



HAL
open science

An improved taxonomic sampling is a necessary but not sufficient condition for resolving inter-families relationships in Caridean decapods

Laetitia Aznar-Cormano, Julien Brisset, Tin-Yam Chan, Laure Corbari, Nicolas Puillandre, José Utge, Magali Zbinden, Dario Zuccon, Sarah Samadi

► **To cite this version:**

Laetitia Aznar-Cormano, Julien Brisset, Tin-Yam Chan, Laure Corbari, Nicolas Puillandre, et al.. An improved taxonomic sampling is a necessary but not sufficient condition for resolving inter-families relationships in Caridean decapods. *Genetica*, 2015, Next-generation sequencing for Biodiversity, Taxonomy, Evolution and Ecology, 143 (2), pp.195-205. 10.1007/s10709-014-9807-0 . hal-03922675

HAL Id: hal-03922675

<https://hal.science/hal-03922675>

Submitted on 10 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **An improved taxonomic sampling is a necessary but not sufficient**
2 **condition for resolving inter-families relationships in Caridean**
3 **decapods.**

4 **Aznar-Cormano L¹, Brisset J², Chan T-Y³, Corbari L¹, Puillandre N¹, Utge J⁴, Zbinden M⁵,**
5 **Zuccon D^{1,4}, Samadi S¹.**

6

7 1 ISYEB - UMR 7205 – CNRS, MNHN, UPMC, EPHE, Muséum national d’Histoire naturelle, Sorbonne Universités,
8 CP26, 57 rue Cuvier, F-75231 Paris cedex 05, France.

9 2 Direction des collections, Muséum National d’Histoire Naturelle, 57 rue Cuvier, 75005 Paris, France.

10 3 Institute of Marine Biology and Centre of Excellence for the Oceans, National Taiwan Ocean University, Keelung,
11 Taiwan

12 4 Service de Systématique Moléculaire, UMS2700 MNHN-CNRS, Département Systématique et Evolution, Muséum
13 National d’Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France.

14 5 Biologie des Organismes Aquatiques et Ecosystèmes, UMR7208, MNHN-CNRS-IRD-UPMC, Muséum National
15 d’Histoire Naturelle, France.

16

17 **Corresponding author:**

18 Pr Sarah Samadi

19 Muséum National d’Histoire Naturelle, Département Systématique et Evolution

20 Institut de Systématique, Evolution, Biodiversité, UMR 7205 CNRS MNHN UPMC EPHE

21 57 rue Cuvier CP 26

22 75231 Paris Cedex 05, France

23 Phone: 33 1 40 79 37 59

24 E-mail: sarah.samadi@mnhn.fr

25

26

27 **Abstract**

28 During the past decade, a large number of multi-gene analyses aimed at resolving the phylogenetic
29 relationships within Decapoda. However relationships among families, and even among sub-families,
30 remain poorly defined. Most analyses used an incomplete and opportunistic sampling of species, but
31 also an incomplete and opportunistic gene selection among those available for Decapoda. Here we
32 test in the Caridea if improving the taxonomic coverage following the hierarchical scheme of the
33 classification, as it is currently accepted, provides a better phylogenetic resolution for the inter-
34 families relationships. The rich collections of the Muséum National d'Histoire Naturelle de Paris are
35 used for sampling as far as possible at least two species of two different genera for each family or
36 subfamily. All potential markers are tested over this sampling. For some coding genes the
37 amplification success varies greatly among taxa and the phylogenetic signal is highly saturated. This
38 result probably explains the taxon-heterogeneity among previously published studies. The analysis is
39 thus restricted to the genes homogeneously amplified over the whole sampling. Thanks to the
40 taxonomic sampling scheme the monophyly of most families is confirmed. However the genes
41 commonly used in Decapoda appear non-adapted for clarifying inter-families relationships, which
42 remain poorly resolved. Genome-wide analyses, like transcriptome-based exon capture facilitated by
43 the new generation sequencing methods might provide a sounder approach to resolve deep and rapid
44 radiations like the Caridea.

45

46

47 **Keywords:** Caridea, phylogeny, museum specimens.

48

49

50

51 **Introduction**

52

53 During the last decade, many efforts have been employed to resolve, using multi-gene
54 approaches, the deeper nodes in the phylogeny of Decapoda (e.g. Martin et al. 2010; Ahyong et al.
55 2011; Tsang et al. 2014). However, the comparison among molecular studies gives a picture as
56 contradictory as are the morphological studies. As pointed out for example by Shen et al. (2013) or
57 Tsang et al. (2014) unbalanced taxon sampling is a major source of incongruence among multi-gene
58 analyses. Moreover, most of the available datasets differ both in taxon sampling and in gene sampling.
59 As a consequence it is very difficult to combine and compare the results among studies. The objective
60 of this paper is to test if the incongruence comes primarily from incomplete taxon and gene sampling
61 or from a lack of phylogenetic signal in the genetic markers used in published datasets. For testing
62 this hypothesis, we take advantage of the large Decapoda collection of the Muséum National
63 d'Histoire Naturelle of Paris (MNHN) which benefit from forty years of marine expeditions as part
64 of the ongoing Tropical Deep-Sea Benthos program, led by the MNHN and the Institut de Recherche
65 pour le Développement (IRD). This program has generated an important flow of new material from
66 the Pacific and Indian Oceans that have been studied by an active network of taxonomists (Richer de
67 Forges et al. 2013). This collection has also been shown to be an adequate source of specimens for
68 molecular analysis and was used in large DNA-barcoding projects (Puillandre et al. 2012; Zuccon et
69 al. 2012).

70 To perform such an analysis we selected the Caridea that, with more than 3 200 extant species,
71 is after the Brachyura, the second most speciose infraorder of Decapoda (De Grave et al. 2009).
72 Caridean shrimps are present worldwide, from tropical to polar regions, both in marine and freshwater
73 environments. If the definition of Caridea is not questioned (e.g. De Grave et al 2009; Bracken et al.
74 2010; Shen et al. 2013), the classification at the superfamily and family levels is far from being
75 resolved. Several studies use a molecular phylogenetic approach to discuss the classification.
76 However, the results remain unsatisfactory because none of the studies combined a wide taxonomic

77 coverage of Caridean diversity and a phylogenetic signal in the sampled genetic markers that provides
78 support for the family-level relationships. For example, the study of Bracken et al. (2009) includes
79 31 of the 38 currently accepted families, with 17 families represented by at least 2 species, but only
80 2 non-coding genes (mitochondrial 16S and nuclear 18S). Conversely, Li et al. (2011) use 5 nuclear
81 genes (18S and 4 coding genes: NaK, PEPCK, H3 and enolase) but only 20 Caridean families with
82 only 10 of them represented by at least 2 species. The aim of the present study is to provide a multi-
83 gene analysis at the family-level based on an improved taxonomic sampling of the Caridea. To
84 enhance the genetic sampling we also tested all the genetic markers used in the recent Decapoda
85 literature and explored the public genetic databases to identify potential new markers. The main
86 sources of markers were the recent articles on Caridea (Bracken et al. 2009, Chan et al. 2010, Li et
87 al. 2011, Kou et al. 2013) and Decapoda phylogeny (Toon et al. 2009). The taxonomic sampling
88 follows the hierarchical scheme of the latest revisions of the classification of Decapoda (*i.e.* De Grave
89 et al. 2009, 2014; De Grave and Fransen 2011; Short et al. 2013). The sampling is designed to test
90 the monophyly at the family level and to infer inter-familial relationships using within each family as
91 far as possible at least two species from two distinct genera.

92

93

94 **Materials and methods**

95

96 *The state-of-the-art of Caridean classification and taxon selection*

97

98 Several classifications have been proposed for the Caridea. The classification used in WoRMS
99 (World Register of Marine Species: www.marinespecies.org) is a synthesis of current proposals. This
100 classification is primarily based on the classification proposed by De Grave et al. (2009) using a
101 comparative morphology approach. The *Carideorum Catalogus* (De Grave and Fransen 2011) and
102 recent molecular phylogenies (Page et al. 2008 b; Bracken et al. 2010; Chan et al. 2010; De Grave et

103 al. 2010, 2014; Short et al. 2013) provided five main modifications to the classification of De Grave
104 et al. (2009). First, the Procarididae were excluded from the Caridea and included in a distinct
105 infraorder, the Procarididea (Bracken et al. 2010). Second, the Oplophoridae were shown to be
106 polyphyletic (Chan et al. 2010) and thus separated into two distinct families, the Acanthephyridae
107 and the Oplophoridae *s.s.*. These two families were included in the same superfamily, the
108 Oplophoroidea. Third, two genera, *Eugonatonotus* and *Galatheacaris*, were synonymized and placed
109 in the Eugonatonotidae (De Grave et al. 2010). Fourth, the family Kakaducarididae was synonymized
110 with Palaemonidae (Short et al. 2013). The fifth is a just published work (De Grave et al. 2014)
111 separating Hippolytidae into five families, namely Merguiddae, Bythocarididae, Thoridae,
112 Hippolytidae *s.s.* and Lysmatidae. In this revised classification, 14 superfamilies and 38 families are
113 considered as valid (table 1). Because the goal of the study was to test the monophyly of these families
114 and the relationships among them, we tried to include in the sampling at least two genera, each
115 represented by at least two species, for as many families as possible.

116 The MNHN collections were screened to select specimens based on their taxonomic
117 identification but also based on the availability of field data (i.e. precise location, depth, habitat, etc.).
118 To increase the success of the DNA sequencing, we also considered additional criteria such as the
119 sampling dates and the apparent conservation state of the specimen. Specimens that might have been
120 previously fixed in formaldehyde were tentatively excluded although this information is generally
121 lacking. For the taxa not present in the MNHN collections or for which the sequencing success rate
122 was too low, we supplemented our dataset with sequences from GenBank.

123 Whenever available, for each taxon we selected several specimens sampled in different
124 collecting events (i.e. cruises or expeditions). This strategy allows us to check the reliability of the
125 sequences obtained from the same taxon and easily detect potential contaminations, and also to
126 identify within each taxon the sample providing the DNA extract of better quality (see also below).

127 Three species from other infraorders of Decapoda were selected in GenBank and used as
128 outgroup: *Litopenaeus vannamei* (Dendrobrachiata), *Uroptychus parvulus* (Anomura, Pleocyemata)

129 and *Cancer pagurus* (Brachyura, Pleocyemata). Because the Procarididea were considered Caridea
130 by some authors, two species of Procarididae from GenBank were also added to the matrix to
131 corroborate the hypothesis that Procarididea is a distinct infraorder from the Caridea.

132

133 *Genetic markers selection*

134 The most documented genes in the public databases for the Caridea are the fragment of the COI
135 mitochondrial gene used in DNA-barcoding projects and fragments of the 18S RNA and 28S RNA
136 nuclear genes. The sequences available in GenBank (and BOLD for the COI) cover most of the
137 Caridean families. We thus use this set of genetic markers to detect possible contamination of the
138 DNA extracts and/or sequences resulting from potentially poor DNA quality in part of the museum
139 specimens. These markers are used to select among the museum specimens those that will provide
140 the better quality DNA extracts.

141 In a second step, six additional gene fragments (16S, EPRS, H3, NaK, PEPCCK and TM9SF4),
142 used in various phylogenetic studies in groups of Caridea and/or Decapoda (e.g. Bracken et al. 2009,
143 Toon et al. 2009, Li et al. 2011), were tentatively amplified and sequenced. Because these genes are
144 more difficult to amplify and sequence, contaminations at the PCR step were expected. However,
145 their detection should be facilitated by the fact that potential contamination at the previous step (DNA
146 extraction) would have been ruled out. Since nuclear coding genes are supposed to increase the
147 resolution of the phylogeny at this scale (Tsang et al. 2008), GenBank was explored for additional
148 coding gene markers. However, only few genomic data are available for Caridea and only one marker,
149 previously not used in Caridean phylogeny, was identified as a new candidate gene: a fragment of a
150 β -actine gene, for which six sequences for four distinct species (*Exopalaemon carinicauda*,
151 *Macrobrachium amazonicum*, *Macrobrachium rosenbergii* and *Palaemonetes pugio*) were available
152 in GenBank (ref : JX948081, JQ045354, AF221096, AY651918, AY626840, AY935989).

153

154 *DNA amplification and sequencing*

155 To extract DNA, a pleopod was used because these structures are generally uninformative in Caridean
156 taxonomy (Chace 1992, Bracken et al. 2009). Pleopods III or IV were preferentially used since for
157 many Carideans the second and sometime the first pleopods may provide diagnostic characters for
158 males. Total genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen) or the
159 NucleoSpin 96 Tissue kit (Macherey-Nagel) with the automated pipetting system epMotion 5075,
160 according to the manufacturer instructions. The PCR reactions were performed in 20 μ L reaction
161 volume, containing a final concentration of 1X reaction buffer, 3.4 mM MgCl₂, 0.26mM dNTP, 0.3
162 mM of primers, 5% DMSO and 1.2 units of Qiagen Taq polymerase, plus 1.5 μ L of DNA extract.
163 The amplification thermal profiles consisted of an initial denaturation for 5 min at 94°C, followed by
164 cycles of denaturation at 94°C, annealing, extension at 72°C and a final extension at 72°C for 5 min.
165 The number of cycles, duration for denaturation, annealing and extension steps, and annealing
166 temperature for each fragment are provided in table 3 (Supplementary material). All markers were
167 amplified in a single fragment, except for the 18S RNA gene which was amplified in three
168 overlapping fragments. The PCR products were visualised on a 2% agarose gel stained with ethidium
169 bromide and the positive PCR products were purified and sequenced in both directions using the
170 Sanger method by an external sequencing facility (Génoscope, Evry, France). The sequence
171 chromatograms were assembled using CodonCode Aligner (<http://www.codoncode.com>).

172

173 *Curation and quality control of sequence data*

174

175 The sequences obtained for the first set of markers (COI, 18S, 28S), well represented in sequence
176 databases, were analyzed to detect contaminations at the DNA extraction step by using BLAST (Basic
177 Local Alignment Search Tool) in GenBank (www.ncbi.nlm.nih.gov) or the identification tool in
178 BOLD (for the COI gene) (www.barcodeoflife.org). If, by removing the contaminated DNAs, a taxon
179 is removed from our dataset, a second sampling in the MNHN collection is performed to maintain
180 the expected taxonomic coverage of the Caridea. Specimens from this second sampling are, as the

181 first one, sequenced for the first set of genes to check the quality of the DNA extracts. Additional
182 sampling is performed until enough specimens have been gathered such as to get as close as possible
183 to the expected taxonomic coverage. Then, the other selected genes from the literature (16S, EPRS,
184 H3, NaK, PEPCK, TM9SF4 and β -actine) are sequenced.

185 For each gene-fragment, the sequences were aligned using MUSCLE (Edgar 2004)
186 implemented in MEGA 5.2 (Tamura et al. 2011), and the alignment accuracy was adjusted by eye.
187 Some regions of the non-coding markers (16S, 18S and 28S) were extremely divergent and therefore
188 difficult to align: GBlocks 0.91b (Castresana 2000) was used to omit poorly aligned positions. Single
189 gene-fragment trees were then generated through a Neighbor-Joining (NJ) analysis under MEGA 5.2,
190 with the Maximum Composite Likelihood method and a 100 bootstrap replicates to detect potential
191 incongruences between trees that might indicate the presence of contaminated sequences. Notably,
192 trees for the three fragments of the 18S gene were compared to avoid the creation of chimeric 18S
193 sequence with contamination for one or two fragments of the marker only. For the 18S the final set
194 of selected sequences has thus no supported incongruence (support > 70%) and thus, the fragments
195 were assembled using CodonCode Aligner (www.codoncode.com).

196

197 *Assembly of the taxa x genes matrix*

198

199 Some taxa, because of a relatively high amplification success rate and low contamination rate, were
200 comparatively over-represented in the matrix. To obtain a more balanced matrix for the phylogenetic
201 analysis, only one specimen per species was retained, selecting the specimen for which the highest
202 gene number and/or a species level identification was available. For this selection, a first analysis of
203 the concatenated dataset is done using a NJ tree, with a bootstrap of 100 replicates. The matrix is then
204 rebalanced at the family scale, with a selection of at least two genera per family or per well-supported
205 group for non-monophyletic families, following the same criteria (highest number of genes and/or

206 best identification). To limit artifacts resulting from missing data, only specimens for which at least
207 half of the selected genes have been obtained were retained in the final analysis.

208

209 *Phylogenetic analysis*

210

211 As no well-supported incongruence between single gene NJ trees was observed, the retained gene-
212 fragments were concatenated using CodonCode Aligner (<http://www.codoncode.com>). The dataset
213 was partitioned by gene, and, for the coding genes, by codon position. We applied a GTR+G model
214 for RAxML and GTR+I+G model for MrBayes. All Maximum Likelihood (ML) analysis and
215 Bayesian Inference (BI) analyses were performed using the CIPRES Science Gateway
216 (www.phylo.org/index.php/portal). The ML analysis was conducted under RaxML-HPC2 on XSEDE
217 (v.8.0.9) (Stamatakis 2014). Confidence in the resulting topology was assessed using non-parametric
218 bootstrap estimates (Felsenstein 1985) with 1000 replicates. The BI analysis was done using Mr
219 Bayes 3.2.2 on XSEDE (Ronquist et Huelsenbeck 2003), with the following parameters: 30,000,000
220 generations, 8 chains, 5 swaps, temperature of 0.02, tree sampling frequency of 10,000; all other
221 parameters are set to default. The first 3,000,000 (10%) generations were discarded as “burn-in”.
222 Stability of each parameter (ESS values superior to 200) was checked with Tracer 1.5 (Rambaut and
223 Drummond 2009).

224

225 **Results**

226

227 *Taxonomic coverage*

228 A total of 28 families (out of 38) from 13 superfamilies (out of 14) were included in our analysis
229 (table 1). The families Agostocarididae, Merguiidae, Physetocarididae and Pseudochelidae, each of
230 them only comprising a single genus, were nor represented in the MNHN collections, neither
231 available from our network of collaborators. For the families Bythocarididae (four genera),

232 Desmocarididae (one genus), Euryrynchidae (one genus), Gnathophyllidae (five genera),
233 Ogyrididae (one genus), Thalassocarididae (two genera), Typhlocarididae (one genus) and
234 Xiphocarididae (one genus), although some specimens were available we failed at obtaining at least
235 half of the selected sequences. Sequences of Gnathophyllidae and Xiphocarididae were available in
236 public databases to complement our dataset. Unfortunately, the sequences of Bythocarididae and
237 Merguiidae from De Grave et al. (2014) do not covered the final set of genes selected for the analysis
238 and were thus not included in the present dataset. Similarly for Desmocarididae, Euryrynchidae and
239 Thalassocarididae, sequences were available for only two genes (18S and 16S, Kou et al. 2013) and
240 were thus not added in the analysis.

241

242

243 *Data selection*

244 Our working hypothesis was that the DNA degradation in collection specimens occurs gradually.
245 Following this hypothesis the success rate of PCR amplification should decrease with the specimen
246 age, and the collection date and the number of genes successfully amplified should be positively
247 correlated. This correlation was statistically significant ($R = 0.525$; $n=288$; $\alpha=0.05$) over the whole
248 dataset (Fig. 1). However, the distribution is not uniform over the entire range of dates. The graphic
249 suggests a discontinuity around 1990. Indeed, when testing the correlation for specimens collected
250 before 1990, there is no significant correlation between the date of collect and the amplification
251 success ($R=-0.083$; $n=101$; $\alpha=0.05$), and the success rate is globally low (mean success rate before
252 1990: 41%, and after 1990: 76%). Conversely, for the specimens collected from 1990, the correlation
253 is significantly positive ($R=0.320$; $n=185$; $\alpha=0.05$). This analysis suggests that the degradation of
254 DNA mainly occurs during the first 25 years after collection. An alternative hypothesis is that the
255 fixation protocols in the field and later in the museum were improved during the last 25 years.
256 The recursive sampling in the MNHN collection resulted in 455 extracted specimens. Among those,
257 138 were discarded from the final dataset because neither the COI nor the 18S or the 28S sequences

258 were obtained. As these three genes fragments are easy to amplify, a failure was supposed to result
259 from low-quality DNA. For those specimens, we considered that the quality and/or the quantity of
260 the DNA extracts were too low and the other markers were not tested. For each of these three genes,
261 about 13 % of all sequences obtained were identified as resulting from contaminations. No evidence
262 of intra-individual variability was observed neither for the 28S nor for the 18S fragments. The success
263 rates for each gene are detailed in table 2. Also the DNA extracts that provided contaminated
264 sequences were considered of too low quality and were thus excluded from the subsequent analyses.
265 Using these two criteria 179 DNA extracts were removed from the dataset.

266 For the 276 remaining DNA extracts, we amplified the 16S mitochondrial gene using the
267 universal primers that are usually used for Decapoda. Using these primers about 31% of the sequences
268 matched with human 16S sequences in GenBank. This high rate of contamination is probably
269 explained by the combination of the low quality and quantity of the DNA extracts with poorly specific
270 primers. Therefore, we designed new primers, based on the comparison of available sequences of 16S
271 for several Caridea and *Homo sapiens*, to amplify Caridea DNA preferentially to human DNA. As
272 expected, the success rate of this new pair of primers was largely enhanced and we were able to obtain
273 90% of the specimens for which a human sequence was amplified with the first pair of primers.

274 For some of the tested nuclear coding genes, the sequencing success was very low (table 2).
275 The analysis of the available sequences in GenBank shows that these genes are highly saturated on
276 the 3rd position of the codon, potentially leading to important mismatches with the primers and
277 possibly explaining the low success rate of PCR amplifications. Among the nuclear coding genes,
278 only the H3 gene-fragment was successfully amplified and sequenced in more than half of the
279 specimens. For the β -actine gene 43% of the specimens were successfully amplified. Unfortunately
280 the sequences revealed that several copies of the same size were present in the amplification,
281 suggesting that the designed primers amplify not a single gene but several genes from a multigenic
282 family.

283 At this taxonomic scale, and using specimens from museum collections, we were able to obtain
284 reliable data for only five genes (16S, 18S, 28S, COI and H3) that are then retained for the final
285 phylogenetic analysis. Only for 207 specimens at least three of these five genes were obtained, and
286 among them, the taxonomic coverage was still very unbalanced because of a higher success in some
287 groups. The over-represented taxa were thus subsampled, leading to a final selection of 70 specimens.
288 Taxa with poor coverage in this dataset were re-equilibrated using sequences of 29 species from
289 GenBank.

290

291 *Phylogenetic analysis*

292

293 For non-coding genes, GBlocks deleted a large proportion of the data, even with the least stringent
294 conditions. For example, in the alignment of the 16S sequences, up to 41% of the alignment obtained
295 with MUSCLE was deleted. The MUSCLE alignments were thus checked by eye using GBlocks
296 results as a guide for suppressing poorly aligned parts. The NJ analysis of the single-gene dataset
297 displayed no significant incongruence and the genes were concatenated in an alignment of 4282 bp.
298 No supported incongruence between the trees obtained with ML and BI analysis was detected. The
299 BI tree was used to summarize the results (Fig. 2) because it displayed more supported nodes. Nodes
300 with high support in ML (bootstrap > 70) are also well supported in BI, but several nodes were only
301 supported in BI and are labelled on Fig. 2 with little stars.

302 For some families only a single species was available in our analysis, preventing us to test
303 properly the monophyly of Bresiliidae (two genera), Disciadidae (one genus), Campylonotidae (one
304 genus), Gnathophyllidae (five genera), Hymenoceridae (two genera), Psalidopodidae (one genus) and
305 Xiphocarididae (one genus). However, none of the samples belonging to these families are recovered
306 nested within any of the other family clades. Monophyly is well-supported (BI > 0.96 and/or ML >
307 70) for 19 families: Acanthephyridae (seven genera), Alpheidae (more than ten genera),
308 Alvinocarididae (seven genera), Anchistioidea (one genus), Atyidae (more than ten genera),

309 Barbouriidae (three genera), Bathypalaemonellidae (two genera), Crangonidae (more than ten
310 genera), Eugonatonotidae (one genus), Glyphocrangonidae (one genus), Nematocarcinidae (four
311 genera), Oplophoridae (three genera), Pandalidae (more than ten genera), Pasiphaeidae (seven genera),
312 Processidae (five genera), Rhynchocinetidae (two genera), Stylodactylidae (five genera) Thoridae
313 (eight genera). Three families are non-monophyletic: the Lysmatidae (five genera) Hippolytidae
314 (more than ten genera) and the Palaemonidae (more than ten genera). Relationships among family-
315 level clades are at best poorly resolved with the exception of the superfamily Palaemonoidea that is
316 well-supported (BI: 1, ML: 100). Families are though distributed among two major multi-familial
317 clades, with a high support in BI (BI: 0.99 both), with the exception of Bathypalaemonellidae,
318 Disciadidae and Rhynchocinetidae which have a basal position. The first multi-familial clade (Clade
319 I) includes Atyidae, Psalidopodidae, Stylodactylidae and Xiphocarididae. All other families belong
320 to the second clade (Clade II).

321

322

323 **Discussion**

324 The aim of the study was also to test if the relationships among families may be resolved using
325 currently used markers sequenced over a dense taxonomic coverage. The availability of specimens in
326 the MNHN collection allowed us to enhance the taxonomic coverage of Caridean families for a
327 multigene phylogenetic analysis. However, some taxa are still missing either because they were rare
328 or not available at all in MNHN collections or because the DNA was too degraded in specimens
329 available in MNHN collection. With 28 families analyzed over 38, among which 21 are represented
330 by at least 2 species, our taxonomic coverage was comparable with that of Bracken et al. (2009). The
331 number of genetic markers analysed was comparable with that of Li et al. (2010) but included both
332 nuclear and mitochondrial markers corresponding either to coding or non-coding genes. Among the
333 ten lacking families most include a very small number of species and/or are associated to poorly
334 sampled habitats. For example, the family Desmocarididae only contains one genus and two species

335 (*Desmocarid* *bisliniata* and *D. trispinosa*) both living in estuaries in Africa. Also these families are
336 rare and the scarce material available in the MNHN collections was generally collected before 1990,
337 explaining the poor amplification success.

338 At the family rank, the inclusion within the same dataset of an adequate sampling of major
339 families of Caridea allows us to support with high confidence the monophyly hypothesized in the
340 literature for twelve families (Acanthephyridae, Alpheidae, Alvinocarididae, Atyidae, Crangonidae,
341 Glyphocrangonidae, Nematocarcinidae, Pandalidae, Processidae, Rhynchocinetidae, Styrodactylidae
342 and Thoridae). We also provide new supports for Bathypalaemonellidae and Ophiophoridae but also
343 for two monogeneric families Anchistioidae and Eugonatonotidae. The monophyly of Ophiophoridae,
344 that was suggested in the analysis of Bracken et al. (2009) and then explored in more detail in Chan
345 et al. (2010), is here corroborated thanks to a better coverage of other Caridean families. We also
346 confirm the results of Bracken et al. (2009), Li et al. (2011), Short et al (2013), recovering
347 Palaemonidae polyphyletic and supporting the synonymizing of Kakaducarididae with Palaemonidae.
348 The species traditionally included in Palaemonidae form two well supported clades (noted A and B
349 on Fig. 2), one comprising the genera *Macrobrachium*, *Cryphiops*, and *Leptopalaemon* (previously
350 belonged to the now abandoned family Kakaducarididae), and a second clade for the genera *Leander*,
351 *Palaemon* and *Periclimenes*, sister to the Anchistioididae, Gnathophyllidae and Hymenoceridae
352 families. As Palaemonidae is a very large family containing more than 100 genera with members
353 exhibiting very different morphology, in depth morphological comparison and analysis are necessary
354 to re-build a natural classification system to reflect the relationships within the superfamily
355 Palaemonoidea.

356 The status of Hippolytidae remains unresolved. Among the species sampled in this family,
357 specimens attributed to the genera *Lysmata* and *Ligur* form a clade with the Barbouriidae species of
358 the genus *Parhippolyte*, whereas the position of the eight other Hippolytid species remains
359 unresolved. These results are congruent with the study of Li et al. (2011) in which *Lysmata* is closely
360 related to *Janicea antiguesis* (Barbouridae). A very recent paper (De Grave et al. 2014) explored the

361 status of Hippolytidae using several genes and a larger taxonomic coverage within clade B. This study
362 confirmed that Hippolytidae as defined in De Grave and Fransen (2011) are not monophyletic. The
363 strongly supported close relationships of *Eualus gaimardii* and *Lebbeus polaris* revealed by the
364 present study agrees with the resurrection of the family Thoridae by De Grave et al. (2014). Although
365 Barbouriidae is confirmed to be closely related to *Lysmata* spp. and *Ligur ensiferus* – two genera
366 attributed to the family Lysmatidae recently resurrected by De Grave et al. (2014) – the present result
367 revealed that *Ligur ensiferus* is sister to Barbouriidae with very strong support.

368 The Pasiphaeidae (with seven genera) are found monophyletic. However, since we were not
369 able to obtain sequences for the second genus sampled in the MNHN collection (*Leptochela*), this
370 result does not challenge the results of Bracken et al. (2009). Similarly, Barbouriidae (with four
371 genera) is recovered monophyletic but since only two species of the same genus are included this
372 result remains to be confirmed with additional sampling of species from other genera.

373 The relationships among families are generally poorly resolved even so two major clades are
374 supported. A first clade includes four families (Clade I: Atyidae, Psalidopodidae, Stylodactylidae and
375 Xiphocarididae). Within clade I, two sister lineages are supported: Psalidopodidae and
376 Stylodactylidae on one hand and Atyidae and Xiphocarididae on the other. The close relationship
377 between the two latter families was already pointed out by Page et al. (2008 a) and Bracken et al.
378 (2009). Clade II is more diverse and includes 23 major lineages, most of them corresponding to
379 traditionally defined families. However, the deeper nodes are not supported. Clade II includes the
380 Palaemonoidea which is the only well-supported superfamily in the Caridea. Within this superfamily
381 two clades may be distinguished. First a well-supported clade (BI: 1, ML: 98) includes
382 *Macrobrachium*, *Cryphiops* and *Leptopalaemon* (currently all attributed to Palaemonidae).
383 Anchistioididae, Gnathophyllidae, Hymenoceridae and all other Palaemonidae genera form a second
384 clade. Within the latter clade, Gnathophyllidae and Hymenoceridae are closely related. However, the
385 sampling within these two families is still too restricted to validate their close relationships.

386 The five latest major taxonomic revisions within the Caridean classification are almost all
387 corroborated. (1) Although we do not test the position of the Procarididae among other infraorder of
388 Decapoda, the position of the two Procarididae included in the dataset as sister-group of the Caridea
389 is congruent with their reclassification as a distinct infraorder. This node is well-supported (BI: 1,
390 ML: 100) and the branch lengths do not suggest that this position results from a reconstruction artifact.
391 Indeed branch lengths for the two Procarididae are close to that of other included outgroups, notably
392 that of *Cancer pagurus*. (2) The Acanthephyridae and the Oplophoridae are two distinct lineages with
393 high support (BI: 1, ML: 100 for both). However, the relationship between these two families is not
394 resolved and thus the definition of the superfamily Oplophoroidea cannot be rejected. (3) Within the
395 family Eugonatonotidae, the recent synonymization of *Galatheacaris* with *Eugonatonotus* is
396 supported (De Grave et al. 2010). (4) The recent synonymization of Kakaducarididae with
397 Palaemonidae (Short et al. 2013) is also well supported. (5) Only the very recent separation of
398 Hippolytidae by De Grave et al. (2014) is not completely supported. Although the present result also
399 reveals that Lysmatidae is close to Barbouriidae, the former is showed to be polyphyletic. The
400 monophyly of the resurrected Thoridae is strongly supported but Hippolytidae s.s. may still be
401 polyphyletic.

402 Our analysis shows that additional data are needed to revise some families, notably
403 Palaemonidae, Hippolytidae *s.l.* and Pasiphaeidae. However such a revision needs additional
404 taxonomic sampling. Although not fully resolved at the deeper nodes, the phylogenetic hypothesis
405 presented here may provide a guide for the taxonomic sampling of future family-level revisions and
406 notably for the selection of adequate outgroups.

407 One of our working hypotheses was that higher taxonomic coverage will improve the
408 phylogenetic resolution of multi-gene analyses. This approach allowed us to corroborate the proposed
409 monophyly of some families but did not provide significant resolution for the deeper nodes of the
410 Caridea tree. The poor success of amplification of nuclear coding genes points to the limits of PCR
411 amplification associated with Sanger sequencing in multigene phylogenetic analyses. A

412 phylogenomic analysis based on the generation of a large amount of through next-generation
413 sequencing might provide the data needed for a better resolution between the oldest Caridean
414 lineages. However, at present the genomic data available for the Caridea are rather scarce. No
415 Caridean genome has been sequenced and only few transcriptome datasets are available for two
416 Palaemonidae species, *Macrobrachium rosenbergii* (Mohd-Shamsudin et al. 2013) and *M.*
417 *nipponense* (Ma et al. 2012, Jin et al. 2013). The most documented type of genomic data remains the
418 mitogenomes that are available for 11 Caridean species (Miller et al. 2005, Ivey et al. 2007, Shen et
419 al. 2009, Qian et al. 2011, Kim et al. 2013, Yang et al. 2012, Yang et al. 2013). Indeed, sequencing
420 complete mitochondrial genomes is easier and more affordable than sequencing complete genomes or
421 transcriptomes. However, mitogenomics will provide the reconstruction of mitochondrion
422 genealogies that might be only partly correlated to organisms' genealogy. The sequencing of
423 complete genome for comparative studies is hampered by the very large and variable genome size
424 within Caridea (Rees et al. 2008). Moreover, the analysis of the variation of the genome size across
425 populations revealed unexpected intraspecific variation in the Alvinocarid shrimp *Mirocaris*
426 *fortunata* (Bonnivard et al. 2009). In this context and with the objective of resolving deeper
427 phylogenetic nodes, the sequencing of transcriptomes might be a way to define loci that are widely
428 shared and not too variable to provide the adequate phylogenetic signal at this phylogenetic depth.

429

430

431 **Acknowledgement**

432 We are grateful to Bertrand Richer de Forges and Philippe Bouchet, cruise leaders of several deep-
433 sea cruises of the Tropical Deep-Sea Benthos program on board R/V Alis, that generated most of the
434 samples used in this study. We also thank Philippe Keith for providing freshwater species and Pierre
435 Chevaldonné for the specimens of *Bresilia saldanhai*. All material has been collected under
436 appropriate collection permits and approved ethics guidelines. This project was supported by the
437 network "Bibliothèque du Vivant" funded by the CNRS, the Muséum National d'Histoire Naturelle,

438 the INRA and the CEA (Genoscope), the French-Taiwanese project TF-DeepEvo funded by ANR
439 (ANR 12-ISV7-0005-01) and Ministry of Science and Technology, Taiwan, R.O.C., and the project
440 “Taxonomie moléculaire: DNA Barcode et gestion durable des collections” funded by the Muséum
441 National d'Histoire Naturelle.

442

443

444 **Conflict of interest**

445

446 The authors declare that they have no conflict of interest.

447

448

449 **References**

450 Ah Yong ST, Schnabel KE, Macpherson E (2011) Phylogeny and fossil record of marine squat lobsters
451 in In: Poore GB, Ah Yong ST & Taylor J (Eds.), *The Biology of Squat Lobsters*. CSIRO Publishing.
452 73-104.

453

454 Bonnivard E, Catrice O, Ravaux J, Brown SC, Higuét D (2009) Survey of genome size in 28
455 hydrothermal vent species covering 10 families. *Genome* 52: 524-536.

456 Bracken HD, De Grave S, Felder DL (2009) Phylogeny of the infraorder Caridea based on
457 mitochondrial and nuclear genes (Crustacea: Decapoda). In: Martin JW, Crandall KA, Felder DL
458 (Eds.) *Decapod Crustacean Phylogenetics*. Crustacean Issues 18. CRC Press, Boca Raton, FL,
459 USA, pp 274–300.

460 Bracken HD, De Grave S, Toon A, Felder DL, Crandall KA (2010) Phylogenetic position, systematic
461 status, and divergence time of the Procarididea (Crustacea: Decapoda). *Zool Scripta* 39: 198–212.

462 Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in
463 phylogenetic analysis. *Mol Biol Evol* 17: 540-552.

464 Chace FA Jr (1992) On the classification of the Caridea (Decapoda). *Crustaceana* 63: 70–80.

465 Chan TY, Lei HC, Li CP, Chu KH (2010) Phylogenetic analysis using rDNA reveals polyphyly of
466 Oplophoridae (Decapoda: Caridea). *Invertebr Syst* 24: 172-181.

467 De Grave S, Li CP, Tsang LM, Chu KH, Chan TY (2014) Unweaving hippolytoid systematics
468 (Crustacea, Decapoda, Hippolytidae): resurrection of several families. *Zool Scripta* 43: 496-507.
469 DOI: 10.1111/zsc.12067

470 De Grave S, Chan T-Y, Chu KH (2010) On the systematic position of *Galatheacaris abyssalis*
471 (Decapoda: Galatheacaridoidea). *J Crust Biol* 30: 521–527.

472 De Grave S, Fransen CHJM (2011) Carideorum catalogus: the recent species of the dendrobranchiate,
473 stenopodidean, procarididean and caridean shrimps (Crustacea: Decapoda). *Zoologische*
474 *Mededelingen*, 85: <http://www.zoologischemededelingen.nl/85/nr02/a01> .

475 De Grave S, Pentcheff ND, Ahyong S, Chan T-Y, Crandall KA, Dworschak P, Felder DL, Feldmann
476 RM, Fransen CHJM, Goulding LYD, Lemaitre R, Low ML, Martin JW, Ng PKL, Schweitzer CE,
477 Tan SH, Wetzer R (2009) A classification of living and fossil genera of decapod crustaceans.
478 *Raffles Bull Zool Suppl* 21: 1–109.

479 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.
480 *Nucleic Acids Res* 32:1792-1797.

481 Ivey JL, Santos SR (2007) The complete mitochondrial genome of the Hawaiian anchialine shrimp
482 *Halocaridina rubra* Holthuis, 1963 (Crustacea: Decapoda: Atyidae). *Gene* 394: 35-44.

483 Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*
484 39: 783 – 791.

485 Jin S, Fu H, Zhou Q, Sun S, Jiang S, Xiong Y, Gong Y, Qiao H, Zhang W (2013) Transcriptome
486 Analysis of Androgenic Gland for Discovery of Novel Genes from the Oriental River Prawn,
487 *Macrobrachium nipponense*, Using Illumina Hiseq 2000. *PLoS ONE* 8: e76840.

488 Keane TM, Creevey CJ, Pentony MM, Naughton, TJ, McInerney JO (2006). Assessment of methods
489 for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for
490 choice of matrix are not justified. *BMC Evol Biol* 6, 1–17.

491 Kim SJ, Pak SJ, Ju SJ, (2013) Mitochondrial genome of the hydrothermal vent shrimp *Nautilocaris*
492 *saintlaurentae* (Crustacea: Caridea: Alvinocarididae). Mitochondrial DNA:
493 doi:10.3109/19401736.2013.815169.

494 Kou Q, Li X, Chan T-Y, Chu KH, Gan Z (2013). Molecular phylogeny of the superfamily
495 Palaemonoidea (Crustacea: Decapoda: Caridea) based on mitochondrial and nuclear DNA reveals
496 discrepancies with the current classification. *Invertebr Syst* 27: 502-514.

497 Li CP, De Grave S, Chan T-Y, Lei HC, Chu KH (2011) Molecular systematics of caridean shrimps
498 based on five nuclear genes: implications for superfamily classification. *Zool Anz* 250: 270-279.

499 Ma K, Qiu G, Feng J, Li J (2012) Transcriptome analysis of the oriental river prawn, *Macrobrachium*
500 *nipponense* using 454 pyrosequencing for discovery of genes and markers. *PLoS ONE* 7: e39727.

501 Martin JW, Crandall KA, Felder DL (2010) Decapod crustacean phylogenetics. CRC Press, Boca
502 Raton, FL, USA

503 Miller AD, Murphy NP, Burrige CP, Austin CM (2005) Complete mitochondrial DNA sequences
504 of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii*
505 (Palaemonidae). *Mar Biotechnol* 7:339-49.

506 Mohd-Shamsudin MI, Kang Y, Lili Z, Tan TT, Kwong QB, Liu H, Zhang G, Othman RY, Bhassu S
507 (2013) In-depth transcriptomic analysis on giant freshwater prawns. *PLoS ONE* 8: E60839.

508 Page TJ, Cook BD, von Rintelen T, von Rintelen K, Hughes JM (2008 a) Evolutionary relationships
509 of atyid shrimps imply both ancient Caribbean radiations and common marine dispersals. *J N Am*
510 *Benthol Soc* 27: 68–83.

511 Page TJ, Short JW, Humphrey CL, Hillyer MJ, Hughes JM (2008 b) Molecular systematics of the
512 Kakaducarididae (Crustacea: Decapoda: Caridea). *Mol Phyl Evol*, 46: 1003-1014.

513 Puillandre N, Bouchet P, Boisselier-Dubayle MC, Brisset J, Buge B, Castelin M , Chagnoux S,
514 Christophe T, Corbari L, Lambourdière J, Lozouet P, Marani G, Rivasseau A, Silva N, Terryn Y,
515 Tillier S, Utge J, Samadi S (2012) New taxonomy and old collections: integrating DNA barcoding
516 into collections curation processes. *Mol Ecol Resour* 12: 396-402.

517 Qian GH, Zhao Q, Wang A, Zhu L, Zhou K, Sun H (2011) Two new decapod (Crustacea,
518 Malacostraca) complete mitochondrial genomes: bearings on the phylogenetic relationships within
519 the Decapoda. *Zool J Linn Soc* 162: 471-481.

520 Rambaut A, Drummond AJ (2009) Tracer v1.5 README [Documentation file]. Available with the
521 installation files at <http://tree.bio.ed.ac.uk/software/tracer/>

522 Rees DJ, Belzile C, Glémet H, Dufresne F (2008) Large genomes among caridean shrimp. *Genome*
523 51: 159-163.

524 Richer de Forges B, Chan T-Y, Corbari L, Lemaitre E, Macpherson E, Ahyong ST, Ng PKL (2013)
525 The MUSORSTOM-TDSB deep sea Benthos exploration programme (1976-2012): An overview
526 of crustacean discoveries and new perspectives on deep-sea zoology and biogeography. In Ahyong
527 A, Chan T-Y, Corbari L, Ng P (eds) *Tropical Deep-Sea Benthos*, Volume 27, pp 13-66

528 Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed
529 models. *Bioinformatics* 19:1572-1574.

530 Shen X, Sun M, Wu Z, Tian M, Cheng H, Zhao F, Meng X (2009) The complete mitochondrial
531 genome of the ridgetail white prawn *Exopalaemon carinicauda* Holthuis, 1950 (Crustacean:
532 Decapoda: Palaemonidae) revealed a novel rearrangement of tRNA genes. *Gene* 437:1-8.

533 Shen H, Braband A, Scholtz G (2013) Mitogenomic analysis of decapod crustacean phylogeny
534 corroborates traditional views on their relationships. *Mol Phyl Evol* 66: 776-789.

535 Short JW, Humphrey CL, Page TJ (2013) Systematic revision and reappraisal of the Kakaducarididae
536 Bruce (Crustacea : Decapoda : Caridea) with description of three new species of *Leptopalaemon*
537 Bruce & Short. *Invertebr Syst* 27: 87-117.

538 Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large
539 Phylogenies. *Bioinformatics* 30: 1312–1313.

540 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular
541 Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and
542 Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731-2739.

543 Toon A, Finley M, Staples J, Crandall KA (2009) Decapod phylogenetics and molecular evolution.
544 In: Martin JW, Crandall KA, Felder DL (Eds.) *Decapod Crustacean Phylogenetics*. Crustacean
545 Issues 18. CRC Press, Boca Raton, FL, USA, pp 15–30.

546 Tsang LM, Ma KY, Ahyong ST, Chan T- Y, Chu KH (2008) Phylogeny of Decapoda using two
547 nuclear protein-coding genes: origin and evolution of the Reptantia. *Mol Phyl Evol* 48: 359-368.

548 Tsang LM, Schubart CD, Ahyong ST, Lai JC, Au EY, Chan TY, NG PKL, Chu KH (2014)
549 Evolutionary History of True Crabs (Crustacea: Decapoda: Brachyura) and the Origin of
550 Freshwater Crabs. *Mol Biol Evol* 31: 1173-1187.

551 Yang CH, Tsang LM, Chu KH, Chan TY (2012) Complete mitogenome of the deep-sea hydrothermal
552 vent shrimp *Alvinocaris chelys* Komai and Chan, 2010 (Decapoda: Caridea: Alvinocarididae).
553 *Mitochondrial DNA* 23:417-9.

554 Yang JS, Lu B, Chen DF, Yu YQ, Yang F, Nagasawa H, Tsuchida S, Fujiwara Y, Yang WJ (2013)
555 When did decapods invade hydrothermal vents? Clues from the Western Pacific and Indian
556 Oceans. *Mol Biol Evol* 30: 305-309.

557 Zuccon D, Brisset J, Corbari L, Puillandre N, Samadi S (2012) Optimized protocol for barcoding
558 museum collections of Decapoda crustaceans: a case-study for a 10-40 years old collection.
559 *Invertebr Syst* 26: 592–600.

560
561
562

563 **Figure captions**

564

565 Fig. 1. Histograms of the number of specimens per date of collect, for which respectively the
566 sequences of 0, 1, 2, 3, 4 or 5 genes were successfully obtained (i.e. number obtained after excluding
567 the potentially contaminated sequences). Correlation between the date of collect of the specimens,
568 and the number of validated sequences for the five selected genes.

569

570 Fig. 2. Caridea phylogeny: IB topology, with posterior probability shown when > 0.94 .

571 *: nodes only supported in BI. All other nodes have similar support both in ML and in BI (bootstrap
572 > 70 in ML and posterior probability > 0.96 in BI).

573

574

575

576

578 **Table 1.** Taxonomic sampling, in number of specimens per family, in the articles of Bracken et al.
 579 (2009), Li et al. (2011) and in this present study. Family status according to each study is indicated
 580 as follow: Yes: monophyly / No: non-monophyly / ?: monophyly non tested / -: no data.
 581

Superfamily	Family	Bracken et al. 2009		Li et al. 2011		This study	
		Nb sp	Monophyly	Nb sp	Monophyly	Nb sp	Monophyly
Alpheoidea	Alpheidae	15	Yes	3	Yes	3	Yes
	Barbouriidae	-	-	1	?	2	Yes
	Bythocarididae	-	-	-	-	-	-
	Hippolytidae	8	No	1	?	6	No
	Lysmatidae	4	Yes	1	?	3	No
	Merguiidae	-	-	-	-	-	-
	Ogyrididae	2	Yes	-	-	-	-
Thoridae	1	?	-	-	2	Yes	
Atyoidea	Atyidae	10	Yes	2	Yes	8	Yes
Bresilioidea	Agostocarididae	1	?	-	-	-	-
	Alvinocarididae	4	Yes	1	?	8	Yes
	Bresiliidae	-	-	-	-	1	?
	Disciadidae	1	?	-	-	1	?
	Pseudochelidae	-	-	-	-	-	-
Campylonotoidea	Bathypalaemonellidae	1	?	1	?	2	Yes
	Campylonotidae	-	-	1	?	1	?
Crangonoidea	Crangonidae	3	Yes	2	Yes	4	Yes
	Glyphocrangonidae	2	?	2	Yes	2	Yes
Nematocarcinoidea	Eugonatonotidae	1	?	1	?	3	Yes
	Nematocarcinidae	3	Yes	3	Yes	5	Yes
	Rhynchocinetidae	1	?	4	Yes	2	Yes
	Xiphocarididae	1	?	-	-	1	?
Oplophoroidea	Acanthephyridae	5	Yes	-	-	2	Yes
	Oplophoridae	2	?	1	?	2	Yes
Palaemonoidea	Anchistioidea	1	?	-	-	2	Yes
	Desmocarididae	1	?	-	-	-	-
	Euryrhyndidae	1	?	-	-	-	-
	Gnathophyllidae	3	No	1	?	1	?
	Hymenoceridae	1	?	1	?	1	?
	Palaemonidae	18	No	2	No	13	No
	Typhlocarididae	1	?	-	-	-	-
Pandaloidea	Pandalidae	6	Yes	2	Yes	9	Yes
	Thalassocarididae	1	?	-	-	-	-
Pasiphaeoidea	Pasiphaeidae	5	No	2	Yes	4	Yes
Physetocaridoidea	Physetocarididae	-	-	-	-	-	-
Processoidea	Processidae	4	Yes	-	-	2	Yes
Psolidopodoidea	Psolidopodidae	1	?	-	-	1	?
Stylodactyloidea	Stylodactylidae	2	Yes	3	Yes	3	Yes

583 **Table 2.** Sequencing success rates for each gene, based on the number of sequences obtained post-
 584 detection of contaminations. The first three genes are the reference markers. The other genes are
 585 tested only for specimens successfully amplified for the references genes.
 586

Gene			Success
18S	Nuclear	non-coding	71%
28S	Nuclear	non-coding	60%
COI	Mitochondrial	Coding	67%
16S	Mitochondrial	non-coding	54%
16S (new primers)	Mitochondrial	non-coding	78%
β -actine	Nuclear	Coding	43%
EPRS	Nuclear	Coding	0%
H3	Nuclear	coding	62%
NaK	Nuclear	coding	4%
PEPCK	Nuclear	coding	4%
TM9SF4	Nuclear	coding	0%

587



