Ascothoracida and the phylogenetic position of the Facetotecta (Maxillopoda: Thecostraca) using 18S rDNA sequences. *Journal of Crustacean Biology* , *22* (3), 661-669.
31. Petrunina, A. S., Neretina, T. V., Mugue, N. S., & Kolbasov, G. A. (2014). Tantulocarida versus Thecostraca: inside or outside? First attempts to resolve phylogenetic position of Tantulocarida using gene sequences. *Journal of Zoological Systematics and Evolutionary Research* , *52* (2), 100-108
32. Petrunina, A. S., Neretina, T. V., Mugue, N. S., & Kolbasov, G. A. (2014). Tantulocarida versus Thecostraca: inside or outside? First attempts to resolve phylogenetic position of Tantulocarida using gene sequences. *Journal of Zoological Systematics and Evolutionary Research* , *52* (2), 100-108
33. Regier, J. C., Shultz, J. W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., ... & Cunningham, C. W. (2010). Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences.*Nature* , *463* (7284), 1079-1083.
34. Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large

model space. *Systematic biology* , *61* (3), 539-542
35. Schizas, N. V., Street, G. T., Coull, B. C., Chandler, G. T., & Quattro, J. M. (1997). An efficient DNA extraction method for small metazoans. *Molecular Marine Biology and Biotechnology* , *6* (4), 381
36. Schram, T.A. (1972). Further records of nauplius y type IV Hansen from Scandinavian waters. *Sarsia* , *50* , 1-24.
37. Schwentner, M., Combosch, D. J., Nelson, J. P., & Giribet, G. (2017). A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Current Biology* , *27* (12), 1818-1824
38. Tsang, L. M., Chan, B. K. K., Shih, F. L., Chu, K. H., & Chaoulun, A. (2009). Host-associated speciation in the coral barnacle *Wanella milleporae* (Cirripedia: Pyrgomatidae) inhabiting the *Millepora*coral. *Molecular Ecology* , *18* (7), 1463-1475
39. von Reumont, B. M., Jenner, R. A., Wills, M. A., Dell'Ampio, E., Pass, G., Ebersberger, I., ... & Misof, B. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Molecular biology and evolution* , *29* (3), 1031-1045.
40. Watanabe, H. K., Senokuchi, R., Shimanaga, M., & Yamamoto, H. (2016). Comparison of the efficiency of three methods of DNA extraction for deep-sea benthic copepods. *JAMSTEC Report of Research and Development* , *23* , 52-59

FIGURE LEGENDS

Fig. 1. **Multilocus phylogeny estimation for Facetotecta with single-specimen resolution. A** Maximum-likelihood tree from an alignment spanning 6730bp and six loci with Ascothoracida chosen as outgroup, showing relationships of Facetotecta and its division into at least two major clades. **B** Putative life cycle of Facetotecta consisting of seven y-nauplius stages, one y-cyprid stage, one or more ypsigons, and as yet unknown adults. **C** Examples of morphological information extracted from Protocols 1, 2, and 3, represented by branch colors (red, grey, and yellow, respectively). Protocol 2 lumps different subprotocols that lack live image data, here illustrated by three examples.

Fig. 2. ***Comparison of primer amplification efficacy with two filter-free DNA extraction protocols for single-larva systematics.*** **A** Amplification success of all primers using GeneReleaser and Simplified DNeasy extraction kits. **B** Amplification success of all primers partitioned to locus using GeneReleaser and the Simplified DNeasy extraction kits. The large difference for 16S and H3 may be caused by the different primers used for each extraction protocol (Table 1S). **C** density plot of amplification success of all primers for both extraction kits combined, showing uniformly higher amplification rates for 12S, 18S, and COX1 and uniformly lower amplification rates for 16S and H3.

Fig. 3. ***Amplification success and efficacy of "legacy" primers .*** **A** Combined amplification success using both GeneReleaser and Simplified DNeasy extraction kits. **B** Summary of **A** but with amplification success partitioned to locus, hence the comparatively larger difference for 28S and 18S primers.

AUHTOR CONTRIBUTIONS

Niklas Dreyer (ND) designed and conceived the study. ND participated in sampling, co-designed and tested primers, extraction methods and protocols. ND performed all sequence, phylogenetic and statistical analyses, co-funded the project, designed graphical outputs, took specimen photos and wrote the first draft. Ferran Palero supervised the project, designed primers, assisted statistical analyses, participated in sampling and edited manuscript drafts. Mark J. Grygier (MJG) participated in sampling, co-developed protocols, originally developed the exuvium-voucher strategy and edited manuscript drafts. Alexandra S. Savchenko provided the valuable Azores-material and edited the manuscript drafts. Gregory A. Kolbasov provided the valuable White Sea-material and edited the manuscript drafts. Ryuji J. Machida co-supervised the project, designed DNA extraction methods, assisted primer testing, and edited manuscript drafts. Benny K. K. Chan supervised the project, led sampling in Taiwan and co-funded the project. Jorgen Olesen supervised the project, co-conceived the study, led sampling in Japan, participated in sampling in Taiwan, co-designed protocols, co-funded the project, assisted phylogenetic analyses, designed graphical outputs, took specimen photos,

and edited manuscript drafts.

NIKLAS DREYER[1,2,3,4], FERRAN P. PALERO[5+*], MARK J. GRYGIER[6,7], ALEXANDRA S. SAVCHENKO[8], GREGORY A. KOLBASOV[9], RYUJI J. MACHIDA[1], BENNY K. K. CHAN[1+*] & JORGEN OLESEN[1+*]

TABLES

Table 1. Amplification success of all primers with single larval Facetotecta specimens using Protocols 1 and 2.

Locus	Total tested	Total amplified	% Success
COX1	495	296	59.79
18S	589	515	87.43
28S	426	361	84.74
12S	139	73	52.51
16S	300	168	56
H3	107	70	65.42
Total	**2056**	**1483**	**67.65**

Table 2. Amplification success of preferred "legacy" loci for Facetotecta phylogenetics.

Locus	Primers	# Tested	# Amplified	% Success
COX1	F1new/R1 + mlCOIInF/jgHCO2198	126	117	92.85714286
18S	Face1F/Face1R	132	110	83.33333333
18S	Face2F/Face2R	109	107	98.16513761
18S	Face1F/Face2R	76	65	85.52631579
28S	Face3F/Face3R	154	145	94.15584416
28S	28S_83F	111	82	73.87387387
12S	3F	12	10	83.33333333
16S	16S 137F/548R	181	115	63.5359116
16S	SF/SR	68	41	60.29411765
H3	H3af/H3ar	96	67	69.79166667

Table 3 . Alignment and nucleotide substitution moedls chosen for phylogeny estimation after GBLOCKS.

Locus	# Sequences	# Sites	# Informative sites	# Invariable sites	Model
12S	21	404	122	250	TVM+F+I+G4
16S	52	321	160	138	TVM+F+I+G4
18Sface1	46	1383	299	1037	Tne+R2
18Sface2	70	240	36	196	K2P+I+G4
28S	72	648	76	552	K2P+I+G4
COX1	33	419	173	214	TVM+F+I+G4
H3	25	260	73	169	TIM2+F+I+I+R2
Total	**294**	**3415**	**866**	**2387**	**N/A**

ELECTRONIC SUPPLEMENTARY TABLES (can be obtained directly through FigShare: https://figshare.com/s/9a08fa4635dd4f5c4b14)

Table S1. Overview of specimens, collection data and GenBank accession numbers of specimens used in the present study.

Table S2 . Primer amplification success for the two DNA extraction methods.

DATA ACCESSIBILITY STATEMENT

All new sequences have been deposited in NCBI GenBank under accession numbers XXXX-YYYY.

All figures, tables and supplementary material can be found at: https://figshare.com/s/9a08fa4635dd4f5c4b14

BENEFIT-SHARING STATEMENT

Benefits generated: This project is part of a larger research collaboration seeking to understand the biodiversity and evolution of the enigmatic crustacean group Facetotecta. The collaboration was developed with scientists from diverse countries, all of which provided unique samples and are co-authors of the present paper. Our research targets providing molecular resources for understanding y-larva evolution, but the protocols should work for any invertebrate taxon with molting larvae. Our project will therefore benefit the wider community in evolutionary ecology. Our research group is committed to broad, diverse, and inclusive scientific collaborations, evidenced by our team members and samplings originating from diverse countries.