



University of  
**Salford**  
MANCHESTER

**Understanding the evolution of cymothoid isopod parasites using comparative  
genomics and geometric morphometrics**

Charles Baillie

January 2020

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Environment & Life Sciences

University of Salford, Manchester



# Table of Contents

|   |            |
|---|------------|
| <b>Table of Contents</b> . . . . .  | <b>iii</b> |
| <b>Abstract</b> . . . . .   | <b>v</b>   |
| <b>Preface</b> . . . . .  | <b>vii</b> |
| <b>Acknowledgements</b> . . . . .   | <b>ix</b>  |
| <b>List of Tables</b> . . . . .   | <b>xi</b>  |
| <b>List of Figures</b> . . . . .  | <b>xii</b> |
| <b>1 Introduction</b> . . . . .   | <b>1</b>   |
| 1.1 Parasitism . . . . .  | 2          |
| 1.2 Cymothoidae (Leach, 1818) . . . . .   | 3          |
| 1.3 Museomics . . . . .   | 12         |
| 1.4 Thesis Structure . . . . .  | 14         |
| 1.5 Aims and Objectives . . . . .   | 15         |
| <b>2 Hooked on you: shape of attachment structures in cymothoid isopods<br/>reflects parasitic strategy</b> . . . . . | <b>17</b>  |
| 2.1 Introduction . . . . .  | 18         |
| 2.2 Methods . . . . .   | 21         |
| 2.3 Results . . . . .   | 27         |
| 2.4 Discussion . . . . .  | 32         |
| 2.5 Conclusions . . . . .   | 37         |
| <b>3 Mitochondrial genomes and isopod phylogeny</b> . . . . .   | <b>38</b>  |
| 3.1 Introduction . . . . .  | 39         |
| 3.2 Materials and Methods . . . . .   | 43         |

---

|   |            |
|---|------------|
| 3.3 Results . . . . .   | 54         |
| 3.4 Discussion . . . . .  | 67         |
| 3.5 Conclusion . . . . .  | 72         |
| <b>4 Ancient DNA analysis supports multiple transitions between parasitic strategies in the evolution of cymothoid isopods, and a single colonisation of freshwater in South America. . . . .</b> | <b>74</b>  |
| 4.1 Introduction . . . . .  | 75         |
| 4.2 Materials and Methods . . . . .   | 79         |
| 4.3 Results . . . . .   | 92         |
| 4.4 Discussion . . . . .  | 98         |
| 4.5 Conclusion . . . . .  | 104        |
| <b>5 Conclusions . . . . .</b>  | <b>105</b> |
| <b>Bibliography . . . . .</b>   | <b>111</b> |
| <b>Appendix . . . . .</b>   | <b>143</b> |

## Abstract

One of the single most extraordinary examples of host-parasite co-evolution is shown by the isopod family Cymothoidae, of which all species are obligate parasites of fishes, including many commercially important fish species. Cymothoids are one of the most diverse isopod groups, with 400 species across 43 genera, and they exhibit striking parasitic strategies. Some species, for example, are known to supplant their host's tongue. In addition to mouth attachment, cymothoids also attach externally to the skin, within the gills, or burrow within the body cavities of their host. The majority of species use only one of these attachment strategies, and attachment location is also largely conserved within cymothoid genera. Yet, there is variation in microhabitat use between species with the same parasitic mode, because distinct locations or orientations are used. As well as specificity in parasitic strategies, cymothoids are highly host specific, with most species restricted to a few host species.

Due to their bizarre life-histories and their large size relative to their hosts, cymothoids have been studied since the early 19th century. However, the majority of this work is related to traditional taxonomy and, hampered by high intraspecific variation and lack of an evolutionary framework for the group, species boundaries and relationships between taxa remain unclear. Preliminary molecular evidence suggests that the evolution of attachment in cymothoids has a complex history, but to better understand cymothoid-host co-evolution a more densely sampled phylogeny is required.

Here we examine several aspects of cymothoid evolution using museum specimens. First, we use a geometric morphometric approach to quantify morphological variation

in attachment appendages, and relate this to the different parasitic strategies and phylogeny. In the process of producing mitogenome reference sequences for cymothoids, we investigate the placement of Cymothoidae within Isopoda, test the monophyly of several isopod subclades, and investigate the origin of isopod terrestrialisation. Finally, using our reference mitogenome sequences, we recover further mitogenome sequences from museum specimens using a low-coverage shotgun sequencing approach, and use these data to reconstruct cymothoid phylogeny.

We show that, after accounting for shared ancestry, attachment morphology is strongly influenced by parasitic strategy. Whole mitogenome sequences are not suitable for resolving isopod inter-familial relationships. We find some evidence that cymothoids, and possibly all isopods, have an atypical mitogenome structure (previously characterised), structural variation in which causes compositional biases and long-branch artefacts. Sophisticated modelling and data treatment can alleviate some of these effects but we caution the validity of inferred relationships based on mitogenome data. We successfully obtain sequence data from liquid-preserved museum specimens; find that the evolution of attachment in cymothoids is complex, and that freshwater species from South America likely colonised rivers in a single event.

# **Preface**

## **Statement of Originality**

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due references are made.

## **Ethical Approval**

This research was classed as 'Type I' and did not require ethics approval. Confirmation of the activities to be undertaken as part of this thesis and a signed Research Ethics Form was submitted to the School in January 2016.

## **Funding**

This work was primarily funded through the University of Salford Postgraduate Research Program. Robin M. D. Beck and Charles Baillie were also awarded a research grant from the University of Salford of £1,600. Charles Baillie was also awarded grants from:

- Santander Bank, Student Travel Award: £5,000
- Systematics Association, Student Research Award: £1,000

- European Society of Evolutionary Biologists, Godfrey Hewitt Mobility Award: £1,400
- Genetics Society, Heredity Fieldwork Grant: £1,500

## **Publications**

The second chapter of this thesis has been published in *BMC Evolutionary Biology* with the following DOI:10.1186/s12862-019-1533-x. Charles Baillie, Robin Beck and Stefano Mariani conceived the study. Kerry Hadfield completed the imaging and Charles Baillie and Rachel Welicky, the landmarking. Charles Baillie, Kerry Hadfield, and Rachel Welicky completed the molecular work. Charles Baillie carried out statistical analyses and wrote the manuscript, and all authors reviewed and edited the final version.

# Acknowledgements

Sarah, I would not have completed a fraction of this work without your love, extraordinary tolerance, and boundless support. It has been a tough, yet rewarding, few years. Thank you for being so attentive to our little guys when it was difficult for me to be there, for creating the space that allowed me to pursue my goals, for pretending to listen when I talked about cymothoid evolution, and for keeping me smiling — I really can't thank, or love you enough.

Of course, none of this would have been possible without my wonderful supervisors, Robin, Stefano, and, latterly, Ian. Thank you for your many wise words, advice, direction, for picking me up in the hard times, and giving me a shot in the first place. I have learned so much from you, and I promise to pass it on.

For showing me the aDNA ways, sharing great laughs, and for getting me out of a few scrapes, Konstantina you have been an all-round star and I'm very happy to count you among my best friends and collaborators.

Beth Shapiro and Terry Brown showed me incredible kindness, and freely opened their labs, and expertise, for my research. I'd especially like to thank Beth (I did visit twice after all!). I would not have been able to complete any of my US sample work without you and your team, and I learned a lot about the research world during my short visits: you and Ed have built a model lab, in my opinion.

I was lucky enough to work in some incredible institutions during the completion of this thesis, and there are many, many curators, friends, and collaborators to thank. In chronological order in which I met them, these wonderful people include: Laure Corbari

at MNHN; Célio Magalhães and Rafaela Ota at INPA; Ananda, Milla, Stefani, Sarita Loureiro, and Fabricio at UFPA; Miranda Lowe at NHM; Karen Osborn and Karen Reed at USNM; all the gang at the Paleogenomics Lab, UCSC, but especially Nedda, Jannine, and Sabrina; and last but by no means least, Kerry, Nico, and Rachel at NWU — still lots of work to be done together. I should also mention Donal, just because (he'll understand). Special thanks to Jonathan, Wilsea, Bethany, and Felix, who took me in while I was working in Belém, and showed me such hospitality, even in difficult times.

For supporting, and joining, our worldly adventures for the past four years, for looking after our little adventures every week, and for your unfathomable ability to spot a misplaced comma at twenty paces, Val we love you and we would be a mess without you.

Dad, without your love, steadiness, appreciation of the natural world, and veiled appreciation of health and safety, which allowed me and my sisters to rampage around the West Coast in a rib from early on, I probably would not have considered Biology. Thanks. And, to my mum, who 'doesn't believe in evolution', I probably would not have considered Biology if you did.

# List of Tables

|     |   |    |
|-----|---|----|
| 2.1 | List of Landmarked Specimens . . . . .  | 22 |
| 2.2 | Results from OLS regressions of the full $P1$ and $P7$ datasets . . . . .     | 29 |
| 2.3 | Results from PGLS Regressions of $P1_{phy}$ and $P7_{phy}$ Datasets . . . . . | 31 |
| 3.1 | Species and Mitogenome Sequence Analyses . . . . .                            | 44 |
| 3.2 | Assembly Strategy . . . . .   | 48 |
| 3.3 | Mapping and Coverage Statistics . . . . .                                     | 54 |
| 3.4 | Mitogenome Annotation . . . . .   | 56 |
| 3.5 | Results of Posterior Predictive Analysis . . . . .                            | 64 |
| 3.6 | Saturation and Composition Tests . . . . .                                    | 67 |
| 4.1 | Specimen Collection . . . . .   | 80 |
| 4.2 | Ancient DNA Mapping . . . . .   | 93 |

# List of Figures

|     |  |     |
|-----|--|-----|
| 1.1 | Number of marine Cymothoidae in biogeographic regions . . . . .  | 6   |
| 1.2 | Images of Different Cymothoid Parasitic Strategies . . . . .   | 8   |
| 2.1 | Landmark Design for Cymothoid Dactyli . . . . .  | 23  |
| 2.2 | Morphospaces of <i>P1</i> and <i>P7</i> Dactyli . . . . .  | 28  |
| 2.3 | Phylogeny and Phylomorphospaces of <i>P1<sub>phy</sub></i> and <i>P7<sub>phy</sub></i> Dactyli . . . . . | 33  |
| 3.1 | Dotplot of <i>Anilocra physodes</i> and <i>Ligia exotica</i> Draft Assemblies . . . . .                  | 57  |
| 3.2 | Mitochondrial Genome Map — <i>Ceratothoa italica</i> . . . . .   | 59  |
| 3.3 | Mitochondrial Genome Maps — <i>Cinusa tetrodontis</i> . . . . .  | 60  |
| 3.4 | Mitochondrial Genome Maps — <i>Anilocra physodes</i> . . . . .   | 61  |
| 3.5 | Mitochondrial Genome Maps — <i>Ligia oceanica</i> . . . . .  | 62  |
| 3.6 | Site-heterogeneous Dayhoff-recoded Phylogeny . . . . .   | 65  |
| 4.1 | Influences on DNA Yield . . . . .  | 95  |
| 4.2 | Metagenomic Analysis . . . . .   | 97  |
| 4.3 | Museum Mitogenome Phylogeny . . . . .  | 99  |
| 4.4 | Museum COI Phylogeny . . . . .   | 100 |
| 1   | ML Partitioned Trees (nucleotides) . . . . .   | 144 |
| 2   | BI Partitioned Trees (nucleotides) . . . . .   | 145 |
| 3   | Amino Acid Partitioned Trees . . . . .   | 146 |
| 4   | Site-heterogeneous Trees . . . . .   | 147 |

# Introduction

## 1.1 Parasitism

At the origin of Charles Darwin's 'entangled bank' (Darwin, 1859) are the most important and intriguing, yet invisible, aspects of evolutionary biology. Not the organisms themselves but the processes by which they come to be — the exchanges and associations between individuals and the outcomes of those interactions. Parasitism, perhaps the most intimate of interactions, has evolved with high frequency throughout history in species from across the tree of life and is the most widely adopted, and successful, mode of life amongst living organisms (Poulin and Morand, 2000; Dobson et al., 2008). Highlighting this success, the single most speciose metazoan order, Hymenoptera, comprises thousands of species of parasitoid wasp, which outnumber even the Coleoptera. Yet, despite the ubiquity of parasitism, it is not straightforward to define, and parasites are variously categorised by where they attach, and for how long, by what they eat, by the developmental stage in which they are infectious, by how many host species are required to complete a life cycle, by the way they are transmitted, and by how harmful they are. Brood parasites do not even live on, or in, their hosts, as with most parasites, but exploit parental investment to dupe other bird species into rearing their young. Broadly, however, there are six evolutionary modes of parasite, crossed between whether the host survives, whether the host dies, by the number of hosts parasitised, and whether the hosts can reproduce following infestation.

Understanding the evolution of parasitism is important not least because of their numbers, but also because of their influence on ecosystems, and roles in moderating consumer-resource interactions of free-living species. The nature of parasite evolution is complex, but some broad patterns emerge. Parasites typically have shorter generation times and larger population sizes than their hosts and, therefore, often show elevated evolutionary rates compared with free-living species. Red Queen dynamics, reciprocal

adaptations between host immune response and parasite virulence, shape the evolution and diversification of parasite species — an arms race which over evolutionary time, in different contexts, can trend toward optima increasing parasite virulence or producing more benign relationships, or mutualism (van Valen, 1973). Patterns of parasite diversity are similarly influenced by the traits of both hosts and parasites: the ability of parasites to switch hosts, host characteristics allowing intrahost speciation, host and parasite life history traits, and population dynamics, all inter-play to influence parasite species richness and biogeography.

It is through these complex co-evolutionary dynamics that parasites have evolved some truly remarkable adaptations. Wingless flies (Nycteribiidae), ant-zombie creating fungi (Cordyceps), few-celled jellyfish with tiny genomes (Myxozoa), worm-resembling intestinal snails (Enteroxenos), and sprouting-limb causing flatworm larvae (*Ribeiroia* sp.) are but a few evolutionary products associated with this fascinating lifestyle. Yet, there is one particular group which perhaps fuels evolutionary curiosity more than any other, and that is the cymothoid isopods, including the famous ‘tongue-biters’.

## **1.2 Cymothoidae (Leach, 1818)**

### **Life-history**

Cymothoidae is a monophyletic taxon of marine and freshwater isopods, all members of which are obligate ectoparasites of fishes (Smit et al., 2014). All cymothoids are sexually-reproducing, sequential, protandrous hermaphrodites, and have a complex biology and life-cycle. In common with all isopods, embryo and larval development takes place within the marsupium, or brood pouch, of the adult female. By the time cymothoids are released from the brood pouch, they have undergone two metamorphoses but are as yet sexually undifferentiated. Now termed ‘mancae’, young cymothoids are temporarily free-living,

and use their well-developed eyes and long setae, which facilitate swimming, to actively hunt for a suitable host. Cymothoids must find a host within approximately seven days though some species can use intermediate hosts before permanently attaching to a definitive host (Brusca, 1978; Varvarigos, 2003). Once established, mancae undergo a further moult to become 'juveniles', during which male gonads develop. Transition to becoming an adult female is thought to be triggered by a complicated combination of cues but is strongly dependent upon the presence of an established female, which appears to suppress transformation into females of other individuals (Bowman, 1960; Smit et al., 2014). Under natural conditions, intensity of infestation is, therefore, low and most commonly a single mating pair occupy a host (Brusca, 1981).

Large size differences between the sexes are usual, as they compete for the same resource, and in general cymothoids are large compared with their hosts. Cymothoid length has been shown to scale linearly in external- and mouth-attaching adult cymothoids, but not so for gill-attaching species, or for juveniles (Welicky et al., 2019). The effects of infestation also vary considerably between species. Reduced host condition factors, growth, aerobic capacity, parasitic castration, tissue damage, and secondary infections have all been reported as effects of parasitism by numerous cymothoid species (Horton and Okamura, 2001; Fogelman et al., 2009; Welicky et al., 2018). For other species the effects of parasitism on the host are largely benign with hosts perhaps being able to compensate by increasing energy intake (Thatcher, 1988; Östlund-Nilsson et al., 2005; Carrassón and Cribb, 2014; Hua et al., 2017). Lack of effect has also led some researchers to suggest that some cymothoid-host relationships are commensal, where cymothoids might actually feed on the prey of their hosts (Thatcher, 1988). However, most cymothoids primarily feed on host fluids and tissues, and externally-attaching species are haematophagous specialists with mouthparts adapted for sucking blood (Nagler and Haug, 2016). It is also likely that ovigerous females do not feed at all, due to compression

of the organs while bearing developing embryos (Brusca, 1981).

## **Diversity**

Cymothoids are among the most diverse isopod families with over 400 described species in 43 genera (Boyko et al., 2019). Most of this biodiversity is concentrated in the tropics, with the largest number of species known to occur in the Western and Central Indo-Pacific regions (Smit et al. 2014; Figure 1.1). Smit et al. (2014) also highlight the unevenness of this distribution, showing that, although 41 species are known from the Atlantic, most species are from the West Atlantic and the Caribbean (perhaps, in part, because of a lack of research attention in the Eastern Atlantic). Cymothoids are mainly found on hosts from epipelagic, shelf habitats but are also found from pelagic hosts, and in the deep sea (Yamauchi, 2009; Smit et al., 2014). The range of different species is also diverse, some having restricted and others cosmopolitan distributions, and this appears unrelated to host specificity (Martin et al., 2015). The vast majority of freshwater taxa are found in South America, with at least 27 species, in nine genera, compared with four species from Africa, and two from South East Asia (Thatcher, 2006). Prevalence is generally low in natural populations (Smit et al., 2014), but with the right conditions, especially in areas of high human disturbance and in aquaculture settings, infection rates can reach between 30% – 100% (Horton and Okamura, 2001; de Lima et al., 2005; Sala-Bozano et al., 2012).

## **Attachment**

Cymothoids have gained significant notoriety because of their striking parasitic strategies — even taking the lead role in a cult horror film (*The Bay*, 2013). Certain species, the so-called ‘tongue-biters’, attach to their host’s tongue and take a blood meal, which causes the tongue to atrophy (Brusca and Gilligan, 1983). As well as the mouth, cymothoids

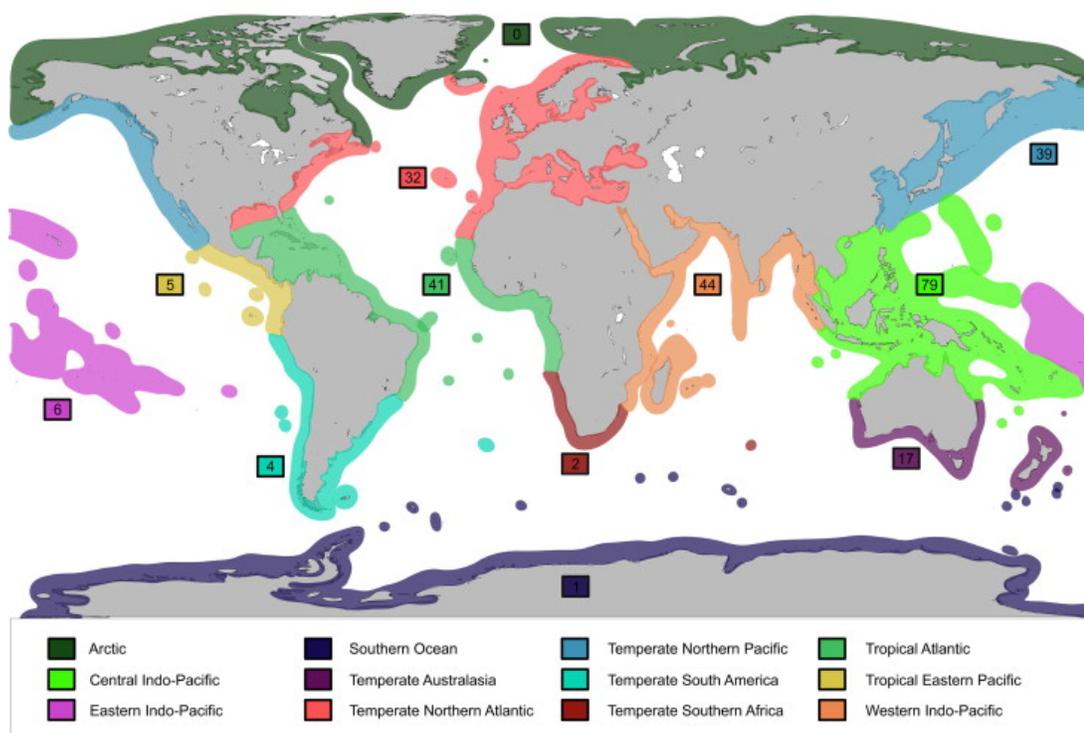


Figure 1.1: Number of marine Cymothoidae in biogeographic regions (Marine Ecoregions of the World). *Reproduced directly from Smit et al. (2014).*

either attach externally to the skin, in the gill chamber, or burrow within the body cavities of their host (Smit et al., 2014). The majority of species are highly site-specific, using only one of these attachment strategies. Attachment location is also largely conserved within cymothoid genera (Bruce, 1990; Hadfield et al., 2013; Welicky et al., 2017). Yet, within these discrete locations, there is considerable variation in microhabitat use between species with the same parasitic mode (Morton, 1974). For example, whereas *Anilocra pomacentri* attaches underneath the eye, *Anilocra haemuli* is found attached under the pectoral fin (Bunkley-Williams and Williams Jr., 1981; Adlard and Lester, 1995). Similarly, *Cymothoa exigua* and *Ceratothoa italica* attach in the mouth, to the tongue, but *Cinusa tetrodontis* and *Olencira praegustator* attach to the upper palette (Brusca and Gilligan, 1983; Trilles, 2007; Hadfield et al., 2010). Most mouth-attaching species also attach anteriorly facing out of the mouth, but *Asotana magnifica*, *Enispa convexa*

and a new genus and species we discovered in the collections at the Instituto Nacional de Pesquisas da Amazônia, Manaus, face inward toward the throat. Gill attachment is equally diverse and it is usually possible to establish where in the gill-chamber a species attaches by the distorted shape of the body. *Anphira branchialis*, for example, attaches in the uppermost corner, oriented in the same plane and direction as that of its serrasalmid host, producing an extreme dorsal vaulting (Figure 1.2 H). All isopods exhibit biphasic moulting (George, 1972) meaning cymothoids do not need to completely detach from the host during ecdysis: an innovation that may have been key to isopods evolving ectoparasitic lifestyles (Nagler and Haug, 2016).

## **Host Specificity**

Most cymothoids are highly host-specific, restricted to just one or two host species (Brusca, 1981; Bruce, 1987). Yet, some species are able to parasitise hosts from several families of fishes. *Elthusa raynaudii* is recorded from over 40 different fish species (van der Wal et al., 2019). Host-specificity does not appear to be related to parasitic strategy, with some generalists recorded among cymothoid species representing each attachment mode. However, some genera do show particular conservatism, such as the genus *Cymothoa* (Hadfield et al., 2013). Even where the ranges of potential host species of similar size and ecologies overlap, preferences have been shown in experimental (Fogelman and Grutter, 2008) and natural populations (Nagasawa, 2017). Several unusual associations have also been recorded, from Squaliformes (Williams et al., 2010), sea snakes (Saravanakumar et al., 2012), and squid (Trilles and Öktener, 2004), though these are thought to be accidental, rather than permanent.

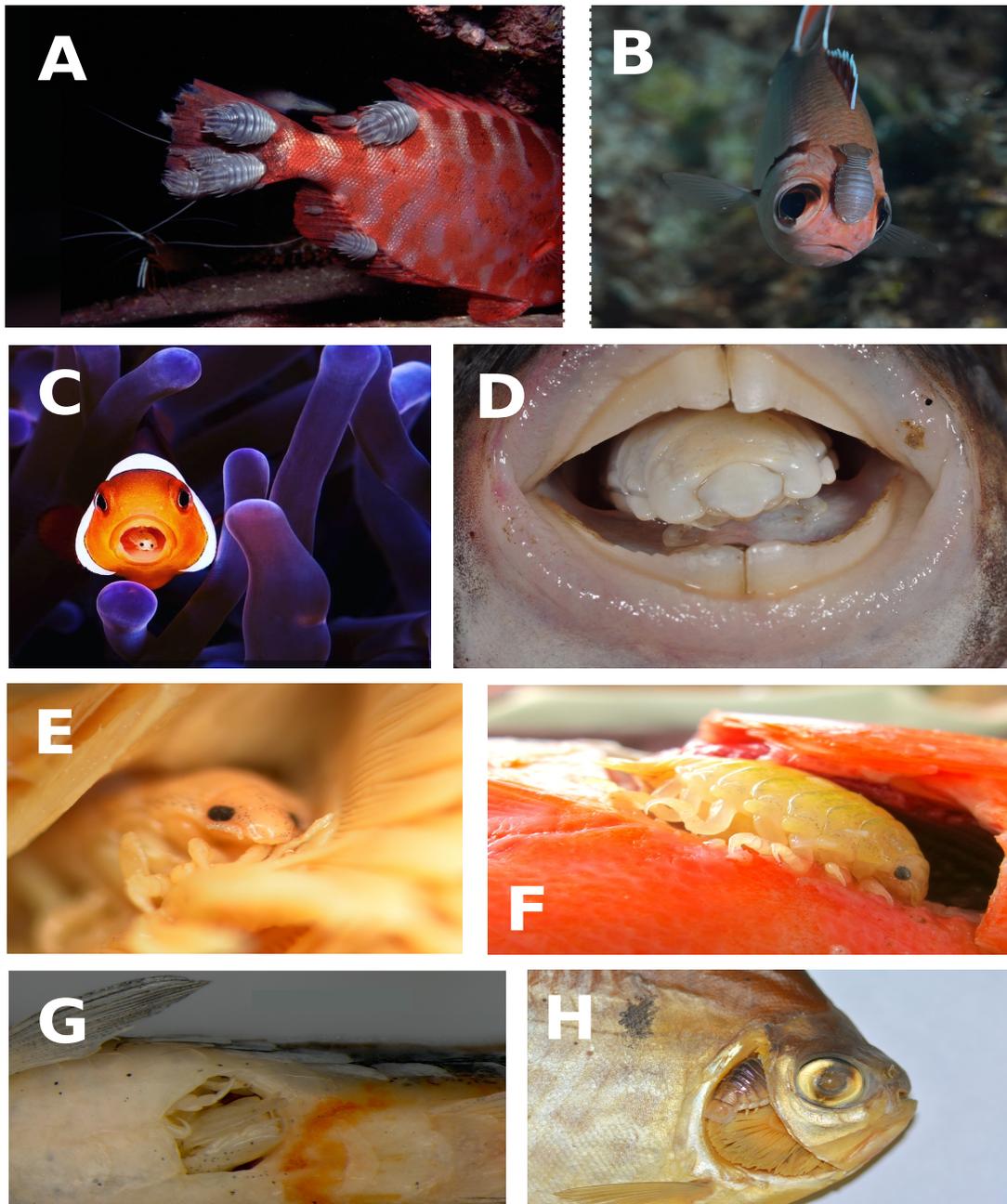


Figure 1.2: **A.** Rare example of high infestation on a single host individual by *Nerocila acuminata* (Credit: Peter Wirtz); **B.** *Anilocra myripristis* on *Myripristis jacobus* (Credit: Wolfram Sander; **C.** *Ceratothoa* sp. in the mouth of *Amphiprion polymnus* (Credit: Els van der Borre); **D.** *Cymothoa* sp. on the mouth of a pufferfish (Credit: Arthus Anker); **E.** Gen. nov. et sp. nov. (discovered in this study) facing posteriorly into the throat of it's host *Metynnus maculatus*; **F.** *Braga patagonica* attached in the gill cavity of *Pygocentrus nattereri*; **G.** *Artystone minima* burrowed within it's favoured host *Nannostomus beckfordii*; **H.** *Amphira branchialis* within the dorsal portion of the gill cavity of *Serrasalmus spilopleura*.

## Phylogenetics

The majority of cymothoid research has concentrated on alpha-taxonomy, yet many cymothoid species boundaries remain unclear and the relationships between cymothoid taxa are uncertain. Taxonomic resolution of cymothoids is particularly problematic because species are highly variable or polymorphic, and intraspecific morphological variation is easily confounded with interspecific differences (Horton, 2000). Morphology in pre-adult stages is also highly homogeneous across cymothoids, restricting species descriptions to adult females, and compared with many free-living isopods, adult cymothoids are phenotypically less complex, thus providing potentially fewer discriminatory characteristics. Furthermore, the high species diversity of cymothoids means there exists extensive variation to untangle.

To compound these challenges, many early diagnoses prioritised body shape, provided overly simplistic illustrations and incomplete descriptions, paid limited attention to within species character variation (often only examining a single specimen), and regularly lacked host data (Bruce, 1987). Fryer (1968, p. 16) succinctly outlined: "So vague are the generic criteria that have been used that the present material might equally well be assigned to any of several genera as they are currently defined." A number of major taxonomic revisions within the past 40 years have sought to rectify these issues, and modern standards have identified more consistent sets of discriminatory characters for many cymothoid taxa (Brusca, 1981; Bruce, 1986, 1987, 1990; Hadfield et al., 2013, 2014; Martin et al., 2013; Welicky et al., 2017). Molecular sequence data have also begun to be incorporated into cymothoid species descriptions to provide a more objective means by which to separate homology from homoplasy of anatomical features (Welicky et al., 2017; van der Wal et al., 2017). However, these studies have been geographically restricted, or only covered certain genera, and species definitions are still regularly based

on a collection of many small morphological differences, some of which often overlap with other species. Most importantly, without a clear evolutionary framework of cymothoid species relationships, these descriptions are inherently phenetic.

Attachment location has long influenced the classification of cymothoid taxa. Brusca (1981) framed the first evolutionary hypothesis for cymothoids, proposing that the burrowing-, skin-, and the mouth- plus gill-attaching genera are subclades which correspond to ecomorphological, adaptive lineages. Bruce (1986, 1987, 1990) also recognised subclades within Cymothoidae, but highlighted greater complexity than Brusca's (1981) model, synonymising the subfamily Livonecinae (gill-attaching) with Anilocrinae (skin-attaching), based on brood pouch, cephalon, pleopod, and pleon morphology. Bruce (1990) cautioned that apparent homologies between taxa sharing a particular attachment mode are likely to be convergent adaptations to similar ecological demands and are not necessarily reflective of shared ancestry. Bruce (1990) did not speculate as to the directionality of the evolution of attachment, but Williams Jr. and Bunkley-Williams (1994) proposed the inverse evolutionary scenario to Brusca (1981), suggesting that the ancestral cymothoid might actually have attached in the mouth and, due to overcrowding, cymothoids were forced to explore ecological opportunities outside the relatively safe confines of the buccal cavity.

Our current understanding of cymothoid phylogeny is based on three published molecular phylogenetic studies of cymothoid relationships (Ketmaier et al., 2008; Jones et al., 2008; Hata et al., 2017). In contrast to most earlier hypotheses (not based on formal phylogenetic analysis) their findings suggest that the ancestral cymothoid was likely gill-attaching, and that there have been multiple transitions between different parasitic strategies. For example, *Ceratothoa* and *Cymothoa*, two mouth-attaching genera, evolved separately, the first likely evolving from a gill-attaching ancestor, and the second from a skin-attaching ancestor (Hata et al., 2017). The most densely sampled of these studies

(Hata et al., 2017), also finds that several genera, as accepted by current taxonomic definitions, are paraphyletic. This includes the genera *Elthusa* and *Ceratothoa*, which have been the subject of recent major revisions (Hadfield et al., 2014; van der Wal et al., 2019). The monophyly of the genus *Anilocra* is also uncertain as evidenced by genetic distances between Caribbean and Eastern Atlantic species, and, if these do transpire to represent two distinct lineages, there is likely considerable cryptic diversity yet to be discovered in both (Welicky et al., 2017; Welicky and Smit, 2019). Only two freshwater species from South America, *Riggia paranensis* and an *Artystone* species, have been included in past phylogenies (Dreyer and Wägele, 2001; Hata et al., 2017), the first of which was primarily concerned with inter-familial relationships. Taken together, a linear evolutionary pathway for attachment is probably not supported, but exactly how different parasitic strategies have evolved remains an outstanding question, as does establishing the plesiomorphic parasitic strategy for the family as a whole, and establishing the timing and pattern of freshwater colonisations. Each of the three phylogenetic studies were based on two gene sequences (mitochondrial Cytochrome Oxidase subunit I and 16S ribosomal DNA), or were not well-sampled, either because they were preliminary analyses or certain groups, particularly the freshwater taxa, were not available. Taxon sampling is crucial for accurate phylogeny reconstruction (Heath et al., 2008), and for cymothoids, in particular, increased taxonomic coverage is essential if we are to answer these important questions. Furthermore, while the monophyly of Cymothoidae is well supported by morphological and molecular phylogenetic analyses (Dreyer and Wägele, 2001; Wetzer, 2002; Wilson, 2009) the position of cymothoids within Isopoda is less clear: historically all parasitic isopods were placed within the same polyphyletic group, Cymothoida.

## 1.3 Museomics

A major challenge for cymothoid systematics is sampling. Many species are still only known from one or two specimens such as in *Elthusa nierstraszi* (Hadfield et al., 2016b). As stated previously, infestation prevalence is highly irregular, and is dependent upon a number of factors including the cymothoid species, geographic location, the host species parasitised and host size, but generally, infestation rates are low in natural populations (Bakenhaster et al., 2006; Yamano et al., 2011; Roche et al., 2013; Welicky and Sikkel, 2014). Their wide biogeographic distribution also makes obtaining broad taxonomic representation from sampling in the wild extremely difficult. However, given the longevity of cymothoid research, specimens from around the world have been amassed in museums, thus concentrating representative cymothoid diversity to a few key collections. Combined with modern sequencing this affords the opportunity to retrieve genome-scale molecular data from these archival tissues — a method that has been successful for herbarium specimens (Zedane et al., 2015), birds (Hung et al., 2013), sharks (Nielsen et al., 2017), mammals (Guschanski et al., 2013), and beetles (Sproul and Maddison, 2017).

Collection material has its own challenges, including: specimen identification, missing host and locality information, and lack of detailed preservation history. Foremost, however, is the issue of DNA degradation that necessitates particular approaches for sequence retrieval and analysis originally designed for ancient DNA (aDNA) (Burrell et al., 2015). For museum specimens, the causes of DNA degradation are post-mortem processes and preservation methods, with types of damage that include single strand nicks, base modifications (deamination and depurination), oxidisation, crosslinking, and fragmentation (Zimmermann et al., 2008). DNA yields are always much lower than fresh tissues, and fragment length distributions are typically in the range of 40 – 150bp (Knapp

and Hofreiter, 2010). No relationship between specimen age and fragment length has been found for aDNA (Dabney et al., 2013b) with the implication that recently collected museum specimens are equally susceptible to damage: the time between collection and preservation, and storage conditions are critical.

Traditionally, aDNA analyses have taken advantage of the power of the Polymerase Chain Reaction (PCR) to amplify target loci. However, there are clear limitations with this approach, not least that, to obtain meaningful sequence data for alignment and comparison, only a very small percentage of molecules at the high end of the fragment length distribution can be targeted (Knapp and Hofreiter, 2010). PCR enrichment is also highly prone to contamination where relatively intact and abundant exogenous molecules are co-extracted and preferentially amplified raising issues about authenticity of recovered sequence data (Rizzi et al., 2012). The types of preserved tissue that are typically used in archaeological and some museum specimens, such as bones and teeth, can also contain higher levels of PCR inhibitors (like chitin-rich crustacean tissues) that are carried over from extraction (Rohland and Hofreiter, 2007). Finally, crosslinks and oxidative damage can also lead to amplification failure, and high nonamplification rates have been observed for museum specimens of branchiopods (Wall et al., 2014).

High-throughput sequencing has revolutionised aDNA research. The clearest advantage to these methods is that most high-throughput sequencers already rely on massively parallel sequencing of short insert sequences. Many samples can also be multiplexed in a single run, library preparation is largely the same as for modern DNA, and as little as 1ng of DNA is sufficient for some library preparations, therefore requiring lower extraction yield. Due to the potentially low number of reads representing target sequences, an enrichment step is often included in library preparation but designing capture probes can be difficult and costly for non-model organisms. Shotgun sequencing, on the other hand, can still be used to target loci for phylogenomics — particularly high copy number regions

such as mitochondrial genes and nuclear ribosomal RNAs. This method also delivers the simplest workflows and is cost-effective when the volume of information produced is factored in (Ripma et al., 2014). However, there are potential pitfalls with shotgun sequencing to be aware of, not least that contamination from exogenous DNA can be overrepresented in a sequencing library, thus lowering the depth of reads for the target organism (Burrell et al., 2015). DNA extraction, library preparation, and computational steps all need to be appropriately planned to maximise the final information content of otherwise poor-quality DNA (Dabney et al., 2013b).

## 1.4 Thesis Structure

This research had three main aims designed to investigate different aspects of cymothoid evolution and prioritised maximising the taxonomic coverage of specimens. Therefore, we naturally turned to museum collections, balancing the risks and challenges of poor specimen preservation for molecular work, with the potential increased power of obtaining high representative diversity. Following this outline introduction, which highlights aspects of cymothoid biology, ecology, and evolution, relevant to subsequent chapters, in Chapter 2 we examine the relationship between a functional trait (attachment morphology) and parasitic strategy. This chapter has been accepted for publication in *BMC Evolutionary Biology* and is presented in the form exactly as it has been accepted, only excluding details and headings particular to that journal. In Chapter 3 we investigate the relationship of cymothoids to other isopod families using whole mitogenome sequences, including those we produced to act as reference sequences for our study of cymothoid phylogeny. Estimation of cymothoid phylogeny is the principal objective of Chapter 4, but we also appraise methodological approaches to obtaining molecular sequence data from museum specimens. We have followed publication style throughout the thesis albeit with the

addition of considerable background and methodological detail in chapters three and four.

## 1.5 Aims and Objectives

### 1. Investigate relationship between the morphology of attachment appendages (dactyli) and parasitic strategy.

Objectives:

1. Quantify dactyli shape variation using geometric morphometrics.
2. Assess whether shape variation is influenced by parasitic strategy (ecology) or phylogeny.

### 2. Produce reference cymothoid mitogenomes.

Objectives:

1. Generate whole genome sequencing data for three cymothoid species using illumina short read and nanopore long read data.
2. Use these data to assemble and annotate high-quality whole mitogenome references.
3. Use publicly available short read data to produce whole mitogenome sequence for *Ligia exotica*.
4. Estimate isopod phylogeny using whole mitogenome data to assess the monophyly of major subclades above the genus level.
5. Assess the evidence for multiple transitions to terrestrialisation in the evolutionary history of isopods.

**3. Expand the scale and taxonomic coverage of sequence data for cymothoids targeting whole mitogenomes and South American freshwater taxa.**

Objectives:

1. Test a low-coverage, shotgun sequencing approach for recovery of mitogenome sequences from liquid-preserved museum specimens.
2. Assess factors influencing the recovery of DNA.
3. Quantify the metagenomic content of sequenced DNA with a focus on sources of likely contamination.
4. Estimate cymothoid phylogeny with whole mitogenome data.
5. Assess whether a linear evolutionary pathway for site-attachment is supported.
6. Assess whether colonisation of freshwater in South America occurred as a single event, or whether there were multiple transitions to this habitat.

## **Hooked on you: shape of attachment structures in cymothoid isopods reflects parasitic strategy**

## 2.1 Introduction

Permanent ectoparasites derive almost all of their energy and habitat requirements from a single host source (Bunkley-Williams and Williams Jr., 1998). Thus, traits for attachment function, which are imposed by this lifestyle, are critical for parasite survival and reproduction. These traits are often ecomorphologically significant, segregating species between different host niches (Gorb, 2008; Poulin, 2009). Morphologies shared by parasite species are likely, then, to reflect similarities in their host use. In a co-evolutionary context, attachment traits may also drive specialisation of location upon a host, partly define the limits of a parasite's host range (i.e. the breadth of host species it could infect), or enable permanent switches to different host species entirely (Araujo et al., 2015).

A remarkable example of permanent ectoparasitism is seen in the family Cymothoidae, of which all known species are obligate parasites of fishes (Hadfield et al., 2014). Individuals of the infamous *Cymothoa exigua* supplant their hosts' tongues — the only example in nature of 'anatomical replacement' by another organism (Brusca and Gilligan, 1983). Four attachment modes can be recognised within Cymothoidae: (1) mouth-attaching species (including the 'tongue-biters'), (2) gill-attachers found in the branchial chamber, (3) skin-attachers, those attached externally to the scales or skin, and (4) flesh-burrowing species that encapsulate themselves within their host's body cavities (Smit et al., 2014). Whereas gill-, mouth-, and flesh-attachers are predominantly marine species, flesh-burrowing cymothoids are typically freshwater species.

Parasitic strategy is largely conserved within cymothoid genera (Brusca, 1981; Hadfield et al., 2014), but there is substantial variation in microhabitat use between species with the same parasitic mode, because distinct locations or orientations are used by different species. For example, at the genus level, *Anilocra* and *Nerocila* comprise

exclusively skin-attaching species, but are found on the anterior and posterior regions of their hosts, respectively (Morton, 1974; Bunkley-Williams and Williams Jr., 1981). Within *Anilocra*, species are often site-specific across the anterior region, for example *A. haemuli* is always found attached near the eye, while *A. acanthuri* attaches under the mouth (Bunkley-Williams and Williams Jr., 1981). Very few species are known to use more than one attachment mode, with all such examples recorded from atypical host associations. For example, the most common Brazilian freshwater species, *Braga patagonica*, is a branchial parasite of several fish species but on cultured *Colossoma macropomum* it is regularly recorded externally attached behind the dorsal fin (Tavares-Dias et al., 2014). The evolution of parasitic mode within Cymothoidae remains unclear, but there is some consensus that each mode has evolved more than once, and that skin-attachment is unlikely to be ancestral for the group as a whole (Ketmaier et al., 2008; Hata et al., 2017).

Cymothoids are well adapted for ectoparasitism on their mobile fish hosts: a thickened cuticle affords protection from crushing forces; increased surface area of gill-bearing pleopods facilitates oxygen transfer in gill- and mouth-attaching species, and modified mouthparts enable the acquisition of blood meals in skin-attachers (Smit et al., 2014; Nagler and Haug, 2016). Externally-attaching species, relative to other parasitic modes, exhibit dorso-ventral flattening, which reduces drag and minimises the energy expenditure of their hosts (Nagler and Haug, 2016). Crypsis is also displayed by some externally-attached species as a strategy to avoid predators such as cleaner fish (Grutter, 2002). The appendages cymothoids use for attachment are particularly characteristic: each of their prehensile pereopods, 'walking' limbs, terminate in a recurved dactylus. The presence of these structures on the posterior pereopods is a synapomorphy of the obligate parasitic groups Cymothoidae and Bopyridae (Brusca, 1981). Cymothoids and their close relatives exhibit a wide spectrum of trophic dependency from free living species, through temporary to obligate parasites, and cymothoids are thought to have evolved

from either a cirolanid-like or an aegid-like ancestor (Brusca, 1981; Bruce, 1987; Smit et al., 2014). Species in the families Cirolanidae and Aegidae do not possess recurved dactyli on their posterior pereopods, and adults retain their ability to swim (Nagler et al., 2017). In contrast, cymothoids lose the ability to swim after they have infested a suitable host, which drastically reduces the probability of finding another host in the event of being displaced (Bunkley-Williams and Williams Jr., 1998). Loss of swimming appendages may have evolved in concert with the origination of cymothoid dactylus morphology as a trade-off between an increasing reliance on host resources and maintenance of traits for the acquisition of new host individuals (Nagler and Haug, 2016).

As an important trait for facilitating obligate ectoparasitism in cymothoids, we hypothesised that variation in dactylus shape would reflect differences in the functional demands of parasitising hosts in different locations. Gill- and mouth-attaching cymothoid species use their dactyli to penetrate host tissue, but also as hooks to clasp gill-rakers, tongues, or the upper palate (Hadfield et al., 2010; Martin et al., 2013; Hadfield et al., 2013, 2014). In contrast, externally-attaching cymothoids use dactyli to anchor themselves to host musculature and dermal tissues and are subject to greater hydrodynamic forces. We predicted that the externally-attaching species would have dactyli that are relatively longer, thinner, and ‘needle-like’, adapted for piercing flesh, while those of gill- and mouth-attaching species will be stouter, more recurved, and strengthened for ‘gripping’. To test these predictions we used a geometric morphometric approach to quantify dactylus shape and assessed the influence of parasitic mode, size allometry, and phylogeny on shape variation.

## 2.2 Methods

### Specimens

Cymothoid specimens used in this study are from collections at the Water Research Group, North Western University, Potchefstroom, South Africa, and the University of Salford, Manchester, UK. We took images of the first pereopod (*P1*) from 124 individuals across 18 species, and from 135 individuals of 19 species for the seventh pereopod (*P7*). Only *P1* and *P7* were measured, since these are recognised as the most useful for taxonomic studies because of considerable morphological variation between species; *P2* – 6 show much less shape variation between species (Bruce, 1990; Hadfield et al., 2013; Welicky et al., 2017). All specimens were adult females, each species was represented by at least three individuals, and there was a minimum of 26 individuals for each of three parasitic modes (Table 2.1). Flesh-burrowing specimens were not included due to insufficient sample numbers and because the seventh pereopods do not possess a recurved dactylus, only a simple stub (Thatcher, 2002). All specimens are known to occur in the ocean around southern Africa except for *Anilocra chromis* and *Anilocra physodes* which are found in the Caribbean and Mediterranean, respectively.

### Image and Landmark Acquisitions

We captured high resolution digital images of the *P1* and *P7* dactyli for each individual using a Nikon DS-Fi1 camera fitted to a Nikon SMZ1500 stereoscopic microscope. For each of *P1* and *P7*, we plotted 39 semi-landmarks to describe two curves between three fixed landmarks (Figure 2.1) with *tpsDig2* (Rohlf, 2015). The first fixed landmark was located at the medial junction with the propodus, and the second landmark was placed at the distal tip of the dactylus. Due to differences between individuals in how the

Table 2.1: Parasite specimens landmarked for this study. E = External, G = Gill, M = Mouth. GB Acc. are Genbank accession numbers for molecular sequences. Newly generated sequences are in bold typeface.

| Parasitic Mode | No. of Individuals |           | No. of Species |           |
|----------------|--------------------|-----------|----------------|-----------|
|                | <i>P1</i>          | <i>P7</i> | <i>P1</i>      | <i>P7</i> |
| External       | 26                 | 38        | 5              | 6         |
| Gill           | 37                 | 36        | 6              | 6         |
| Mouth          | 61                 | 61        | 7              | 7         |
| Total          | 124                | 135       | 18             | 19        |

| Species                      | <i>P1</i> | <i>P7</i> | Mode | GB Acc.         |
|------------------------------|-----------|-----------|------|-----------------|
| <i>Anilocra capensis</i>     | 11        | 11        | E    | <b>MK652475</b> |
| <i>Anilocra chromis</i>      | 3         | 4         | E    | KY562736        |
| <i>Anilocra physodes</i>     | 3         | 3         | E    | <b>MK652476</b> |
| <i>Ceratothoa africanae</i>  | 10        | 9         | M    | <b>MK652477</b> |
| <i>Ceratothoa carinata</i>   | 6         | 6         | M    | <b>MK652479</b> |
| <i>Ceratothoa famosa</i>     | 10        | 9         | M    | Not Available   |
| <i>Ceratothoa retusa</i>     | 7         | 10        | M    | <b>MK652478</b> |
| <i>Cinusa tetrodontis</i>    | 10        | 10        | M    | <b>MK652480</b> |
| <i>Cymothoa eremita</i>      | 8         | 8         | M    | <b>MK652481</b> |
| <i>Cymothoa sodwana</i>      | 10        | 9         | M    | <b>MK652482</b> |
| <i>Elthusa raynaudii</i>     | 9         | 7         | G    | <b>MK652487</b> |
| <i>Elthusa sp.</i>           | 3         | 4         | G    | Not Available   |
| <i>Mothocya affinis</i>      | 9         | 8         | G    | <b>MK652484</b> |
| <i>Mothocya plagulophora</i> | 3         | 3         | G    | <b>MK652483</b> |
| <i>Mothocya renardi</i>      | 7         | 7         | G    | <b>MK652485</b> |
| <i>Nerocila depressa</i>     | 4         | 5         | E    | MH425627        |
| <i>Nerocila sigani</i>       | 0         | 10        | E    | Not Available   |
| <i>Gen. nov. et sp. nov.</i> | 5         | 5         | E    | Not Available   |
| <i>Norileca indica</i>       | 6         | 7         | G    | MF628259        |

propodus overlaps and obscures the dactylus, we drew a line between these first two landmarks, and another at a  $5^\circ$  angle from this. The third fixed landmark was placed on the lateral edge of the dactylus at the intersection of the  $5^\circ$  line, thus removing joint shape information from the same relative point in each specimen. The first curve was plotted between the first and second landmarks, along the medial edge of the dactylus, re-scaled by length with 13 semi-landmarks. We used the same method for the second curve, between the second and third landmarks, but with 26 semi-landmarks as this edge is between 150-200% the length of the medial curve. Thirty images were plotted a second time to calculate landmarking error.

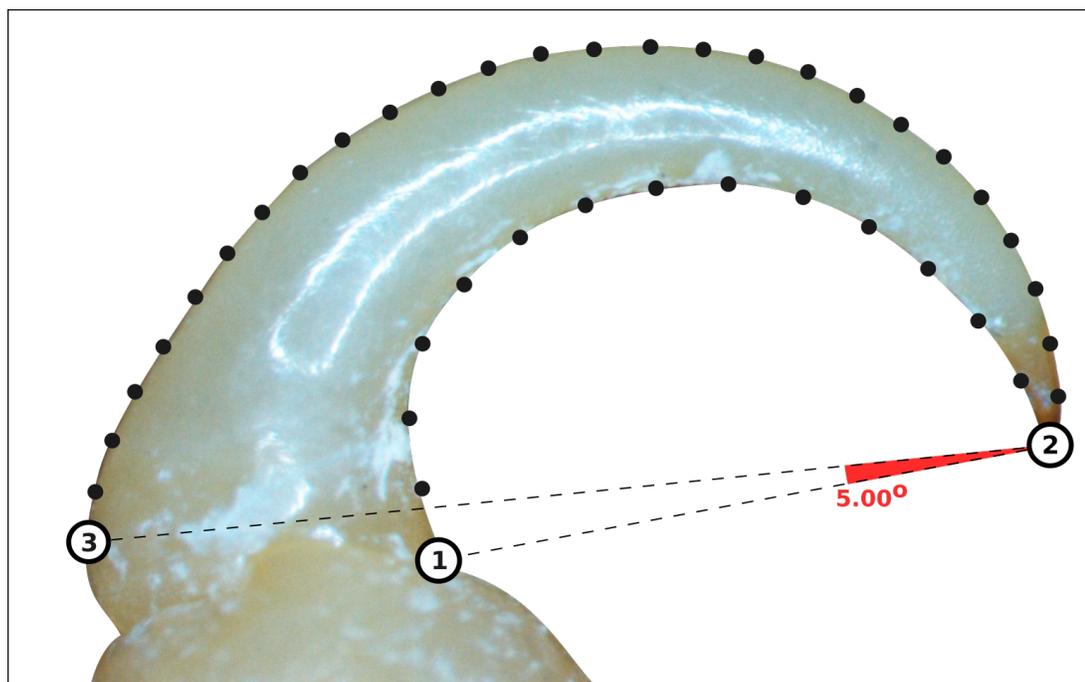


Figure 2.1: Landmark design for cymothoid dactyli. Numbered circles represent location of full landmarks and black points are semi-landmarks. Dashed lines measure a  $5^\circ$  angle between landmarks 1 and 2, from which landmark 3 was positioned.

## Phylogeny Reconstruction

We generated a new phylogenetic tree using molecular sequence data from the barcode region of mitochondrial Cytochrome Oxidase subunit I (COI: Folmer et al., 1994). A sequence was obtained from one representative of each species, for which an image was captured, except *Ceratothoa famosa*, *Elthusa sp.*, *Nerocila sigani*, Gen. nov. et sp. nov. There are no publicly available sequences for these species, and the preservation condition of our specimens did not produce DNA sufficient for PCR. For better preserved specimens, we extracted DNA from a single pereopod using a Machery-Nagel spin column kit, before amplification with Polymerase Chain Reaction (PCR) following the protocol in Welicky et al. (2017). Each PCR product was purified, then sequenced in both directions on an ABI 3630 Genetic Analyzer, and we generated consensus sequences with Geneious R10 (<https://www.geneious.com>). We added sequences for *Nerocila depressa* (MH425627) and *Anilocra chromis* (KY562736), and used the aegid, *Aega psora* (FJ581463), as an outgroup. We aligned nucleotide sequences with TranslatorX (Abascal et al., 2010) and used Gblocks v0.91b (Castresana, 2002) to remove ambiguously aligned sites resulting in a trimmed alignment of 581bp.

Due to the length of alignments and our concern that phylogenetic signal might be limited, we applied a backbone constraint to our tree searches based on the latest available molecular phylogeny for cymothoids (Hata et al., 2017). This allowed us to better place taxa for which we have no phylogenetic information and to calculate branch lengths. Specifically, these included hard constraints on genera represented by more than two species, while restricting the potential placement of other taxa within those clades. The best fitting substitution model for the alignment (GTR +  $\Gamma$ ) was selected using corrected Akaike's Information Criterion (AICc) in jModelTest 2.1.10 (Darriba et al., 2012; Guindon and Gascuel, 2003). Tree searches were performed under Maximum Likelihood (ML)

using RAxML v8 (Stamatakis, 2014), and topological support was assessed with 500 rapid bootstrap replicates. The topology of the constrained tree was then compared against the unconstrained topology using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999). We ultrametricised the tree using the function ‘chronos’ of the *ape* package in R under a correlated trait model, and an optimised value for the smoothing parameter (from 5,000 starting values) was selected as that which produced the tree with the highest penalised log-likelihood (Sanderson, 2002).

## Statistical Analyses

All subsequent analyses used R v3.4.4 (R Core Team, 2018) and the packages *ape* 5.2 (Paradis and Schliep, 2018), *geomorph* 3.0.7 (Adams and Otárola-Castillo, 2013), *phytools* 0.6 (Revell, 2012), *nlme* 3.1 (Pinheiro et al., 2018), and *stats* (R Core Team, 2018). We obtained individual shape variables from our raw landmark coordinates by Generalised Procrustes Analysis (GPA; (Rohlf and Slice, 1990)), using Procrustes distance to optimise locations of semi-landmarks (Bookstein, 1997). This superimposition produces a set of scaled and aligned Procrustes coordinates that minimises location, orientation, and size differences between samples, thus retaining only information related to geometric shape. The mean consensus shapes for each species in both the *P1* and *P7* configurations were aligned in a second GPA to produce species level Procrustes coordinates (Olsen, 2017). We then performed Principal Components Analysis (PCA) on the aligned coordinates to visualise shape differences plotted as morphospaces and PCA backtransformations (Olsen, 2017). The PCA scores were also used as shape variables in regression analyses, for which we retained the first twelve principal components with non-zero eigenvalues. Together these explained over 99% of variation in both our *P1* and *P7* datasets. For phylogenetic analyses we generated sets of shape variables for *P1* and *P7* with the same method including only the species present on our phylogeny,

and using phylogenetic PCA (pPCA) –  $P1_{phy}$  and  $P7_{phy}$  datasets.

At this point, we quality-checked our full  $P1$  and  $P7$  datasets to identify landmark or analytical problems. First, for each dataset, we calculated the mean Procrustes distance of each sample to the  $P1$  or  $P7$  global consensus shape, where outlying data points might indicate landmark error. Outliers were retained if they were consistent within a species. For example, in the  $P1$  dataset individuals of *Norileca indica* all appear above the upper quantile, which reflects genuine shape information, rather than error. Error in our data acquisition steps was assessed by nested ANOVA between our 30 repeated landmark sets to calculate the ratio between total Mean Squared Error (MSE) and that contributed by our replicates. We found our digitisation to be over 97% repeatable, and, therefore, that measurement error did not significantly influence variation in shape.

For both our  $P1$  and  $P7$  datasets, we first fitted a ‘full’ model including all covariates and their interactions, with single factor non-parametric Multivariate Analysis of Covariance (Procrustes npMANCOVA). The twelve principal components describing shape formed our response and as predictors we modelled size as a continuous covariate, parasitic mode as a three-level factor, and the interaction between size and mode. For size, we used the mean of log-transformed centroid sizes for each species, which are the square root of the sum of squared distances between the landmarks of each specimen and their centroid (Klingenberg, 2016). We then proceeded with pairwise tests to assess homogeneity of slopes (whether patterns of shape-allometry are common across parasitic modes), and group means (testing shape differences between groups after accounting for variation in size). Finally, for both datasets we also used the residuals from a regression of shape on size (equivalent to allometry-free shapes) to test shape differences between parasitic modes. This latter analysis is only appropriate where there is a common allometry between groups (Outomuro and Johansson, 2017). For each analysis we used 10,000 permutations of the Residual Randomisation Permutation Procedure (RRPP:

Collyer and Adams 2018) to generate empirical sampling distributions for significance testing, and from which effect sizes were estimated as standard deviates (z-scores).

We applied the same modelling approach to our  $P1_{phy}$  and  $P7_{phy}$  datasets but using Phylogenetic Generalised Least Squares (PGLS) rather than Ordinary Least Squares regression (OLS). Data collated from related species violate the assumption of OLS that residual error is independent between observations (Revell, 2010). It is possible to account for this autocorrelation by weighting the error structure of the regression assuming a model of trait evolution to calculate covariance of traits among species (Poulin et al., 2011; Symonds and Blomberg, 2014). The default method for PGLS in `geomorph` (function 'procD.pgls') assumes traits evolve under Brownian motion, which may not be realistic for all datasets. Therefore, we estimated Pagel's lambda as a measure of phylogenetic signal in the residuals of our 'full' models using Restricted Maximum Likelihood (REML) implemented in the function 'gls'. Using these estimated values of lambda ( $P1_{phy} = 0.18$  and  $P7_{phy} = 0.13$ ) we then calculated phylogenetic variance-covariance correction matrices and conducted PGLS with 'procD.lm'. Calculating separate variance-covariance matrices in this way allowed us to complete all regression analyses with `geomorph` functions and the RRPP method, while using a better fitting model. To infer patterns of morphospace occupation across the evolution of parasitic modes we projected our phylogeny into morphospaces to create phylomorphospaces for  $P1_{phy}$  and  $P7_{phy}$ .

## 2.3 Results

### Dactylus Shape Variation

Principal Component Analysis (PCA) clearly separates dactylus shape of externally-attached cymothoids from gill- and mouth-attaching species for both  $P1$  and  $P7$ , where they occupy a distinct region of morphospace (Figure 2.2). The vast majority of shape

variation is accounted for by the first two principal components: 77.3% and 79.3% for  $P1$  and  $P1_{phy}$ , 88.5% and 89.7% for  $P7$  and  $P7_{phy}$ . For  $P1$ , higher Principal Component 1 (PC1) values indicate increased curvature, while Principal Component 2 (PC2) reflects changes in the width of the dactylus, especially the proximal width. Interestingly, all externally-attaching species have positive PC1 values for  $P1$ , and gill-attaching species share similar curvatures with a narrow PC1 range.  $P7$  dactyli are thinner with margins subparallel at lower values of PC1, and are flatter for negative values of PC2 (Figure 2.2). Similar to  $P1$ , gill-attaching species occupy a small area of  $P7$  morphospace, while mouth-attaching species cover the widest area, indicating a broad range of dactylus shapes.

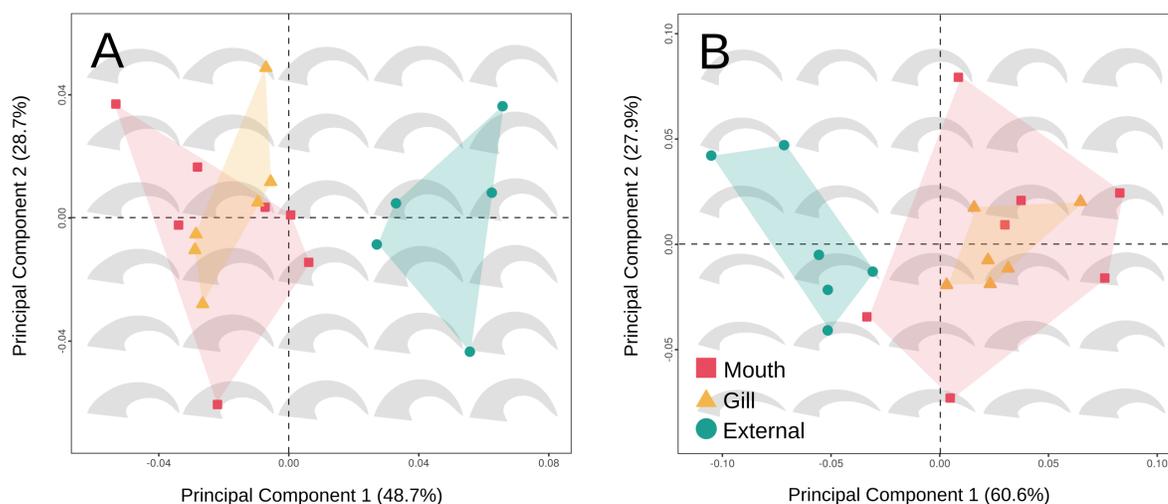


Figure 2.2: Morphospace plot of  $P1$  (A) and  $P7$  (B) dactyli. The first Principal Component is plotted on the x-axis and the second on the y-axis. Percentages included in the axis labels are the variation accounted for by each principal component. Convex hulls are calculated for each parasitic mode. Silhouettes are PCA backtransformations that depict shape across morphospaces.

Table 2.2: Results from OLS regressions of the full *P1* and *P7* datasets. Group Slopes are pairwise comparisons between the allometric vectors of parasitic mode, Group Means are pairwise comparisons of Procrustes distance between modes after accounting for variation in size, and Allometry-free Group Means are the same comparisons as Group Means but using the residuals from regression of shape on size as shape variables. For pairwise comparisons, above the diagonal are p-values, and below the diagonal are z-scores. Below each pairwise comparison table are the values for the model (Type I Sum of Squares). E = External, G = Gill, M = Mouth.

| Model terms                           | P1                                      |         |         | P7                                       |         |         |
|---------------------------------------|---|---------|---------|--|---------|---------|
|                                       | df                                      | z-score | p-value | df                                       | z-score | p-value |
| Shape ~ Size                          | 1                                       | 1.92    | <0.01*  | 1  | 0.81    | 0.23    |
| Shape ~ Mode                          | 2                                       | 2.98    | <0.01*  | 2  | 2.82    | <0.01*  |
| Shape ~ Size * Mode                   | 2                                       | 1.09    | 0.86    | 2  | 1.66    | 0.03*   |
| Pairwise (Group Slopes)               |   |         |         |  |         |         |
|                                       | E                                       | M       | G       | E  | M       | G       |
| E                                     | –                                       | 0.62    | 0.35    | –  | 0.01*   | 0.59    |
| M                                     | 0.54                                    | –       | 0.63    | 1.85                                     | –       | 0.28    |
| G                                     | 0.27                                    | 0.53    | –       | 0.34                                     | 0.75    | –       |
|                                       | z-score = 1.07, Res.df = 12<br>p = 0.86 |         |         | z-score = 1.53, Res.df = 13<br>p = 0.05* |         |         |
| Pairwise (Group Means)                |   |         |         |  |         |         |
|                                       | E                                       | M       | G       | E  | M       | G       |
| E                                     | –                                       | <0.01*  | <0.01*  | –  | <0.01*  | 0.01*   |
| M                                     | 3.78                                    | –       | 0.37    | 3.94                                     | –       | 0.94    |
| G                                     | 3.31                                    | 0.31    | –       | 2.62                                     | 1.33    | –       |
|                                       | z-score = 3.08, Res.df = 14<br>p <0.01* |         |         | z-score = 2.69, Res.df = 15<br>p <0.01*  |         |         |
| Pairwise (Allometry-free Group Means) |   |         |         |  |         |         |
|                                       | E                                       | M       | G       | E  | M       | G       |
| E                                     | –                                       | <0.01*  | <0.01*  | –  | <0.01*  | <0.01*  |
| M                                     | 3.79                                    | –       | 0.37    | 3.82                                     | –       | 0.76    |
| G                                     | 3.32                                    | 0.31    | –       | 3.41                                     | 0.81    | –       |
|                                       | z-score = 3.10, Res.df = 14<br>p <0.01* |         |         | z-score = 2.82, Res.df = 16<br>p <0.01*  |         |         |

## Relationship Between Dactylus Shape, Size, and Parasitic Mode

Dactylus shape is significantly and strongly correlated with parasitic mode, as evidenced by the positioning of species in morphospace and from the results of Procrustes npMANCOVA (Table 2.2). The interaction term between size and mode has a weak and non-significant effect for  $P1$  (z-score = 1.09, p-value = 0.86), whereas both size and mode exhibit large and significant effects (z-score = 1.92 and 2.98, respectively). For  $P7$  we find no significant effect of size on shape (z-score = 0.81, p-value = 0.23), but mode (z-score = 2.82, p-value < 0.001) and the interaction of size and mode (z-score = 1.66, p-value = 0.03) are both significant. Despite the significant interaction term in  $P7$  it accounts for little of the total variance in the model ( $R^2 = 0.09$ ). Pairwise comparisons of angles between group allometric slopes show that the size-mode interaction in  $P7$  is driven entirely by differences between external- and mouth-attaching species (z-score = 1.85, p-value = 0.01). As expected from results of the full model,  $P1$  comparisons of group slopes show no pairwise differences. For both appendages there are significant differences between the mean shapes of externally-attaching cymothoids and the other two modes, but not between gill- and mouth-attaching species.

## Phylogenetic Context

We did not find any significant differences between the log-likelihoods of our constrained and unconstrained topologies using Shimodaira-Hasegawa test (diff = -0.03,  $p = 1$ ) (Figure 2.3A). As expected, the constrained topology is consistent with that of Hata et al. (2017) except we do not recover a sister relationship between *Cymothoa* and *Nerocila*. Evidence from previous work continues to show that evolution of attachment mode in cymothoids is homoplastic, with gill and external attachment likely to have arisen independently at least twice or for there to have been secondary reversals (Ketmaier

Table 2.3: Results from PGLS regressions of the  $P1_{phy}$  and  $P7_{phy}$  datasets. Group Slopes tables are pairwise comparisons between the allometric vectors of parasitic mode and Group Means tables are pairwise comparisons of Procrustes distance between modes after accounting for variation in size. For pairwise comparisons, above the diagonal are p-values, and below the diagonal are z-scores. Below each pairwise comparison table are the values for the model (Type I Sum of Squares). E = External, G = Gill, M = Mouth.

| Model terms             | P1                                       |         |         | P7                                       |         |         |
|-------------------------|--|---------|---------|--|---------|---------|
|                         | df                                       | z-score | p-value | df                                       | z-score | p-value |
| Shape ~Size             | 1  | 1.96    | 0.01*   | 1  | 0.42    | 0.36    |
| Shape ~Mode             | 2  | 2.02    | 0.01*   | 2  | 1.70    | 0.04*   |
| Shape ~Size * Mode      | 2  | 1.66    | 0.96    | 2  | 1.25    | 0.10    |
| Pairwise (Group Slopes) |  |         |         |  |         |         |
|                         | E  | M       | G       | E  | M       | G       |
| E                       | –  | 0.82    | 0.44    | –  | 0.01*   | 0.57    |
| M                       | 0.96                                     | –       | 0.43    | 1.78                                     | –       | 0.24    |
| G                       | 0.10                                     | 0.01*   | –       | 0.27                                     | 0.91    | –       |
|                         | z-score = 1.71, res.df = 9<br>p = 0.96   |         |         | z-score = 1.22, res.df = 9<br>p = 0.11   |         |         |
| Pairwise (Group Means)  |  |         |         |  |         |         |
|                         | E  | M       | G       | E  | M       | G       |
| E                       | –  | 0.01*   | 0.02*   | –  | 0.01*   | 0.04*   |
| M                       | 2.64                                     | –       | 0.36    | 2.77                                     | –       | 0.97    |
| G                       | 2.29                                     | 0.31    | –       | 1.89                                     | 1.49    | –       |
|                         | z-score = 2.31, res.df = 11<br>p = 0.01* |         |         | z-score = 1.61, res.df = 15<br>p = 0.04* |         |         |

et al., 2008; Hata et al., 2017). The first two components of pPCA for  $P1_{phy}$  account for 48.4% and 30.8% of variation in shape respectively, while for  $P7_{phy}$ , pPC1 reflects 47.7% and pPC2 42.2%. Mouth-attaching genera occupy a much greater portion of phylomorphospace than other modes and exhibit similarly striking patterns as our full datasets: *Cymothoa eremita*, for example, exhibits a  $P7$  morphology most similar to external species than to the congener *C. sodwana* (Figure 2.3C). Gill-attaching species have highly similar  $P7$  morphologies but *Norileca indica* has a  $P1$  morphology distinct

from the other gill-attaching species, perhaps due to its unusual attachment orientation on the ventral side of the operculum, facing anteriorly (van der Wal et al., 2017). Most gill-attaching species are found anchored to the gill filaments with their dorsal side facing the operculum. Accounting for phylogeny with PGLS results in large and significant shape differences between parasitic modes for both  $P1_{phy}$  and  $P7_{phy}$  (Table 2.3). There is no evidence of separate allometric slopes between groups for either dataset using PGLS but as with  $P1$ , allometry remains an important influence on shape for  $P1_{phy}$ . We find the same pattern in pairwise comparisons as with OLS regression from our full datasets, where the mean shapes of gill- and mouth-attaching species are significantly different from externally-attaching species, but not from each other.

## 2.4 Discussion

Our results indicate that the shape of cymothoid dactyli are strongly influenced by parasitic mode, corroborating theory that adaptation to hosts and local environments is a driver of parasite phenotypes (Kaltz and Shykoff, 1998). Other empirical studies also link the morphology of parasite attachment to various ecological factors: in relation to host size in feather lice (Bush et al., 2006); host specialisation in platyhelminth fish parasites (Mandeng et al., 2015); host thermal regulation strategy in acanthocephalans (Poulin, 2007); host biogeographic plasticity (Vignon and Sasal, 2010); and, as we also find here, microhabitat (Llewellyn, 1956; Aznar et al., 2018). However, phylogeny may also explain morphological variation where more closely related species are expected to share similar traits. For example, Vignon et al. (2011) and Rodríguez-González et al. (2017) both suggest that integration between parts of monogenean attachment organs results in phylogenetic constraint of shape variation. Both of these studies find that divergent parasite species that infect the same host species possess differently shaped attachment

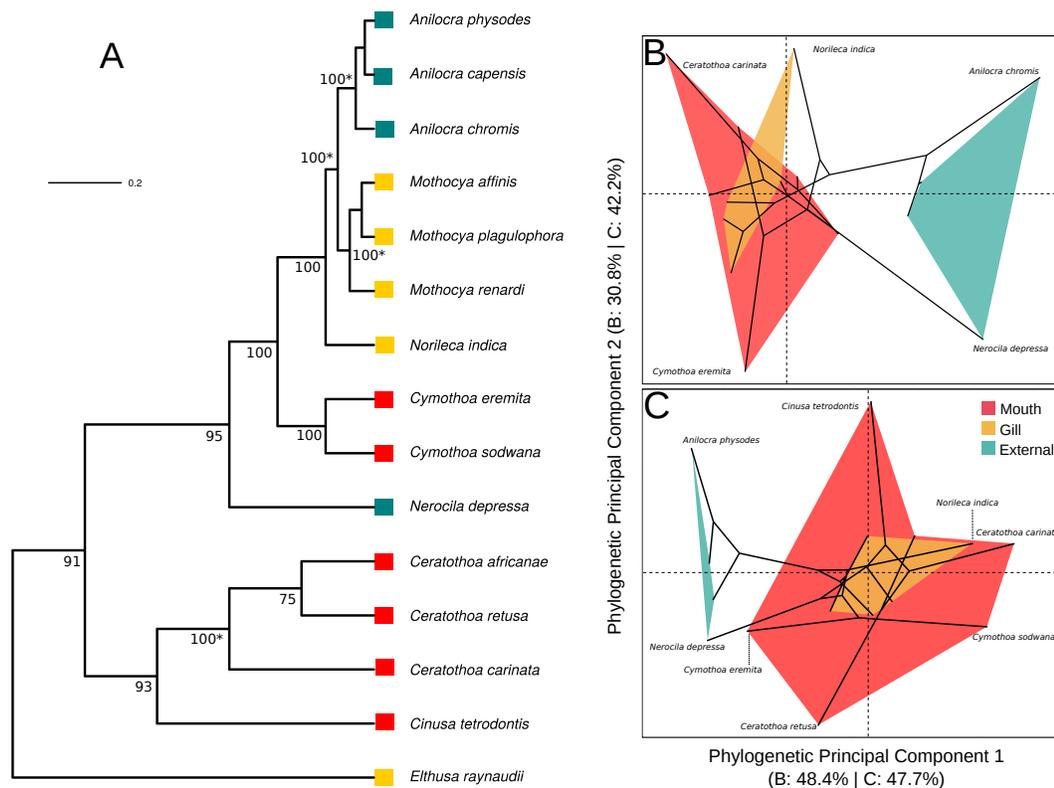


Figure 2.3: Maximum-likelihood phylogeny (**A**), and phylomorphospace plots of  $P1_{phy}$  (**B**) and  $P7_{phy}$  (**C**) dactyli. Node numbers are bootstrap support values and asterisks are constrained nodes. For phylomorphospaces the first Principal Component is plotted on the x-axis and the second on the y-axis. The percentages included in the axis labels are the variation accounted for by each Principal Component. Convex hulls are plotted for each parasitic mode.

organs (called haptors). We have found no evidence that there are clade-specific patterns of association between parasite mode and dactylus shape. Our results imply that different clades can converge on the same dactylus shape, which presumably is well-adapted for that particular parasite mode.

Parasitic mode consistently shows large effect sizes in both OLS and PGLS analyses of  $P1$  and  $P7$  datasets. Underlying this pattern are significant shape differences between externally-attaching species, and both gill- and mouth-attachers. Interestingly, however, the shape dissimilarities of external and internal species in morphospace are also different

between *P1* and *P7* dactyli. We also find that the shape of *P1* dactyli is influenced by size, but this is not the case for *P7*. Altogether, this suggests that anterior and posterior dactyli may function differently, and in externally-attaching species the differences are particularly acute. Certainly, pereopods are broadly arranged in two opposing angles between pereopods 1–3 and 4–7, which is suggested to enhance attachment ability (Brusca, 1981; Nagler and Haug, 2016). *P1* dactylus shapes are highly recurved, whereas for *P7* the shape is flatter and more slender than those of gill- and mouth-attaching species. Gill- and mouth-attaching species, on the other hand, possess more similar *P1* and *P7* morphologies. All isopods exhibit biphasic moulting (George, 1972); a particularly useful trait for cymothoids meaning they do not need to completely detach from the host during ecdysis. It has previously been suggested that this preadaptation may have been key to isopods evolving ectoparasitic lifestyles (Nagler and Haug, 2016). The risk of detachment during ecdysis is presumably far greater for external-attaching species and posterior dactyli could provide secure anchorage if they function to pierce host tissue, as *P7* morphology suggests they might.

Inhabiting the external environment, at least for the few external taxa sampled, seems to result in convergence of *P7* dactylus morphology, where relatively less related species share similar morphologies. External-attaching species have a narrow range of PC1 scores, and *Anilocra capensis* shares a more similar morphology with *A. chromis* and *Nerocila depressa* than its closest relative in our phylogeny, *A. physodes*. Similarly, gill-attaching species share very similar *P7* morphologies. Welicky et al. (2019) and Pawluk et al. (2015) have shown that cymothoid and host size are tightly correlated except for gill-attaching species. This suggests that available space in the opercular cavity does not scale linearly with host size much in the same way that, for eyes, the upper and lower tarsi are near flush to the sclera or cornea regardless of size. Such a restriction could limit growth of gill-attaching species, including dactylus shapes. Morphology of

gill-attaching species appears to neatly fit their location and orientation within the gill cavity, where they are typically asymmetrical, twisted to one side, and either completely flat or extremely concave. Therefore, there could be two constraints operating on *P7* morphology: exposure and detachment risk related to external parasitism, and for gill-attaching species the volume of the opercular cavity.

We observe the opposite pattern in mouth-attaching species, which exhibit highly divergent *P7* forms. Free from hydrodynamic forces and the risks of being brushed or cleaned off, a fish mouth is, perhaps counter-intuitively, a safe environment in which to reside as an ectoparasite. Mouth-attaching species have adapted to 'bite tongues' but they also secure themselves to the upper palate, as in *Cinusa tetradontis* (Hadfield et al., 2010), wrap their dactyli around the tongue, or attach to the inner cheek and gill arches. Most mouth-attaching species are oriented anteriorly, but a few species face posteriorly into the throat, like all species in genus *Isonobula* (Taberner, 1998). Therefore, mouth-attachment as a discrete parasitic mode actually conceals considerable microhabitat variation that could explain the diversity of *P7* dactylus shapes. In addition, host specificity and the manner of parasite speciation might also be important. Except for *Cymothoa eremita*, each mouth-attaching species in our dataset is known from fewer than three host species, and even *C. eremita* shows strong preference toward *Parastromateus niger* (Hadfield et al., 2013, 2014). Host-switching is a common model explaining parasite speciation, but only a subset of potential new hosts may be a suitable match for parasite phenotypes and it is likely that new host environments are initially suboptimal (Engelstädter and Fortuna, 2019). After a host-switch event, subsequent co-adaptation between the parasite and the new (suboptimal) host environment would then refine morphological features, potentially driving shape differences between parasite species.

Another possible determinant of cymothoid attachment morphology is feeding ecology. It has previously been suggested that while externally-attaching species depredate hosts,

feeding on blood and tissue, gill- and mouth-attaching species may additionally take up energy from prey their hosts consume (Maxwell, 1982; Thatcher, 1988; Adlard and Lester, 1995; Roche et al., 2013). Certainly, the relatively benign effects of many gill- and mouth-attaching species could support this view (Marks et al., 1996; Carrassón and Cribb, 2014). Differences in feeding ecology are also consistent with mouthpart morphology, where gill- and mouth-attaching cymothoids possess molar processes and incisors adapted to grind or slice tissues. Externally-attaching species, however, have mouthparts that interlink to form a functional sucking cone, adapted for fluid intake (Nagler and Haug, 2016). In externally-attaching species, the incisor process is narrowed to a point rather than a blade, perhaps for piercing host flesh, while the mouthparts are angled downward toward the host. Furthermore, the primary mouthparts are restricted in their movement, only able to motion in an inward-outward direction. Such a movement requires a suitable counter-force to hold the mouth close to the host (Nagler and Haug, 2016); a role that might be performed by the anterior pereopods and, in part, determine the *P1* shape of externally-attaching species.

Size is also a significant predictor of *P1* shape in OLS and PGLS analyses. As stated previously, the external environment must impose unique challenges, not least of which is hydrodynamic force. We suggest that the directionality of this force could be key to understanding the different influences of size between *P1* and *P7* dactylus shapes. All externally-attaching species orientate anteriorly, parallel with their hosts (Bunkley-Williams and Williams Jr., 1981; Bruce, 1987). In our dataset, they also possess the largest and most recurved *P1* dactyli (mean centroid size = 7.54). The anterior dactyli in cymothoids are angled in a manner ideal for pulling against the direction of flow, which requires less effort than pushing due to reduction of frictional force. Such efficiency is likely complemented in externally-attaching species by increased strength derived from *P1* dactylus shape and size. Drawing parallels from structural engineering, we

can conceive that steeper arches with more symmetrical parabolic geometries are able to resist greater forces in compression (Dym and Williams, 2010). Gill- and mouth-attaching species, on the other hand, are not subject to the same demands of the external environment, and possess *P1* dactyli that are not as recurved and are perhaps better suited to withstand other structural displacements like shear.

## 2.5 Conclusions

We used geometric morphometrics, multivariate analyses, and phylogenetic comparative methods to quantify shape variation in the attachment structures of cymothoid isopods to determine whether shape differences between species correspond to parasite mode. In addition, we assessed the relative influences of allometry and shared ancestry. We found that ecology is the primary driver of dactylus shape. Separate lineages appear to have independently evolved similar dactylus morphologies that are presumably optimal for particular parasitic modes. The clearest differences are between externally-attaching species, and those attached internally in the gill or mouth. Geometric morphometrics is a powerful method for uncovering complex patterns from simple outline shapes like dactyli. Of particular note are the shape differences we found between anterior and posterior pereopod shapes. These anterior-posterior differences are also characteristic of the different parasitic modes and likely reflect adaptations to obligate parasitism and differences in feeding ecology between externally-attaching, and gill- and mouth-attaching cymothoids.

# **Mitochondrial genomes and isopod phylogeny**

### 3.1 Introduction

Mitochondrial DNA sequences are among the most widely used data for molecular systematics across a diversity of taxonomic groups. Mitochondria are attractive markers for phylogenetic analyses due to their high copy number, low levels of recombination, uniparental inheritance, and high mutation rates. For many taxa, mitogenomes are also well characterised, which increases confidence in orthology assessment. Today, sequencing technology has made whole mitogenomes widely accessible and their recovery is now routine even for non-model organisms, and from degraded specimens (Hahn et al., 2013; Gillett et al., 2014; Dodsworth, 2015; Crampton-Platt et al., 2016; Grandjean et al., 2017). Whole mitogenomes are expected to provide better phylogenetic estimates than single mitochondrial genes because, despite physical linkage, individual loci often produce incongruent topologies due to differential selective pressures (Duchêne et al., 2011). Yet, despite the many positive features of whole mitogenomes for phylogenetics, accelerated evolutionary rates — augmented by phenomena like gene rearrangements (Xu et al., 2006) — and functional conservatism can result in compositional skews and substitution saturation, which increases homoplasy, limits character variation, and erodes phylogenetic signal (Wetzer, 2002; Talavera and Vila, 2011; Timmermans et al., 2015). Therefore, phylogenetic inferences based on these data are more susceptible, than nuclear genes, to systematic errors such as long-branch attraction (LBA: Sanderson et al., 2000), particularly when taxa are sparsely sampled (Tarrío et al., 2001; Rosenberg and Kumar, 2003; Foster, 2004; Bergsten, 2005). Previous research across a breadth of taxonomic groups has, therefore, suggested that mitogenome data may only be suitable for accurate resolution of shallower divergences below family level (Curole and Kocher, 1999; Zardoya and Meyer, 2001; Talavera and Vila, 2011; Morgan et al., 2014). Another possible limitation of mitogenome phylogenetics is that in the presence of introgression

and incomplete lineage sorting, an accurately resolved phylogeny may not reflect the species tree, only the evolutionary history of the mitochondria (Moore, 1995; Sota and Vogler, 2001; Cenzi de Ré et al., 2010; Duchêne et al., 2011).

Isopoda is a species-rich, morphologically, and ecologically diverse crustacean order that includes over 10,300 described species (Poore and Bruce, 2012). Isopods are found in most of the world's environments, having successfully colonised terrestrial, freshwater, and subterranean habitats from the ocean in which they still exhibit the greatest species richness (Wilson, 2007; Poore and Bruce, 2012; Sfendourakis and Taiti, 2015). In many of these environments isopods are integral to ecosystem function, most notably as nutrient-cyclers, but they also fulfil ecological roles from scavengers to obligate parasites (Poore and Bruce, 2012). Isopods exhibit gonochoristic and asexual reproduction (*Trichoniscus pusillus* is parthenogenetic), and sequential hermaphroditism (in cymothoids and anthurideans), making them an obvious target for studies of sexual determination systems (Fussey, 1984; Smit et al., 2014). Some species, such as *Limnoria lignorum*, also produce novel proteins for breaking down cellulose and plant material, with potential industrial applications in the synthesis of biofuels (Kern et al., 2013). Additionally, isopods are found to possess peculiar genomes with unusual architectures, and a propensity for heteroplasmy and large-scale gene rearrangements (Doublet et al., 2012; Kilpert et al., 2012; Peccoud et al., 2017). Altogether, isopods are a potentially informative model for a broad range of biodiversity, genomic, and evolutionary research provided their phylogenetic relationships can be robustly resolved.

Isopods are currently divided into eleven suborders: Asellota, Calabozoidea, Cymothoidea, Limnoriidea, Microcerberidea, Oniscidea, Phoratoidea, Phreatoicoidea, Sphaeromatidea, Tainisopidea, and Valvifera (Horton et al., 2019). Yet, there is significant uncertainty in isopod higher relationships due to major topological conflicts between phylogenetic analyses of morphological characters, nuclear and mitochondrial genes,

whole mitogenomes, and combinations of these data types (Brusca and Wilson, 1991; Wetzer, 2002; Brandt and Poore, 2003; Spears et al., 2005; Wilson, 2009; Kilpert et al., 2012; Poore and Bruce, 2012; Hata et al., 2017; Lins et al., 2017; Yu et al., 2018; Hua et al., 2018). The first subordinal studies, based on morphological characters, variously place Asellota, Phreatoicidea, Asellota+Phreatoicidea or Oniscidea (the group containing all terrestrial isopods) sister to other isopod taxa (Wägele, 1989; Brusca and Wilson, 1991; Brandt and Poore, 2003; Schmidt, 2008). This suggests, at least, that the ancestral isopod condition was 'short-tailed' (characterised by short pleotelson and sedentary lifestyles): a hypothesis supported by the earliest known isopod fossil (*Hesslerella shermanii*) from the Upper Carboniferous, which shares a similar morphology to modern Phreatoicidea (Schram, 1970).

Molecular phylogenies, however, consistently contrast with morphological evidence, with only a few studies supporting the 'short-tailed first hypothesis' (Dreyer and Wägele, 2001; Wetzer, 2002; Wilson, 2009; Shen et al., 2017; Yu et al., 2018). Limnoriidea has been found to occupy the sister position using a mito-nuclear dataset and node-dated fossil calibrations (Lins et al., 2017). Cymothoidea, the taxon containing most parasitic species and widely accepted as exhibiting derived characteristics, has also been found to comprise one of the deepest branches within Isopoda (Wilson, 2009; Lins et al., 2012; Hua et al., 2018; Zou et al., 2018). Far from alleviating disagreements surrounding early divergences within the isopod crown group, whole mitogenomes have introduced additional uncertainty into higher isopod phylogeny. To date, published studies have found that Cymothoidea and Oniscidea are never monophyletic when whole mitogenome data are used to estimate phylogeny, and several taxa (e.g. *Eurydice pulchra*, *Gyge ovalis*, *Ligia oceanica*, and *Limnoria quadripunctata*) are found to fluctuate position across the tree or are pulled toward the outgroup (Kilpert et al., 2012; Lins et al., 2017; Shen et al., 2017; Hua et al., 2018; Zou et al., 2018). These results lead to conclusions that

are likely anomalous, such as a polyphyletic Oniscidea or a sister Cymothoidea (Kilpert et al., 2012; Lins et al., 2017; Shen et al., 2017; Hua et al., 2018; Zou et al., 2018). These hypotheses have obvious implications for our understanding of isopod evolution and patterns of habitat colonisation, including terrestrialsation — which, according to mitogenome phylogenies, has occurred twice, contradicting morphological analyses (Wilson, 2009).

Despite previous authors highlighting the limitations of both mitochondrial genes (Wetzer, 2002) and nuclear ribosomal genes (Spears et al., 2005; Wilson, 2009) for accurate resolution of ordinal and subordinal isopod relationships, the same standard models — subsets of General Time-Reversible (GTR) models — have been applied to phylogeny estimation with each newly published isopod mitogenome. Numerous methods have been proposed to overcome systematic bias in molecular phylogenetics. These include: data recoding to mitigate the effects of saturation, compositional biases, and model misspecification (Susko and Roger, 2007); clustering genes or codon positions into partitions that share similar evolutionary rates and applying separate best-fit substitution models to each partition (Frandsen et al., 2015); removal of fast-evolving (likely saturated) and invariant (lacking phylogenetic information) sites (Cummins and McInerney, 2011); application of more realistic substitution models that do not assume stationarity, reversibility, and homogeneity in the evolutionary process across lineages or sites (Moran et al., 2015); and, increased taxon sampling to break up long branches (Heath et al., 2008).

Having examined previous isopod phylogenies and their associated mitochondrial sequences, we believe they have all failed to recover an accurate topology since none account for the compositional biases inherent in the data. We reconstruct the mitogenomes of four more isopod species using high throughput sequencing, including three species of fish parasite from the family Cymothoidea and another oniscidean — *Ligia exotica*.

Additional representatives of Cymothoidea could help to establish their monophyly and, as suggested by Yu et al. (2018), increased *Ligia* representation could validate the sequence data available for *L. oceanica*, and confirm or refute a polyphyletic Oniscidea, and a dual origin of isopod terrestrialisation. Using these new data, and other published mitogenomes, we apply methods that focus on reducing compositional bias to test whether whole mitogenomes are suitable for estimation of deep isopod phylogeny by reducing the topological conflicts described. In particular, we assess the performance of site-heterogeneous models (CATGTR), which have been shown to overcome composition related error by accounting for site-specific evolutionary propensities — the probability of occurrence of different residues at a given site (Lartillot et al., 2007; Feuda et al., 2017; Philippe et al., 2019; Uribe et al., 2019). We use simulation to characterise data biases across genes and posterior predictive analyses to compare partitioned and site-heterogeneous models on different recoded datasets.

## 3.2 Materials and Methods

### Sampling and DNA Extraction

We extracted DNA from two pereopods of three cymothoid specimens — each a different species. *Ceratothoa italica* was found in the mouth of a striped seabream (*Lithognathus mormyrus*) from the Mediterranean by our research group in May 2007 (Sala-Bozano et al., 2012). Pereopods of *Cinusa tetrodontis* were taken from a specimen collected by Dr. Kerry Hadfield, North West University, in March 2016 from the Tsitsikamma Marine Protected Area, South Africa (permit number: MALH-K/2016-005). *Cinusa tetrodontis* was removed from the mouth of the species' only known host, the evil-eyed pufferfish (*Amblyrhynchotes honckenii*). Dr. Chrysoula Gubili, Hellenic Agricultural Association, collected the specimen of *Anilocra physodes* from an annular seabream (*Diplodus*

Table 3.1: Species taxonomy, Genbank accession numbers, mitogenome length (bp), proportion of A+T bases on the majority strand, and the strand AT and GC skews. Outgroup amphipod taxa are in brackets.

| Species                           | Family             | Suborder         | Accession         | Length | A+T  | AT skew | GC skew |
|-----------------------------------|--------------------|------------------|-------------------|--------|------|---------|---------|
| <i>Asellus aquaticus</i>          | Asellidae          | Asellota         | GU130252          | 13639  | 0.38 | 0.00    | -0.12   |
| <i>Janira maculosa</i>            | Janiridae          | Asellota         | GU130255          | 9871   | 0.29 | -0.04   | -0.03   |
| <i>Anilocra physodes</i>          | Cymothoidae        | Cymothoida       | <b>This Study</b> | 16606  | 0.37 | -0.00   | -0.13   |
| <i>Ceratothoa italica</i>         | Cymothoidae        | Cymothoida       | <b>This Study</b> | 16295  | 0.32 | -0.04   | -0.37   |
| <i>Cinusa tetrodontis</i>         | Cymothoidae        | Cymothoida       | <b>This Study</b> | 16278  | 0.38 | -0.03   | -0.34   |
| <i>Ichthyoxenos japonensis</i>    | Cymothoidae        | Cymothoida       | MF419233          | 15440  | 0.27 | 0.03    | -0.38   |
| <i>Bathynomus</i> sp.             | Cirolanidae        | Cymothoida       | KU057374          | 14965  | 0.41 | -0.09   | 0.17    |
| <i>Tachaea chinensis</i>          | Corallanidae       | Cymothoida       | MF419232          | 14616  | 0.27 | 0.06    | -0.35   |
| <i>Cymothoa indica</i>            | Cymothoidae        | Cymothoida       | MH396438          | 14475  | 0.36 | 0.13    | -0.44   |
| <i>Gyge ovalis</i>                | Bopyridae          | Cymothoida       | KY038053          | 14268  | 0.40 | -0.09   | 0.12    |
| <i>Eurydice pulchra</i>           | Cirolanidae        | Cymothoida       | GU130253          | 13055  | 0.44 | -0.05   | 0.20    |
| <i>Limnoria quadripunctata</i>    | Limnoriidae        | Limnoriidea      | KF704000          | 16515  | 0.34 | -0.10   | 0.19    |
| <i>Ligia oceanica</i>             | Ligiidae           | Oniscidea        | DQ442914          | 15289  | 0.39 | -0.04   | 0.13    |
| <i>Ligia exotica</i>              | Ligiidae           | Oniscidea        | <b>This Study</b> | 14829  | 0.40 | -0.04   | 0.12    |
| <i>Mongoloniscus sinensis</i>     | Agnaridae          | Oniscidea        | MG709492          | 16018  | 0.25 | -0.01   | 0.13    |
| <i>Cylisticus convexus</i>        | Cylisticidae       | Oniscidea        | KR013002          | 14154  | 0.32 | -0.04   | 0.19    |
| <i>Trachelipus rathkei</i>        | Trachelipodidae    | Oniscidea        | KR013001          | 14129  | 0.33 | -0.03   | 0.18    |
| <i>Porcellio dilatatus</i>        | Porcellionidae     | Oniscidea        | KX289582          | 14103  | 0.34 | -0.07   | 0.22    |
| <i>Porcellionides pruinosus</i>   | Porcellionidae     | Oniscidea        | KX289584          | 14078  | 0.40 | -0.09   | 0.25    |
| <i>Armadillidium vulgare</i>      | Armadillidiidae    | Oniscidea        | MF187613          | 13955  | 0.29 | -0.04   | 0.18    |
| <i>Armadillidium nasatum</i>      | Armadillidiidae    | Oniscidea        | MF187611          | 13943  | 0.32 | -0.04   | 0.16    |
| <i>Armadillidium album</i>        | Armadillidiidae    | Oniscidea        | KX289585          | 13812  | 0.30 | -0.04   | 0.17    |
| <i>Oniscus asellus</i>            | Oniscidae          | Oniscidea        | KX289581          | 13515  | 0.34 | -0.01   | 0.06    |
| <i>Eophreatoicus</i> sp.          | Amphisopidae       | Phreatoicoidea   | FJ790313          | 14994  | 0.30 | -0.10   | 0.25    |
| <i>Sphaeroma serratum</i>         | Sphaeromatidae     | Sphaeromatidea   | GU130256          | 13467  | 0.46 | -0.07   | 0.22    |
| <i>Idotea baltica</i>             | Idoteidae          | Valvifera        | DQ442915          | 14247  | 0.39 | -0.08   | 0.16    |
| <i>Glyptonotus antarcticus</i>    | Chaetiliidae       | Valvifera        | GU130254          | 13809  | 0.35 | -0.03   | 0.04    |
| ( <i>Metacrangonyx longipes</i> ) | Metacrangonyctidae | Order: Amphipoda | HE861923          | 14117  | 0.25 | -0.01   | -0.04   |
| ( <i>Gammarus fossarum</i> )      | Gammaridae         | Order: Amphipoda | KY197961          | 15989  | 0.35 | 0.02    | -0.26   |

*annularis*) near the Kitros Reef, Greece, in May 2016. Pereopods were homogenised in 200 $\mu$ l of Buffer ATL (Qiagen) using a pestle and mortar, before overnight lysis at 37°C with an additional 200 $\mu$ l of buffer and 40 $\mu$ l of proteinase K. Extraction and purification of DNA proceeded with a DNeasy Blood and Tissue kit (Qiagen). DNA was eluted in 20 $\mu$ l of Tris-HCl pH 8.5, which was applied to the column membrane twice in separate elution steps for a total of 40 $\mu$ l. The DNA concentration of each extract was quantified by fluorometry using a Qubit 3.0 high sensitivity assay (ThermoFisher Scientific). In addition to the new cymothoid sequence data we generated, we also downloaded whole genome shotgun sequence data from NCBI's Short Read Archive (BioProject: DRX049399, Run: DRR054550) for *Ligia exotica*.

### **Library Preparation and Sequencing**

DNA extracts were first sheared by sonication using a Biorupter (Diagenode) to target a modal fragment length of 300bp. Single indexed libraries were then prepared for *C. italica* and *C. tetradontis* using a KAPA HyperPlus kit (Roche) including 12 PCR cycles. A PCR-free library was prepared for *Anilocra physodes* with a TruSeq DNA kit (illumina). We quantified library concentrations with Qubit and library fragment lengths using a TapeStation 4200 (Agilent). *C. italica* and *C. tetradontis* libraries were pooled in equimolar concentrations, and were sequenced for 300 cycles (PE100) using half a lane of a HiSeq 2500 (illumina) by Macrogen, South Korea. The *A. physodes* library was sequenced in a separate run, using the same instrument, number of cycles, and proportion of a single lane.

To test for atypical mitogenome structure in cymothoids (Doublet et al., 2012), we also obtained long read data for *Anilocra physodes* with nanopore sequencing (Oxford Nanopore Technologies). Additional, high molecular weight DNA was purified from the tissue of half our *A. physodes* specimen (~50mg) using the same extraction method

described above. We then prepared a library following the open-source, one-pot ligation protocol (Quick, 2017). Sequencing was stopped after 5Gbp of data had been generated.

## Assembly

We filtered adapter, PhiX, and human contamination from each set of demultiplexed, forward and reverse reads for each isopod species using `bbduk` 38.96 (Bushnell, 2019). We then performed *de novo* assembly using `NOVOPlasty` 3.1 (Dierckxsens et al., 2016) with conspecific mitochondrial Cytochrome Oxidase subunit I (COI) sequences as seeds (Table 3.2). Seed sequences were also used for assembly of each other species, and this experiment was repeated using the largest contigs obtained from the initial draft assemblies. Assemblies were considered complete if we could scaffold contigs into one circularised sequence where ends overlapped by at least 200bp, and the assembly length was between 14Kbp and 20Kbp — within the range of most eukaryotic mitochondrial genomes. Where a draft assembly contained multiple contigs (usually breakpoints within repeat regions), or could not be circularised, we manually located overlaps and scaffolded contigs using `AliView` 1.18 (Larsson, 2014). The filtered short reads were mapped back to the best draft assembly for each species using `BBMap` to obtain coverage statistics, and mapping was visualised with `IGV` 2.5 (Thorvaldsdóttir et al., 2013) to assess assembly quality. Additionally, we assembled reads mapped to the previously published mitogenome for *Cymothoa indica* with `Edena` v3 (Hernandez et al., 2008) to compare against *de novo* results.

Under these criteria the mitogenomes of *A. physodes* and *L. exotica* could only be circularised when their lengths were double that expected: 33,655bp and 29,966bp respectively. In both cases these size discrepancies were due to single, near perfect, palindromic repeats, which we visualised with dotplots using `flexidot` 1.02 (Seibt et al., 2018). To confirm whether the observed palindromes were real mitogenome

features or their presence was due to misassembly, we used a hybrid assembly approach incorporating nanopore long read data for *A. physodes* in an attempt to bridge poorly mapped regions and the junction of the inversion. Nanopore FAST5 files were first base called using `Albacore v2.1.3` (Oxford Nanopore Technologies). We then performed separate error-correction of our *A. physodes* short reads with `Spades 3.12` (Bankevich et al., 2012) and assembled both short and long reads with `Unicycler` (Wick et al., 2017). In a second hybrid assembly we used `minimap2` (Li, 2018) to first map both short and long reads to the 33Kbp circularised *A. physodes* draft assembly. Only the successfully mapped reads were then used as input for `Unicycler`. Long reads were also mapped back to the best *A. physodes* assemblies with `minimap2` and coverage statistics extracted with `samtools 1.8` (Li et al., 2009).

For all species, including *L. exotica*, we also assembled nuclear 18S ribosomal DNA (rDNA), also using `NOVOPlasty 3.1` and a publicly available sequence for *Anilocra physodes* (AF255686; Dreyer and Wägele, 2001). Finally, we also attempted to reconstruct the mitogenomes of host species since it was likely that host cells were present either externally or within the parasite gut, and their DNA may have been co-extracted. For this experiment we, again, used `NOVOPlasty` and the seed sequences KM538406 for *Lithognathus mormyrus*, both JF492826 and NC031325 for *Amblyrhynchotes honckenii*, and MF464087 for *Diplodus annularis*. AT content and strand compositional bias were calculated for the assembled isopod mitogenomes using a custom script with functions from the package `seqinr` (Charif and Lobry, 2007), and the skewness equations from Perna and Kocher (1995).

## Annotation

Given assembly uncertainties of possible whole mitogenome duplications (detailed description presented in the Results) we annotated the best assembly for each species

Table 3.2: Assembly references and parameter settings for each cymothoid species. Datasets are percentage of total illumina paired reads, normalised coverage and the maximum GC content across reads. Unicycler assembly was only used for *Anilocra physodes*. Reference is the species for which a seed reference sequence was obtained, and Source is their Genbank accession number: COI = Cytochrome Oxidase subunit I barcode; Draft = draft assembly or contig produced in this study; Mito = whole mitogenome; 18S = small subunit nuclear ribosomal DNA; ONT = Oxford Nanopore Technologies.

| Sample                    | Reference                 | Source (locus)            | Assembly                                     |
|---------------------------|---------------------------|---------------------------|--|
|                           | <i>Anilocra physodes</i>  | EF455817 (COI)            |  |
|                           | <i>Cinusa tetrodontis</i> | MK652480 (COI)            |  |
|                           | <i>Ceratothoa italica</i> | JN604340 (COI)            |  |
| <i>Anilocra physodes</i>  | <i>Ceratothoa italica</i> | <b>This Study</b> (Draft) |  |
|                           | <i>Cinusa tetrodontis</i> | <b>This Study</b> (Draft) | A. NOVOPlasty                                |
|                           | <i>Cymothoa indica</i>    | MH396438 (Mito)           |  |
|                           | <i>Anilocra physodes</i>  | AF255686 (18S)            | kmer: 23, 33, 49                             |
|                           | <i>Ceratothoa italica</i> | JN604340 (COI)            | Datasets: 20%, 33%, 50X, 100X, GC-filter=0.4 |
|                           | <i>Anilocra physodes</i>  | EF455817 (COI)            |  |
|                           | <i>Cinusa tetrodontis</i> | MK652480 (COI)            | B. Unicycler                                 |
| <i>Ceratothoa italica</i> | <i>Cinusa tetrodontis</i> | <b>This Study</b> (Draft) |  |
|                           | <i>Anilocra physodes</i>  | <b>This Study</b> (Draft) | Error-corrected illumina reads               |
|                           | <i>Cymothoa indica</i>    | MH396438 (Mito)           | Datasets: 20%, 33%, 50X, 100X                |
|                           | <i>Anilocra physodes</i>  | AF255686 (18S)            | ONT reads                                    |
|                           | <i>Cinusa tetrodontis</i> | MK652480 (COI)            |  |
|                           | <i>Anilocra physodes</i>  | EF455817 (COI)            | C. Edena                                     |
|                           | <i>Ceratothoa italica</i> | JN604340 (COI)            | Mapped reads to <i>C. indica</i>             |
| <i>Cinusa tetrodontis</i> | <i>Ceratothoa italica</i> | <b>This Study</b> (Draft) |  |
|                           | <i>Anilocra physodes</i>  | <b>This Study</b> (Draft) |  |
|                           | <i>Cymothoa indica</i>    | MH396438 (Mito)           |  |
|                           | <i>Anilocra physodes</i>  | AF255686 (18S)            |  |

that was the correct length. These were, one of the palindromic arms for *L. exotica*, a GC-filtered assembly for *A. physodes*, and two circularised assemblies for *C. italica* and *C. tetradontis*. Draft assemblies were first BLAST-searched against the complete NCBI nucleotide database to authenticate them as isopod mitogenomes. Assemblies were then annotated using the MITOS web server (Bernt et al., 2013). MITOS annotations for Protein Coding Genes (PCG) were verified and edited by identifying open reading frames with ORF Finder (Wheeler et al., 2003).

Prediction of tRNAs proved difficult due to overlaps and atypical structures, so we used a stand-alone version of `MiTFi 0.1` (Jühling et al., 2011) to search for tRNAs independently of other genome features. We also compared `MiTFi` results to those of `ARWEN v1.2` (Laslett and Canbäck, 2007) — a pattern matching algorithm that searches for diagnostic conformations such as tRNA triple stems. We accepted the boundaries of 12S and 16S produced by the MITOS annotation. We wrote a custom R function based on the package `circlize` (Zuguang et al., 2014) to visualise mitogenome features, their composition, and sequencing coverage (available at <https://github.com/bailliecharles/mitoplotR>).

## Phylogenetic Analysis

### Sequence Alignment

We downloaded complete, or near-complete, isopod mitochondrial genomes over 9Kb and 18S rRNA for the same species (Table 3.1). Sequences for two amphipod outgroup taxa were also downloaded, which in previous analyses were found to be the sister clade to isopods within Peracarida (Richter and Scholtz, 2001; Spears et al., 2005; Wilson, 2009). Where partial mitogenomes did not contain certain genes, we used conspecific sequences for those regions if they were available in Genbank. We did not align tRNA sequences due to uncertainty surrounding their precise locations and their abnormal configurations. PCGs were individually aligned using `TranslatorX` (Abascal et al., 2010), followed by

removal of ambiguously-aligned positions by Gblocks (Castresana, 2002), retaining sites with up to 50% gaps (option 'b5 = h'). We removed ATP8 altogether because it could not be confidently aligned (Gblocks actually returned an empty alignment), and upon visual inspection, we deleted the ND4 gene for *Oniscus asellus* which contains clear annotation errors including spurious insertions and premature stop codons. The mitochondrial and nuclear rRNA genes were aligned separately using 1000 iterations of the `mafft-xinsi` algorithm (Katoh and Standley, 2013), which incorporates information from RNA secondary structure to guide alignment. Ambiguous sites with gaps in more than 25% of the sequences or with a similarity score lower than 0.001 were removed using `trimAl v1.2` (Capella-Gutiérrez et al., 2009).

To assess how different datasets might affect phylogeny estimation, we constructed five concatenated alignments using FASconCAT-G (Kück and Longo, 2014): PCG-123 included the complete PCG; PCG-3RY where third codon positions were recoded as purines and pyrimidines (RY-coding); PCG-3X in which third codon positions were removed; PCG-AA — amino acid translations of the PCG using the invertebrate mitochondrial code; PCG-DAY — amino acid sequences recoded as six Dayhoff groups (Dayhoff et al., 1978). The rRNA alignments were then concatenated with each of the nucleotide alignments including just mitochondrial 12S and 16S, then, additionally, nuclear 18S. Based on results from compositional homogeneity tests (see subsection 'Sequence Analysis and Model Selection') we constructed another concatenated alignment, PCG-HOMO, that included the most compositionally homogeneous genes from the PCG-DAY alignment. Finally, for the amino acid and Dayhoff-recoded supermatrices we also calculated the evolutionary rates of each site using TIGER (Cummins and McInerney, 2011), split these into ten bins using the relative rate method of Rota et al. (2018), and then removed the fastest and invariant partitions (PCG-AA-R and PCG-DAY-R). In total, we analysed 15 different datasets. The full nucleotide alignment, including all rRNAs, was 10,743bp long

(8,114bp without third codon positions), the amino acid and Dayhoff-recoded alignments had 2,629 sites, the homogeneous alignment 1,708 sites, and the two rate-trimmed, amino acid alignments had 1,450 sites and 1,128 sites, respectively.

### **Sequence Analysis and Model Selection**

Evidence of substitution saturation was examined for whole genes and individual codon positions with plots of transition/transversion ratios ( $\kappa$ ) against p-distances using the R package `ape` 5.0 (Paradis and Schliep, 2018) and calculating the C-value using `BaCoCa` v1.1 (Kück and Struck, 2014). `BaCoCa` was also used to analyse nucleotide frequencies and compositional homogeneity of each gene and codon position using  $\chi^2$  tests. However, the  $\chi^2$  distribution is a poor approximation for testing compositional differences because non-independence between samples increases the probability of type II error. Therefore, we also used the method described by Foster (2004) implemented in `P4`, which constructs a null distribution of  $\chi^2$  statistics drawn from simulations under a homogeneous model rather than generating a  $\chi^2$  curve directly from the data.

Best fit partition schemes and site-homogeneous substitution models were evaluated using `PartitionFinder2` (Lanfear et al., 2016) with corrected Akaike's Information Criterion (AICc), linked branch lengths, and the 'greedy' search algorithm. Partitions were selected by codon position in the nucleotide alignments, by gene in the amino acid alignments, rate partition for the rate-trimmed datasets, and each of the rRNA genes were treated as separate partitions. Since different types of substitution model cannot be applied to separate partitions in `RAxML` (though partitions can have individual models of the same type) where different models were found for different partitions, we selected between them by re-running `PartitionFinder2` under each of these models and selected that with the lowest AICc.

### **Phylogeny Estimation and Posterior Predictive Analyses**

One of our aims was to test whether site-homogeneous partitioned analyses or site-heterogeneous mixture models could better account for biases in our mitogenome data and overcome Long Branch Attraction (LBA) artefacts. For partitioned analyses, we estimated phylogeny using Maximum Likelihood (ML) and Bayesian Inference (BI). ML analyses were conducted in `RAxML v8` (Stamatakis, 2014) and BI using a version of `MrBayes` (Ronquist et al., 2012) that includes additional empirical amino acid substitution matrices (`mrBayes5d`: Tanabe, 2008). For ML analyses of the PCG-DAY and PCG-DAY-R datasets, we converted alphanumeric residue codes to zero-based integers and applied the MULTIGAMMA model, and for all datasets topological support was assessed with a number of rapid bootstraps determined by automatic stopping (option 'autoMRE'). Bayesian analyses were initially run for five million generations and included two simultaneous, independent runs, each with one cold and three heated chains. Runs were sampled every 1,000 generations, and the analyses were continued for a further two million generations at a time until the standard deviation of split frequencies was less than 0.01. We considered the analyses had converged when the Potential Scale Reduction Factor (PSRF) was close to one, the variance in PSRF between model parameters was low, and the mean Effective Sample Size (ESS) was over 200 for each parameter. We also plotted parameter traces in `Tracer v1.7` (Rambaut et al., 2018) to assess mixing and determine an appropriate burn-in. The post burn-in set of trees were then summarised as the 95% majority rule consensus and node support was assessed with Bayesian Posterior Probabilities (BPP).

Site-heterogeneous analyses used the CATGTR model implemented in `PhyloBayes MPI 4.1` (Lartillot et al., 2013), and these were conducted on the *Danzek* High Performance Cluster at The University of Manchester. Due to computational demands we only analysed PCG-AA, PCG-DAY, PCG-HOMO, and the nucleotide PCG datasets, i.e. those

without rRNAs. Analyses for each dataset included two independent chains, each set to default priors and model parameters. Convergence and mixing were assessed with ‘bpcomp’ and ‘tracecomp’ (utilities packaged with `PhyloBayes`), and chains were stopped when the largest discrepancies (‘maxdiff’) were less than 0.1 and parameter minimum effective sizes were above 300 (thresholds recommended in the `PhyloBayes` manual). One thousand post burn-in trees were summarised as the 95% majority rule consensus. Trees from all analyses were plotted using the R packages `ape` 5.0 and `phytools` 0.6 (Revell, 2012).

To assess the adequacy of partitioned and site-heterogeneous approaches for modelling isopod mitogenome sequences we conducted three Posterior Predictive Analyses (PPAs). PPAs simulate data from a model using posterior parameter estimates from Bayesian phylogenetic analyses. Statistics calculated for each of these simulations create a null (posterior predictive) distribution against which statistics for the empirical data are compared (Lartillot et al., 2007). Site-specific base diversity (PPADIV) tests how well models estimate the mean number of distinct bases observed at each site (Feuda et al., 2017). Compositional homogeneity (PPAX2) is used to assess the composition of each sequence in the alignment to the average base composition. The multinomial likelihood (PPAMULTI: Goldman, 1993) describes site patterns and is a measure of alignment plausibility (Bollback, 2002). For each test we first simulated 1,000 alignments from randomly drawn sets of post burn-in posterior estimates. Simulations under CATGTR were generated using the `PhyloBayes` function ‘readpb\_mpi’ with the option ‘-ppred’. For partitioned analyses we used P4, with some modified code, to generate simulated alignments from `MrBayes` parameter output files for the same datasets run with `PhyloBayes`. Each PPA was then calculated using a custom R script and the deviation of the empirical test statistic from null expectation was compared with z-scores. The R script and code we modified from P4 are available from <https://github.com/bailliecharles/PPA>.

## 3.3 Results

### Mitogenome Assembly and Annotation

Short read sequencing produced greater than 11-fold whole genome coverage per specimen based on an estimated genome size of 1.6Gbp (Table 3.3). We obtained fully circularised mitogenomes for *Ceratothoa italica* and *Cinusa tetrodontis* within the 14Kbp – 20Kbp range, but both of these assemblies are longer than the mitogenomes of two previously sequenced cymothoids — *Cymothoa indica* (14,475bp; Zou et al., 2018) and *Ichthyoxenos japonensis* (15,440bp; Hua et al., 2018). This size difference is due to the increased length of the Non-Coding Region (NCR) between 12S and NAD1, including an approximately 200bp stretch where an excess of reads map poorly, creating a high coverage spike (Figure 3.2, 3.3, 3.4, 3.5). We also see a longer NCR in our best assembly for *A. physodes*. Such an apparent elongation is perhaps due to structural complexity in these regions causing assembly error, which poorly aligned reads seem to confirm. Similar to the NCR of *Cymothoa indica*, we do not find any evidence that the NCRs are comprised of linear repeats as found in *Ichthyoxenos japonensis* (Hua et al., 2018; Zou et al., 2018).

Table 3.3: Read, mapping, and coverage statistics per species. *A. physodes* includes illumina (ill) short reads and Oxford Nanopore Technologies (ONT) long reads. Reads mapped are pairs of reads for illumina data and ONT data are single reads.

| Species                     | Raw Reads | Raw Bases   | Mean Insert (bp) | Reads Mapped (%) | Mean Coverage (sd) |
|-----------------------------|-----------|-------------|------------------|------------------|--------------------|
| <i>A. physodes</i> (ill)    | 402760752 | 60816873552 | 467.71           | 329937 (0.16)    | 5993.87 (687.91)   |
| <i>A. physodes</i> (ONT)    | 2537847   | 4921077153  | 1939 (500)       | 7252 (0.07)      | 200.94 (45.12)     |
| <i>C. italica</i> (ill)     | 150894892 | 19925016604 | 203.81           | 30221 (0.04)     | 584.59 (1933.88)   |
| <i>C. tetrodontis</i> (ill) | 130854316 | 18076132843 | 230.11           | 49873 (0.07)     | 859.07 (127.41)    |
| <i>L. exotica</i> (ill)     | 37291858  | 8585565501  | 301.03           | 23291 (0.12)     | 729.51 (138.97)    |

For *Anilocra physodes* and *Ligia exotica* we could only assemble circular mitogenomes that were double the expected length (33,655bp and 29,966bp, respectively), which are comprised of near-perfect palindromic sequences (Figure 3.1). Smaller inverted repeat contigs (~2Kbp) were also found for all species where NOVOPlasty parameter combinations resulted in dis-contiguous assemblies of several contigs that could not be overlapped in the same orientation. Long reads and reference guided assemblies could not resolve these phenomena — the best hybrid output for *A. physodes* producing a seven node assembly, the largest contig of which was 16,053bp. Using a GC-filtered short read dataset for *A. physodes* resulted in a circular, single contig assembly of 16,606bp, which aligned exactly with one arm of the palindrome recovered in the double length draft.

Taken together, we tentatively propose that the length of NCRs and the poorly mapped spans within them, as well as the peculiar *A. physodes* assembly, are evidence that species of the family Cymothoidae also possess the atypical isopod mitogenome structure of two identical molecules fused in a head-to-head dimer (Raimond et al., 1999; Marcadé et al., 2007; Peccoud et al., 2017). The locations of the origin of replication (OR) and the junction of the head-to-head dimers are most likely downstream of the NCR (closest to ND1), as seen in oniscids (Marcadé et al., 2007; Doublet et al., 2012; Peccoud et al., 2017). Additionally, *L. exotica* may also possess the same mitochondrial conformation. Previous research found the congener *L. oceanica* to have normal mitogenome structure but those results were inconclusive (Doublet et al., 2012). In addition to the palindromic assembly, we also find evidence of heteroplasmic sites within tRNA-L1 and tRNA-S1 of *L. oceanica*, which have been associated with atypical mitochondria (Peccoud et al., 2017).

Notwithstanding assembly uncertainty, we recovered all PCGs and rRNAs for each cymothoid species, and for *L. exotica* (Figure 3.2, 3.3, 3.4, 3.5). These features are all arranged in the expected order, matching that of other cymothoids, and the congener

Table 3.4: Mitogenome annotation of *de novo* assemblies for the three cymothoid species used in this study. Annotation for *Anilocra physodes* is based on our GC-filtered assembly described in the Results. Start and Stop are base positions for beginning and end of gene features zeroed on COI. Length is the length of a feature in base pairs, and Strand is the strand on which a feature is encoded. Asterisk denotes translocated tRNA-R in *A. physodes*.

| Feature       | Strand | <i>Anilocra physodes</i> |       |        | <i>Ceratothoa italica</i> |       |        | <i>Cinusa tetrodonis</i> |       |        |
|---------------|--------|--------------------------|-------|--------|---------------------------|-------|--------|--------------------------|-------|--------|
|               |        | Start                    | Stop  | Length | Start                     | Stop  | Length | Start                    | Stop  | Length |
| COX1          | +      | 1                        | 1512  | 1512   | 1                         | 1533  | 1533   | 1                        | 1531  | 1531   |
| COX2          | +      | 1615                     | 2262  | 648    | 1582                      | 2265  | 684    | 1586                     | 2276  | 690    |
| tRNA-K (ttt)  | +      | 2279                     | 2333  | 55     | 2433                      | 2494  | 62     | 2266                     | 2324  | 58     |
| tRNA-D (gtc)  | +      | 2334                     | 2394  | 61     | 2493                      | 2550  | 58     | 2325                     | 2381  | 56     |
| ATP8          | +      | 2396                     | 2548  | 153    | 2551                      | 2706  | 156    | 2381                     | 2537  | 156    |
| ATP6          | +      | 2647                     | 3216  | 570    | 2700                      | 3434  | 735    | 2554                     | 3244  | 651    |
| COX3          | +      | 3221                     | 4000  | 780    | 3436                      | 4167  | 732    | 3246                     | 4001  | 755    |
| tRNA-R (tcg)* | +      | NA                       | NA    | NA     | 4214                      | 4268  | 55     | 3993                     | 4051  | 58     |
| tRNA-G (tcc)  | +      | 4049                     | 4118  | 70     | 4165                      | 4220  | 56     | 4051                     | 4101  | 51     |
| NAD3          | +      | 4112                     | 4447  | 336    | 4274                      | 4627  | 354    | 4098                     | 4452  | 354    |
| tRNA-R (tcg)* | +      | 4948                     | 5006  | 59     | NA                        | NA    | NA     | NA                       | NA    | NA     |
| tRNA-A (tgc)  | +      | 5161                     | 5219  | 59     | 4618                      | 4671  | 54     | 4456                     | 4511  | 55     |
| tRNA-V (tac)  | +      | 5354                     | 5411  | 58     | 4671                      | 4730  | 60     | 4512                     | 4572  | 60     |
| tRNA-N (gtt)  | +      | 5796                     | 5859  | 64     | 4729                      | 4790  | 62     | 4571                     | 4632  | 61     |
| 12S rRNA      | +      | 5817                     | 6520  | 704    | 4790                      | 5496  | 707    | 4632                     | 5342  | 710    |
| NAD1          | -      | 7886                     | 8782  | 897    | 6807                      | 7736  | 930    | 7596                     | 8517  | 921    |
| tRNA-L1 (tag) | -      | 8815                     | 8875  | 61     | 7760                      | 7819  | 60     | 8525                     | 8586  | 61     |
| tRNA-E (ttc)  | -      | 8868                     | 8921  | 54     | 7821                      | 7860  | 40     | 8587                     | 8640  | 52     |
| tRNA-L2 (taa) | -      | 8927                     | 8987  | 61     | 7863                      | 7922  | 60     | 8641                     | 8701  | 60     |
| tRNA-S1 (tct) | -      | 8983                     | 9042  | 60     | 7914                      | 7976  | 63     | 8701                     | 8767  | 60     |
| tRNA-W( tca)  | -      | 9053                     | 9106  | 54     | 7964                      | 8032  | 69     | 8768                     | 8837  | 69     |
| CYTB          | -      | 9139                     | 10221 | 1083   | 8031                      | 9155  | 1125   | 8837                     | 9927  | 1090   |
| tRNA-T (tgt)  | -      | 10231                    | 10284 | 54     | 9151                      | 9209  | 59     | 9927                     | 9979  | 52     |
| NAD5          | +      | 10311                    | 11972 | 1662   | 9194                      | 10921 | 1728   | 10005                    | 11664 | 1659   |
| tRNA-F (gaa)  | +      | 11977                    | 12034 | 58     | 10905                     | 10962 | 58     | 11674                    | 11731 | 57     |
| tRNA-H (gtg)  | -      | 12027                    | 12083 | 57     | 10955                     | 11013 | 59     | 11723                    | 11781 | 58     |
| NAD4          | -      | 12097                    | 13293 | 1197   | 11013                     | 12317 | 1305   | 11803                    | 13079 | 1276   |
| NAD4L         | -      | 13395                    | 13637 | 243    | 12311                     | 12598 | 288    | 13087                    | 13339 | 252    |
| tRNA-P (tgg)  | -      | 13686                    | 13744 | 59     | 12611                     | 12667 | 57     | 13381                    | 13438 | 57     |
| NAD6          | +      | 13762                    | 14208 | 447    | 12660                     | 13163 | 504    | 13448                    | 13946 | 498    |
| tRNA-S2 (tga) | +      | 14225                    | 14285 | 61     | 13159                     | 13218 | 60     | 13934                    | 13993 | 57     |
| 16S rRNA      | -      | 14316                    | 15250 | 935    | 13176                     | 14289 | 1114   | 13994                    | 15070 | 1076   |
| tRNA-Q (ttg)  | -      | 15372                    | 15427 | 56     | 14267                     | 14346 | 80     | 15071                    | 15129 | 57     |
| tRNA-M (cat)  | +      | 15432                    | 15492 | 61     | 14338                     | 14397 | 60     | 15120                    | 15183 | 63     |
| NAD2          | +      | 15532                    | 16416 | 885    | 14398                     | 15387 | 990    | 15184                    | 16173 | 989    |
| tRNA-C (gca)  | -      | 16480                    | 16529 | 50     | 15369                     | 15448 | 80     | 16167                    | 16218 | 51     |
| tRNA-Y (gta)  | -      | 16544                    | 16601 | 58     | 15830                     | 15888 | 59     | 16219                    | 16278 | 59     |

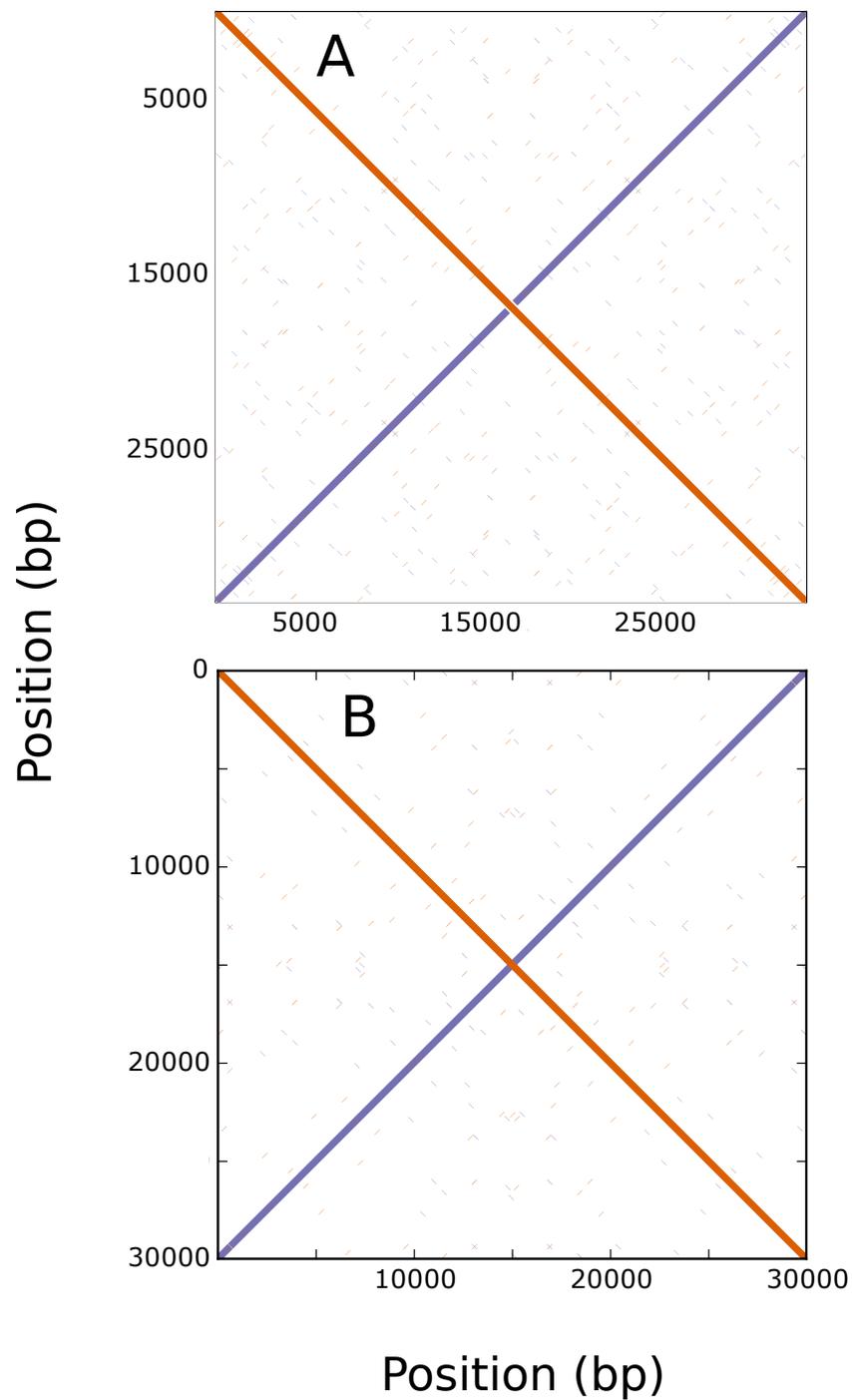


Figure 3.1: Dotplot of fully circularised *Anilocra physodes* (A) and *Ligia exotica* (B) draft assemblies (orange) and their reverse complement (purple). Plots were calculated with a 9bp sliding window. Positional coordinates are in bases (bp).

*Ligia oceanica* in the case of *L. exotica*. There is no evidence of frame shift mutations or premature stop codons, and all PCGs are of expected length. While start codons all conform to the invertebrate mitochondrial code, an alternative stop codon, T(AA), likely terminates transcription of Cytochrome Oxidase subunit II (COII) in *C. tetradontis*. At 14,983bp, our *L. exotica* assembly is 306bp shorter than the mitogenome of the congener *L. oceanica* (15,289bp). This size difference could be due to the fact we did not find the triple tRNA cluster downstream of 12S that is present in *L. oceanica*. In fact, tRNAs proved particularly difficult to annotate for all species due to large gene overlaps and atypical conformations — common properties of isopod tRNAs (Chandler et al., 2015; Doublet et al., 2015). Our *A. physodes* assembly appears to have a tRNA-R translocation and additional NCRs within the tRNA cluster between NAD3 and 12S (Table 3.4; Figure 3.4). In total, we identified 21 tRNAs in each of our cymothoid mitogenomes — the same number found in *Ichthyoxenos japonensis*. Zou et al. (2018) only found 20 tRNAs across the *Cymothoa indica* mitogenome, but did record a 46bp gap between tRNA-L1 and tRNA-L2 where tRNA-E is found in our cymothoid species and *Ichthyoxenos japonensis*.

We were able to assemble and annotate near-complete mitogenomes of *Lithognathus mormyrus* and *Diplodus annularis*, the host fish of *Ceratothoa italica* and *Anilocra physodes*, respectively. Until now, only single mitochondrial genes have been available for these species despite their importance in commercial fisheries. The mitogenome of *Amblyrhynchotes honckenii* could not be assembled, even partially, regardless of the seed sequence used. Our *C. tetradontis* specimen was collected primarily for morphological taxonomy, and we presume the preservation method was not conducive to host cell preservation.



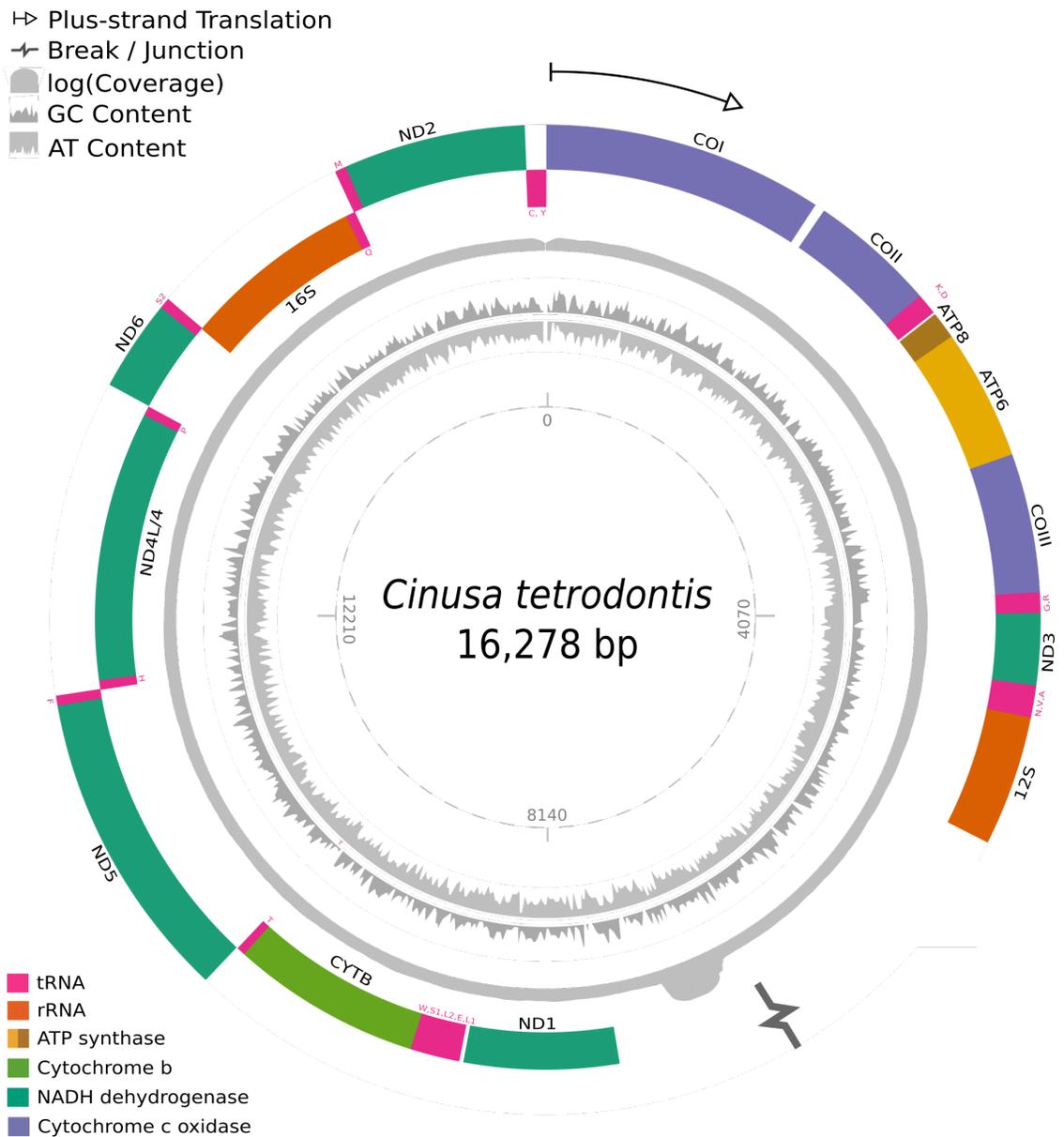


Figure 3.3: Mitochondrial genome map for *Cinusa tetrodontis*. Break points denote the hypothesised junction of the ‘head-to-head’ dimer if assembled mitogenomes are found to have atypical architectures. Majority strand is encoded on the outside track in a clockwise direction.

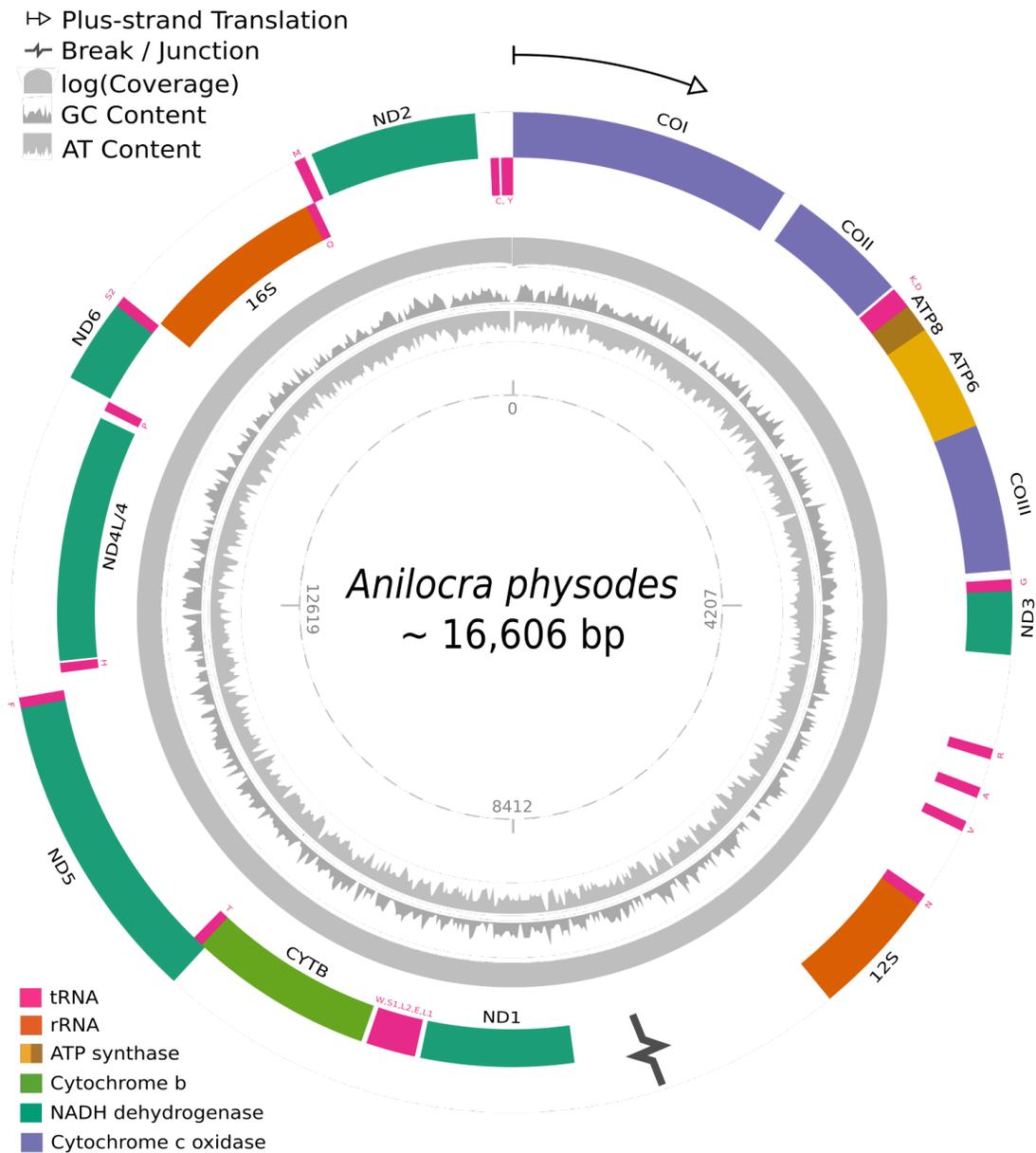


Figure 3.4: Mitochondrial genome map for *Anilocra physodes*. Break points denote the hypothesised junction of the ‘head-to-head’ dimer if assembled mitogenomes are found to have atypical architectures. Majority strand is encoded on the outside track in a clockwise direction.

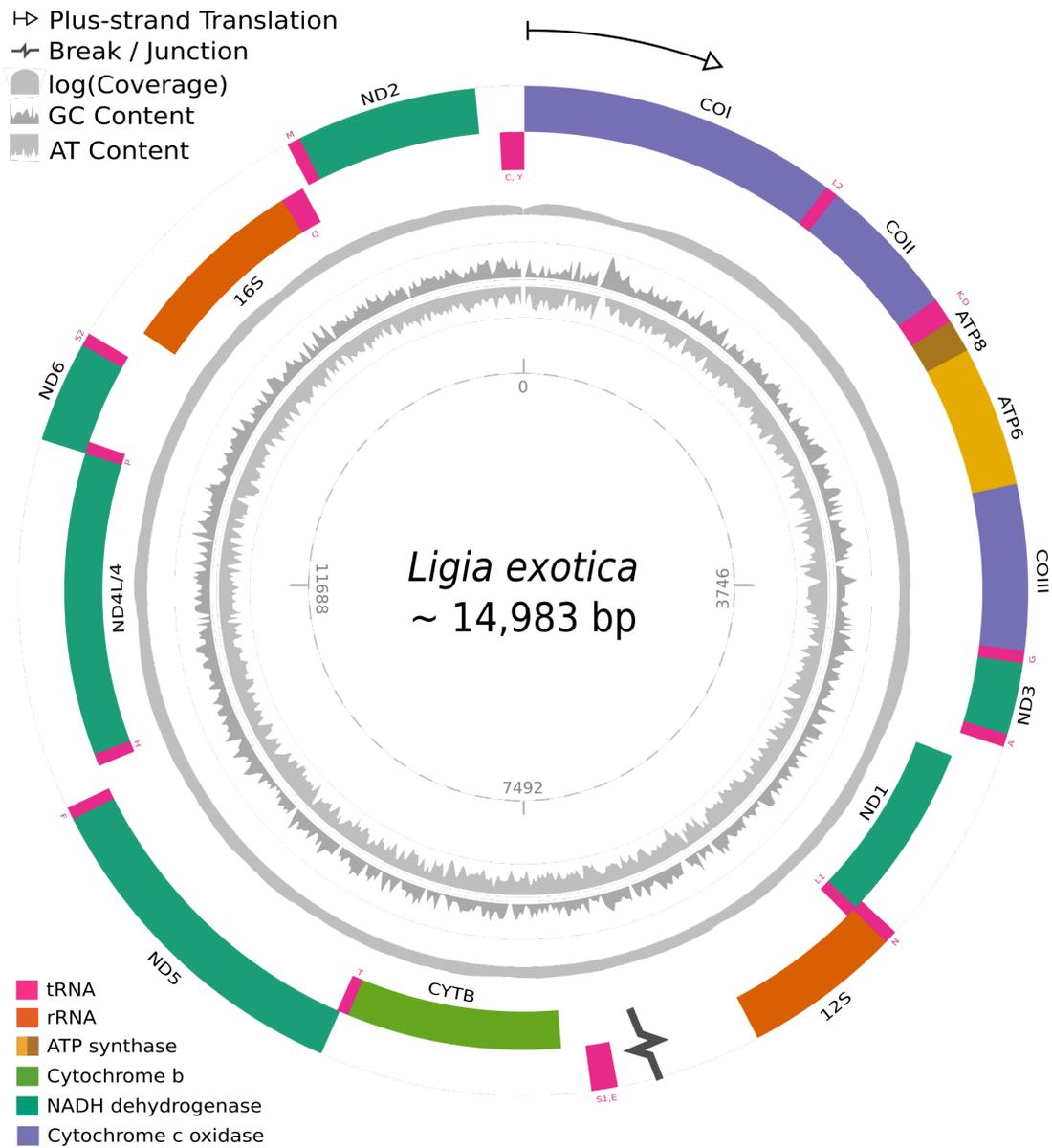


Figure 3.5: Mitochondrial genome map for *Ligia exotica*. Break points denote the hypothesised junction of the ‘head-to-head’ dimer if assembled mitogenomes are found to have atypical architectures. Majority strand is encoded on the outside track in a clockwise direction.

## Gene and Strand Compositional Bias

Our simulation-based homogeneity tests show that standard  $\chi^2$  statistics consistently fail to reject the null hypothesis that sequences are homogeneous between taxa (Table 3.6). For amino acid translations, only two genes were homogeneous with P4, compared to twelve using standard homogeneity tests, and none of the reduced nucleotide genes were found to be homogeneous. Compositional bias is attenuated for amino acids when alphabet size is reduced: we identify eight homogenous genes, from twelve, with recoded Dayhoff groups. As expected, all genes show relatively high levels of substitution saturation (C-value < 4), driven by first and third codon positions (Kück and Struck, 2014).

Results of strand nucleotide composition (Table 3.1) corroborate previous findings that there are differences in strand asymmetry between taxa, and that the strength of strand bias is more pronounced for GC- than AT-skews (Kilpert et al., 2012; Shen et al., 2017). All oniscideans show positive GC-skew, while in the suborder Cymothoidea only cymothoid isopods have a negative strand bias. Both asellotans, which together with cymothoids have previously been recovered as the sister taxon to other isopods, have a negative GC skew. Interestingly, negative GC-skews are also shared by our outgroup amphipod taxa.

## Phylogeny Estimation

PPAs show incremental improvements with increasing compositional homogeneity and improved modelling (Table 3.5). Our best fitting model and dataset combination (CATGTR + PCG-HOMO) appears to better account for mitogenome compositional biases and ameliorates some LBA effects. Only in this analysis, for example, do we recover a monophyletic Oniscidea (Figure 3.6). In all other trees, Ligiidae are mostly found sister to members of a questionable clade of problematic taxa, which change position on the tree

Table 3.5: Results of Posterior Predictive Analyses (PPAs) to compare Bayesian partitioned and profile mixture model adequacy for describing various datasets. Empirical values are those calculated from the data, whereas z-scores are given for posterior simulations for Partitioned and CATGTR results. Lower z-scores show better model adequacy. PPADIV = base diversity, PPAX2 =  $\chi^2$  test of compositional homogeneity, PPMULTI = multinomial of site patterns from Bollback (2002). AA = amino acids, DAY = Dayhoff recoded amino acids, HOMO = homogeneous dataset, 123 = full protein coding dataset, 12R = full protein coding dataset with 3<sup>rd</sup> codon positions RY-coded, 12X = full protein coding dataset with 3<sup>rd</sup> codon positions removed.

| Data     | Empirical |         |           | Partitioned |        |         | CATGTR |        |         |
|----------|-----------|---------|-----------|-------------|--------|---------|--------|--------|---------|
|          | PPADIV    | PPAX2   | PPMULTI   | PPADIV      | PPAX2  | PPMULTI | PPADIV | PPAX2  | PPMULTI |
| PCG-AA   | 3.71      | 742.04  | -19028.14 | 10.79       | 30.97  | 5.22    | 2.06   | 35.12  | 4.11    |
| PCG-DAY  | 2.00      | 154.38  | -16799.82 | 4.06        | 17.15  | 1.99    | 1.08   | 0.78   | 1.58    |
| PCG-HOMO | 1.92      | 92.61   | -9465.55  | 4.06        | 7.83   | 1.61    | 1.08   | 0.41   | 1.46    |
| PCG-123  | 2.63      | 2613.43 | -63204.41 | 15.21       | 318.91 | 0.71    | 1.27   | 355.57 | 1.68    |
| PCG-3RY  | 2.07      | 705.54  | -61597.45 | 17.45       | 34.46  | 0.07    | 0.03   | 105.36 | 1.44    |
| PCG-3X   | 2.20      | 630.62  | -37483.92 | 7.78        | 91.89  | 0.68    | 0.30   | 108.01 | 1.77    |

depending on the analysis used (partitioned and remaining site-heterogeneous trees can be found in the Appendix). We also find *Bathynomus sp.* and *Gyge ovalis* (both Cymothoidea) cluster together in our CATGTR + PCG-HOMO analysis, albeit separately from the main Cymothoidea clade. The morphological expectation in our datasets is that the cirolanids *Eurydice pulchra* and *Bathynomus sp.* should be sister species. However, *E. pulchra* behaves erratically across all datasets and estimation methods, and we recover this pairing only twice, both in partitioned analyses of recoded amino acids: ML + PCG-DAY, BI + PCG-HOMO. Most often *E. pulchra* is grouped with another rogue taxon, *Sphaeroma serratum*, or as sister to *Bathynomus sp.* + *S. serratum* within the problem clade.

*Limnoria quadripunctata*, *Gyge ovalis*, and *Eophreatoicus sp.* remain as unstable as in previous studies, irrespective of the more sophisticated modelling approach applied here. These species are all found to have long branches, and though they seem to loosely group together across analyses, they hold varying relationships with each other

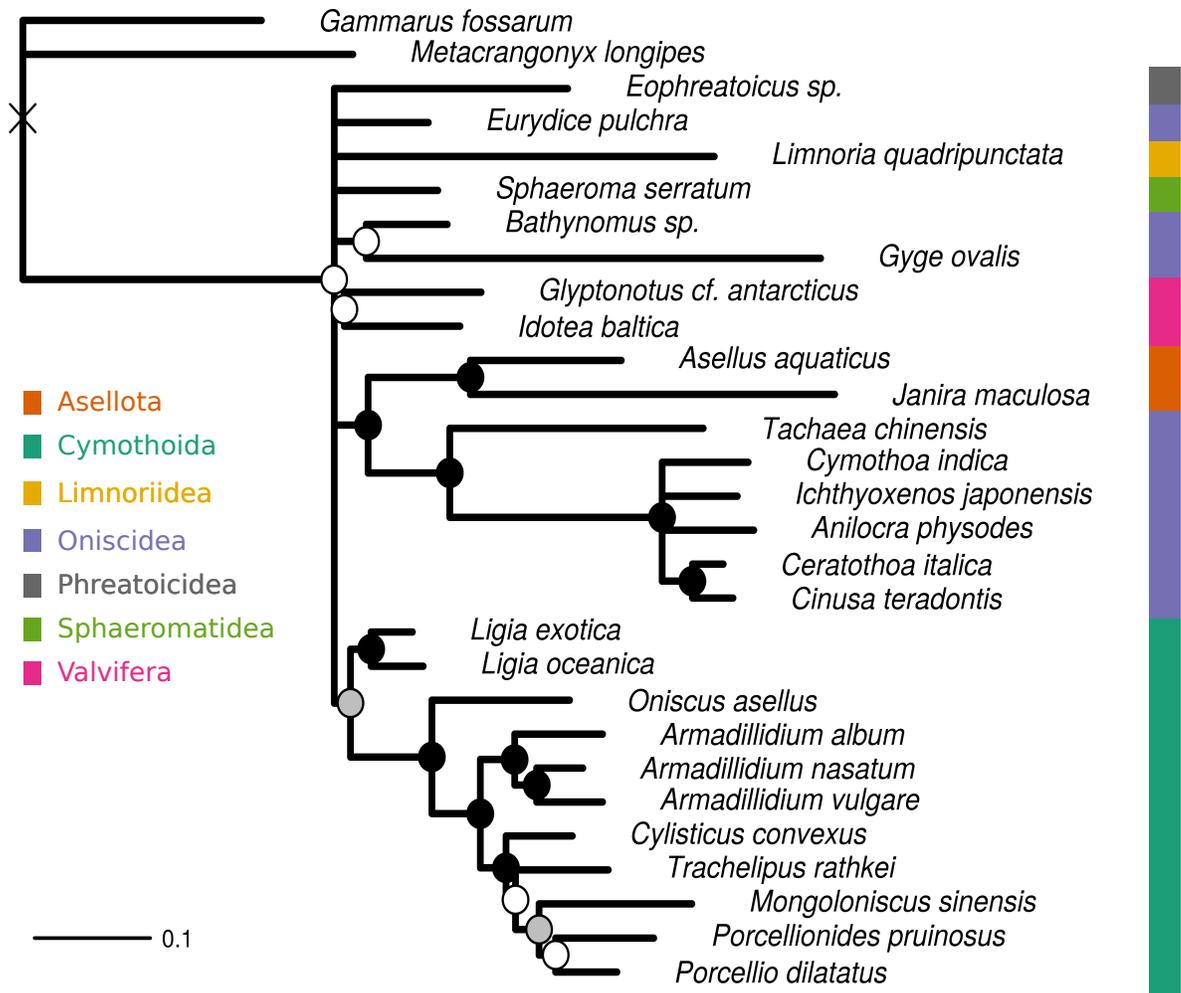


Figure 3.6: Isopod phylogeny estimated with Dayhoff-recoded amino acids, CAT-GTR model. Black circles are Bayesian posterior probabilities (BPP) > 0.9, grey circles are BPP < 0.9 & > 0.7, and white circles are BPP < 0.7.

and to other isopods. In general, nucleotide datasets have *L. quadripunctata* and *G. ovalis* nested within the problem clade, whereas recoded amino acid data place their divergence from other isopods toward the root. These taxa show the same pattern in both site-heterogeneous and partitioned analyses, but resolution of recoded amino acid datasets with the former method are poor and they cannot be separated from a large polytomy. Partitioned analyses of PCG-HOMO support the more traditional placement of *Eophreatoicus* sp. as sister to other isopods, but also confounds that view with *G. ovalis* having an earlier divergence. Such a result is clearly an effect of systematic error — short-tailed isopods did not evolve multiple times from long-tailed, parasitic ancestors.

Despite topological ambiguities, we are able to recover some consistent patterns across analyses (see Appendix for phylogenies for each analysis). The family Cymothoidae is consistently monophyletic. In all our trees cymothoids also group with *Tachaea chinensis*, forming the main Cymothoida clade. However, Cymothoida remains polyphyletic with respect to *Bathynomus* sp., *E. pulchra*, and *Gyge ovalis*. Notably, these three species are the only members of Cymothoida in our dataset that do not exhibit a negative GC skew. Given that fact, the position of Cymothoidae + Corallanidae seems to be dependent on the level of compositional heterogeneity in the alignment rather than the modelling approach: they are always drawn toward the outgroup when nucleotide or non-recoded amino acid data are used. Monophyly of Asellota is consistent across datasets, as is Valvifera except for CATGTR + PCG-HOMO. Asellota is found sister to the problem clade with most datasets, except for the two most homogeneous recoded amino acid alignments, where they group with the negatively GC skewed cymothoidan species.

Table 3.6: Saturation (C-value) and composition homogeneity tests of concatenated alignments. ND4L and ND6 alignments were too small to calculate C-values.  $\chi^2$  are the p-values for  $\chi^2$  tests calculated directly from the data, P4 are p-values estimated from a null generated from 1,000 simulations. 123 = all codon positions; 3RY = third codon position RY-coded; 3X = third codon position removed; AA = amino acid translation; DAY = Dayhoff-recoded amino acid translation. Asterisks denote compositionally homogenous genes.

| Gene | C-value |       |       |      | $\chi^2$ |       |       |      |       |       |       |       | P4   |       |       |      |      |       |       |       |
|------|---------|-------|-------|------|----------|-------|-------|------|-------|-------|-------|-------|------|-------|-------|------|------|-------|-------|-------|
|      | 123     | (1    | 2     | 3)   | 123      | (1    | 2     | 3)   | 3RY   | 3X    | AA    | DAY   | 123  | (1    | 2     | 3)   | 3RY  | 3X    | AA    | DAY   |
| ATP6 | 2.64    | 3.88  | 13.57 | 3.00 | 0.00     | 0.00  | 0.27* | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.00  | 0.11* |
| COX1 | 5.13    | 13.01 | 14.18 | 3.48 | 0.00     | 0.04  | 1.00* | 0.00 | 0.22* | 0.22* | 1.00* | 1.00* | 0.00 | 0.00  | 0.23* | 0.00 | 0.00 | 0.00  | 0.04  | 0.24* |
| COX2 | 4.38    | 6.44  | 11.10 | 3.85 | 0.00     | 0.06* | 0.84* | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.04  | 0.68* |       |
| COX3 | 3.62    | 6.49  | 3.86  | 3.77 | 0.00     | 0.11* | 1.00* | 0.00 | 0.16* | 0.16* | 1.00* | 1.00* | 0.00 | 0.00  | 0.41* | 0.00 | 0.00 | 0.10* | 0.39* |       |
| CYTB | 1.22    | 3.70  | 2.91  | 1.97 | 0.00     | 0.00  | 0.27* | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.11* |       |
| ND1  | 4.41    | 8.00  | 6.11  | 3.76 | 0.00     | 0.00  | 0.02  | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.01  |       |
| ND2  | 1.78    | 2.22  | 3.41  | 2.92 | 0.00     | 0.00  | 0.30* | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.32* |       |
| ND3  | 4.27    | 6.29  | 8.81  | 4.20 | 0.00     | 0.00  | 0.95* | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.02* |       |
| ND4L | 5.18    | -     | -     | -    | 0.00     | 0.81* | 0.98* | 0.00 | 0.43* | 0.43* | 1.00* | 1.00* | 0.00 | 0.63* | 0.01  | 0.00 | 0.00 | 0.12* | 0.09* |       |
| ND4  | 3.24    | 4.10  | 5.41  | 3.11 | 0.00     | 0.00  | 0.00  | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.00  |       |
| ND5  | 2.31    | 3.41  | 4.06  | 2.38 | 0.00     | 0.00  | 0.00  | 0.00 | 0.00  | 0.00  | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.00  | 0.02  |       |
| ND6  | 5.00    | -     | -     | -    | 0.00     | 0.11* | 1.00* | 0.00 | 0.28* | 0.28* | 1.00* | 1.00* | 0.00 | 0.00  | 0.00  | 0.00 | 0.00 | 0.04  | 0.00  |       |
| 12S  |         | 1.51  |       |      |          |       |       | 0.00 |       |       |       |       |      |       |       | 0.00 |      |       |       |       |
| 16S  |         | 1.29  |       |      |          |       |       | 0.00 |       |       |       |       |      |       |       | 0.00 |      |       |       |       |
| 18S  |         | 3.35  |       |      |          |       |       | 0.00 |       |       |       |       |      |       |       | 0.00 |      |       |       |       |

### 3.4 Discussion

Long branch attraction (LBA) negatively impacts phylogenetic inferences by producing false relationships between unrelated species sharing similar evolutionary rates or base compositions (Foster and Hickey, 1999). Mitochondrial DNA is particularly susceptible to the accrual of biases, including compositional heterogeneity, due to elevated mutation rates and directional mutation pressure (Foster and Hickey, 1999; Castellana et al., 2011; Talavera and Vila, 2011; Uribe et al., 2019). LBA effects are evident in previous phylogenetic studies of isopod higher relationships where unrelated taxa have been wrongly grouped, derived taxa are recovered as sister to remaining isopods, and widely accepted clades appear polyphyletic. Here, we focussed on compositional heterogeneities, quan-

tified this bias across isopod mitogenomes and between isopod species, and sought to account for error by various data treatments and application of both partitioned and site-heterogeneous phylogenetic reconstruction methods. Worryingly, phylogenetic analyses recovered mostly unique topologies between different combinations of dataset and method. We obtain some support for a monophyletic Oniscidea but systematic errors obviously persist, highlighting the limitations of mitogenome data, analysed with currently available inference methods, to resolve deep divergences in isopod phylogeny.

The clearest evidence that compositional bias directly affects our tree topologies is the correlation between negative strand bias (i.e.  $C\% < G\%$ ) and the position of Asellota, Cymothoidae, and Cirolanidae: they are most often recovered as sister to other isopods and often form a clade, but from what we know about isopod phylogeny Asellota and Cymothoida should be distantly related (Asellota are short-tailed whereas cymothoidans are long-tailed). Strand compositional skews have been associated with strand-biased deamination, which is most prominent during replication and transcription (Hassanin et al., 2005; Wei et al., 2010). Spontaneous deamination is a common mutational mechanism in DNA, but during replication and transcription, when strands are separated, the lagging strand becomes temporarily single-stranded, which increases exposure to deamination damage (Lindahl, 1993; Marín and Xia, 2008). Furthermore, adenine and cytosine bases naturally carry a higher likelihood of deamination than their compliments but also differ from each other in their mutation rates (Hassanin et al., 2005; Dabney et al., 2013b). It is these combined strand and nucleotide mutational constraints that ultimately lead to strand compositional asymmetry: adenine bases do not occur in equal frequency to thymines, and the number of cytosines does not equal the number of guanines. The consequence for phylogenetic analysis is that common mutation types become rare and rare types become common, and divergent lineages which share the same skew direction will group by virtue of their similar substitution patterns (Hassanin et al., 2005;

Marín and Xia, 2008).

Most pancrustaceans exhibit negative GC-skews on the leading strand, including our amphipod outgroup (Kilpert and Podsiadlowski, 2010; Doublet et al., 2012). This means negative skews likely represent the isopod ancestral condition, with most isopods having acquired a reverse skew (Kilpert and Podsiadlowski, 2006; Kilpert et al., 2012; Shen et al., 2017; Hua et al., 2018). Previously, it has been suggested that global reversal of strand skews is most likely caused by an inversion of the control region (CR), followed by accumulation of substitutions under an opposite mutational constraint (Hassanin et al., 2005; Wei et al., 2010). Certainly, arthropod CRs are characterised by several domains, including small inverted repeats which flank a stem and loop structure (Kilpert and Podsiadlowski, 2006; Doublet et al., 2013); and these features could be amenable to inversions while preserving integrity of replication or translation initiation sites.

We also propose another mechanism to explain reverse skews based on the possibility that all isopods have atypical mitochondrial architectures: CR switching. Under this scenario, replication or transcription initiates preferentially at only one CR copy, perhaps because promoter regions in the reverse orientated dimer cannot easily be recognised by relevant polymerases. Inverse CR sequences, which contain potentially disruptive insertions, have been observed in oniscideans at least suggesting the possibility that only one CR is functional (Doublet et al., 2013). Given the derived position of species with supposedly typical mitochondrial structures (*Sphaeroma serratum* and possibly *Ligia* species; Doublet et al., 2012), their positive skews could be explained by the reversion from dimeric to standard, monomeric 'head-to-tail' forms, in which they necessarily have inherited only one CR. Alternatively, these species could also have atypical structures that have been missed. Our proposal is certainly more plausible if all isopods have atypical mitochondria (i.e. two linear molecules each containing a full complement of genes, fused in a head-to-head configuration), and here we at least provide preliminary

evidence for such mitogenome organisation in another family, Cymothoidae.

Whichever process underlies strand bias reversals, the number and timing of these events is dependent upon which of the conflicting phylogenetic hypotheses we assume for isopods. With the first published isopod mitogenomes, Kilpert and Podsiadlowski (2006) suggested a simple linear process with a switch at the root of Isopoda to positive strand bias. However, as more species' mitogenomes have been sequenced it is clear the picture is more complex, with at least two secondary reversals to negative skews (Kilpert et al., 2012). Lins et al. (2017) and Kilpert et al. (2012) both recovered strand-positive species (*Limnoria quadripunctata* and *Eophreaticoicus* sp. in the sister position to other isopods, requiring a single reversal at the base of Asellota. Alternatively, the original hypothesis of a single event separating positively and negatively skewed mitogenomes is also supported if either Asellota (Yu et al., 2018), or Asellota + Cymothoidae are sister (Hua et al., 2018; Zou et al., 2018). We tend to recover this latter topology with nucleotide data, but it is clearly an artefact of compositional bias and we can safely surmise that isopod negative strand skews are homoplastic. Our PPA results show that applying methods to reduce compositional heterogeneity increases model fit, and that CATGTR + PCG-HOMO does seem to attenuate some LBA artefacts. While we fail to completely resolve the topology in our CATGTR + PCG-HOMO analysis, partitioned and recoded amino acids find cymothoids are not in a sister position, as does past morphological and nuclear evidence (Brusca and Wilson, 1991; Dreyer and Wägele, 2002; Wilson, 2009).

Information from strand reversal events, we suggest, is most consistent with the hypothesis that Asellota are sister to other isopods — a position recovered with 18S (Dreyer and Wägele, 2002) and combined 18S and morphological data (Dreyer and Wägele, 2001; Wilson, 2009). The most often proposed alternative, Phreaticoidea sister, would require at least three skew reversal events in isopods as opposed to two (Wägele, 1989; Brusca and Wilson, 1991). In our analyses, addition of 18S does little to influence

the recovered topology. Large insertions have been observed within 18S and it is possible these impacted our ability to align homologous sites (Spears et al., 2005; Wilson, 2009). In general, the addition of RNAs seems to alter the location of Phreatoicoidea and Asellota to a terminal position, sister to the problem clade. One interesting finding between these two groups is that Asellota cluster with Phreatoicoidea using amino acid translations, but with Dayhoff-recoded datasets are sister to Cymothoidae + Corallanidae — a position most similar to that recovered from nucleotide datasets. This implies that these species share a more similar nucleotide composition than to other isopod species (and is evidenced by their GC-skew), but that the amino acids they encode are quite different. This seems counter to the evidence, and intuition, that nonsynonymous mutations are generally purged by strong purifying selection to maintain essential mitochondrial function (Castellana et al., 2011; Palozzi et al., 2018). However, given the Asellota + (Cymothoidae + Corallanidae) clade obtained from Dayhoff-recoded datasets, these mutations are probably maintained because they result in amino acid replacements with similar physio-chemical properties, and therefore have limited impacts on protein function.

Since most isopods exhibit the same strand bias, we think it unlikely that strand mutational constraints alone can account for the widespread topological discordance we observe between data types, or the instability of particular taxa in mitochondrial trees. As outlined, compositionally similar species certainly cluster in our analyses, but long-branch taxa like *Gyge ovalis* and *Limnoria quadripunctata* also group together despite their supposedly distant relationships, or they are drawn toward the outgroup. Several rogue taxa, including *Eurydice pulchra* and *Bathynomus* sp., also show instability, which is often a mark of conflicting data. Positive selection is perhaps another complimentary explanation for compositional and rate convergence, and it is increasingly recognised to influence mitochondrial genomes (Botero-Castro et al., 2018). Adaptive processes could

also explain our observation that amino acid replacements seem to exceed synonymous substitutions in some lineages. For most animals, life history traits such as longevity, generation time, and metabolic rate can explain mitochondrial substitution rate variation between lineages, but there is no correlation in isopods (Saclier et al., 2018). However, cymothoidans have adopted parasitic, including haematophagous, diets, species of Asellota are found in fresh water, and oniscideans evolved terrestriality. It seems probable that adaptation to such disparate demands has been accompanied by at least short periods of positive selection on mitochondrial genes (Botero-Castro et al., 2018). Therefore, adaptation to different environments or physiological regimes probably contributes something toward convergent mutational rate shifts that manifest LBA artefacts (Bergsten, 2005; Castellana et al., 2011). Additionally, gene rearrangements have been associated with faster evolutionary rates, which ultimately lead to long branches (Xu et al., 2006). Isopod mitogenomes are well known for their labile arrangements, and those branches with the most inferred translocation events (Asellota, *Eurydice pulchra*, Oniscidea) are also the ones we find to be the most unstable or have higher evolutionary rates (Kilpert et al., 2012).

### **3.5 Conclusion**

Convergent substitutions are the primary source of phylogenetic reconstruction errors. To distinguish synapomorphic from homoplastic substitutions it is, therefore, important to apply suitable models to empirical data that can account for the processes underlying increased convergent substitutions. Alternatively, it is possible to recode data to ensure a better fit with available models. Our best dataset and model combination, site-heterogeneous with recoded amino acids, was able to attenuate some of the extreme compositional bias present in isopod mitogenome data, and for the first time molecular

data provides at least some evidence of a monophyletic Oniscidea. Yet, we still find relationships contradictory to established evolutionary concepts in Isopoda, such as a polyphyletic Cymothoida, and as stated by other researchers, we would also caution the use of mitochondrial sequences to resolve isopod higher relationships (Wetzer, 2002). The modelling approach we used here has facilitated resolution of several phylogenetic challenges for invertebrates, including patellogastropod relationships (Uribe et al., 2019), the relationships at the root of the animal tree (Feuda et al., 2017), heteropteran subordinal relationships (Liu et al., 2018), and the position of chelicerates (Lozano-Fernandez et al., 2019). Collectively we find that mutational constraints, rearrangements, combined with atypical mitogenome architectures, and, possibly, positive selection, generates too much homoplastic noise for mitogenomes to serve as a reliable tool to reconstruct the evolutionary history across these ubiquitous crustaceans. Nevertheless, more comforting indications arise from observing the relationships within families, which appear consistent, irrespective of the suite of analytical approaches used. This may provide a platform for mitogenomes to serve as useful phylogenetic indicators at finer taxonomic scales (i.e. family), and hence resolve evolutionary trajectories in some of the most intriguing isopod lineages, such as the parasitic Cymothoidae.

**Ancient DNA analysis supports multiple transitions between parasitic strategies in the evolution of cymothoid isopods, and a single colonisation of freshwater in South America.**

## 4.1 Introduction

One of the single most extraordinary examples of host-parasite co-evolution is shown by cymothoid isopods, which are obligate parasites of fishes, including many commercially important species (Smit et al., 2014). Cymothoids exhibit striking parasitic strategies: for example, some species popularly called "tongue-biters" supplant their host's tongue (Brusca and Gilligan, 1983). As well as the mouth, cymothoids use hook-like structures called dactyli, which terminate their walking limbs (pereopods), to secure themselves within the gill chamber, attach externally to host dermal tissue, or burrow into host body-cavities (Smit et al., 2014). Each species is typically specialised to use only one of these strategies. Most cymothoid species are also highly host specific — the monotypic *Cinusa tetrodontis*, for example, has only been found in the mouth of *Amblyrhynchotes honckenii* (Hadfield et al., 2010). There are exceptions, with some species capable of infesting hosts from several different families, such as *Nerolica orbigny*, which parasitises species of Mullidae, Sparidae, and Serranidae, among others (Bruce, 1987; Kayış and Ceylan, 2011). Similarly, there is considerable biogeographic variation among cymothoid species, some having a restricted distribution and others found to inhabit several of the world's oceans (Smit et al., 2014). Cymothoids are, therefore, a tantalising co-evolutionary system with potential for fruitful insights into a range of fundamental evolutionary patterns and processes, including diversification, specialisation, evolutionary constraints and trade-offs, convergence, and the reversibility of evolutionary changes.

Due to their bizarre life-histories, large size relative to their hosts, and their economic impacts in fisheries and aquaculture, cymothoids have been studied since the mid-18<sup>th</sup> century (Linnaeus, 1758). Yet, despite this extensive research history and the fact that the majority of this work has concentrated on alpha-taxonomy, many species boundaries remain unclear and the relationships between cymothoid taxa are uncertain,

impeding insights into cymothoid-host co-evolution. Taxonomic resolution of cymothoids is particularly problematic because species are highly variable, or polymorphic, and intraspecific morphological variation is easily confounded with interspecific differences (Horton, 2000). For example, *Cymothoa eremita* was once considered two distinct species (Hadfield et al., 2013), and *Ceratothoa parva*, a species first described without attachment location or host data (Richardson, 1905) and placed within a mouth-attaching genus, was recently transferred to *Elthusa* — a genus comprised of gill-attaching species (Hadfield et al., 2016a,b). Morphology in pre-adult stages is also highly homogeneous across cymothoids, restricting species descriptions to adult females. Compared with free-living isopods, adult cymothoids are phenotypically less complex (lack setae and colouration), thus providing potentially fewer discriminatory characteristics. Furthermore, cymothoids are highly diverse in comparison to other isopod families, with approximately 400 species across 43 genera (Boyko et al., 2019), meaning there exists extensive variation to untangle.

Attachment location, in particular, has long influenced the classification of cymothoid taxa. In early works, Schioedte and Meinert (1881; 1883) recognised an association between morphology and attachment location and proposed subfamilial divisions along these lines: Anilocridae (external-attaching), Saophridae (gill-attaching), and Cymothoidae (mouth-attaching). Brusca (1981) largely followed this classification when framing the first evolutionary hypothesis for cymothoids, proposing that the burrowing-, skin-, and the mouth- plus gill-attaching genera are subclades which correspond to ecomorphological, adaptive lineages. Bruce (1986, 1987, 1990) also recognised subclades within Cymothoidae but highlighted greater complexity than Brusca's (1981) model, synonymising the subfamily Livonecinae (gill-attaching) with Anilocrinae (skin-attaching), based on brood pouch, cephalon, pleopod, and pleon morphology. Bruce (1990) cautioned that apparent homologies between taxa sharing a particular attachment mode are likely to be

convergent adaptations to similar ecological demands and are not necessarily reflective of shared ancestry. For example, *Nerocila lomatia* is the single species of an otherwise skin-attaching genus to be found in the gill, and converges with gill-attaching species in asymmetrical body shape, and reduced pleonal processes and coxae. Similarly, *Mothocyca ihi* resembles other mouth-attaching species in its symmetry and dorso-ventral vaulting but is actually placed within a genus of mainly gill-attaching species. Bruce (1990) did not speculate as to the directionality of the evolution of attachment, but did acknowledge the potential of skin-attaching taxa to inhabit the gill, based on the proximity of some species attached on the head and around the pectoral fin to the gill opening. Williams Jr. and Bunkley-Williams (1994), while documenting unusual short-term parasite-host associations, also recognised that such associations might alter host use patterns over evolutionary timescales, and were the first to suggest that the ancestral cymothoid might actually have attached in the mouth: they proposed that overcrowding forced cymothoids to explore ecological opportunities outside the relatively safe confines of the buccal cavity.

Phylogenetic analyses of molecular data suggest a complex history of attachment, but provide some support for an 'internal-first' hypothesis (Ketmaier et al., 2008; Jones et al., 2008; Hata et al., 2017). Gill species are recovered as sister to other cymothoids and a linear evolutionary pathway for attachment is unlikely, with multiple transitions between attachment modes. Genera usually form clades in the molecular trees from these studies, with the exception of a polyphyletic *Ceratothoa*, *Elthusa*, and *Nerocila* (Hata et al., 2017), but ecological subclades above the genus level corresponding to attachment location are definitely not supported. Yet, each of these three studies was based on few gene sequences, or were preliminary analyses including few taxa (Ketmaier et al., 2008). There is also some topological incongruity between trees using Cytochrome Oxidase subunit I (COI) and the large subunit ribosomal DNA (16S) (Hata et al., 2017),

including the position of *Mothocya* (found in terminal position in COI trees). Furthermore, representatives of important clades are missing, which are crucial to answering other outstanding questions in cymothoid phylogeny — key among which is unravelling patterns of colonisation of freshwater. Brusca (1981) proposed that the freshwater fauna in South America adapted to riverine conditions, and hosts, in two separate colonisation events based on the differing biogeography and attachment locations of species of *Asotana* and *Braga*, and the skin-burrowing taxa.

A major challenge for cymothoid systematics is sampling. Many species are still only known from one or two specimens, such as *Elthusia nierstraszi* (Hadfield et al., 2016b). Infestation prevalence is highly irregular and dependent upon a number of factors, including: the cymothoid species, geographic location, the host species parasitised, and host size (Bakenhaster et al., 2006; Yamano et al., 2011; Roche et al., 2013; Welicky and Sikkell, 2014). Prevalence is generally low in natural populations (Smit et al., 2014), but in areas of high human disturbance and in aquaculture settings infection rates can reach between 30% – 100% (Horton and Okamura, 2001; de Lima et al., 2005; Salazar-Bozano et al., 2012). Cymothoids are also distributed globally with the highest diversity in the tropics (Smit et al., 2014). This mean taxonomic representation (essential for accurate phylogeny reconstruction), is difficult to achieve by sampling in the wild (Heath et al., 2008). Yet, cymothoid specimens from around the world have been amassed in museums, thus concentrating the representative diversity to a few key collections. Combined with high-throughput sequencing methods, so-called ‘museomics’ affords the opportunity to retrieve genome-scale molecular data from collection material, thus addressing the need for more data in cymothoid phylogenetics. This approach has been successful in a number of groups, for example: herbarium specimens (Zedane et al., 2015), birds (Hung et al., 2013), sharks (Nielsen et al., 2017), mammals (Guschanski et al., 2013), beetles (Sproul and Maddison, 2017), and spiders (Cotoras et al., 2017).

Here, we collected archived cymothoid tissues from specimens housed in institutes around the world. We particularly focus on South American freshwater taxa for which phylogeny has not been studied, and for which there is only a single sequence, from a single species. This is the first study, of which we are aware, to apply ‘museomics’ to crustacean specimens. We extracted DNA and prepared sequencing libraries using protocols designed for highly-degraded, ancient DNA (aDNA) and generated hundreds of millions of sequencing reads with a low-coverage, shotgun sequencing approach. This method is well suited for non-model organisms, and delivers the simplest workflows, as well as, potentially, the lowest costs for the volume of information produced (Ripma et al., 2014). We successfully assembled partial mitogenomes for 34 specimens using a previously constructed, high-coverage, reference mitogenome for *Ceratothoa italica*. Using mitochondrial genes we conducted phylogenetic analyses in combination with other published mitogenomes with the aim of better understanding cymothoid diversification, the evolution of attachment, and colonisation of freshwater habitats. In addition, we examined factors that might affect the DNA yield recovered from our museum tissues, and assessed the metagenomic content of our DNA extracts by assigning sequencing reads to taxa that might represent sources of contamination.

## 4.2 Materials and Methods

### Sampling

We collected 118 cymothoid specimens from the Natural History Museum, London (NHM), the Muséum National d’Histoire Naturelle, Paris (MNHN), the Invertebrate Zoology Collection of the National Museum of Natural History, Washington D.C. (USNM), the Instituto Nacional de Pesquisas da Amazônia, Manaus (INPA), North West University, Potchefstroom (NWU), and the Universidade Federal do Pará, Belém (UFPA) (Table 4.1). Where

we could not identify specimens to species, we used the existing determination. We also updated all species names to those accepted in the World Register of Marine Species (WoRMS; Horton et al., 2019). For whole adult and pre-adult specimens, we removed pereopods 3 - 6 from one side, since these are taxonomically the least informative. Whole juvenile, or manca, specimens were taken only when there were several in a single lot. Each tissue sample was rinsed with sterile water and then placed in 70% ethanol until DNA extraction.

Table 4.1: Specimen information including species name, institution accession or record number, sex of specimen, host species name, year the specimen was collected, and location it was collected. Sex: F = Female, M = Male, j. = juvenile, t. = transitional stage.

| Species  | Acc./Record  | Sex | Host                                | Year Collected | Location                |
|--|--------------|-----|-------------------------------------|----------------|-------------------------|
| Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus |              |     |                                     |                |                         |
| <i>Anphira branchialis</i>                                 | INPA-09      | F   | <i>Metynnis lippincottianus</i>     | 2013           | Xingu, Brazil           |
| <i>Anphira branchialis</i>                                 | INPA-10      | M   | <i>Metynnis lippincottianus</i>     | 2013           | Xingu, Brazil           |
| <i>Anphira branchialis</i>                                 | INPA-11      | F   | <i>Metynnis lippincottianus</i>     | 2014           | Xingu, Brazil           |
| <i>Anphira junki</i>                                       | INPA-935     | M   | <i>Triportheus albus</i>            | 2000           | Lago Catalao, Brazil    |
| <i>Anphira xinguensis</i>                                  | INPA-12      | F   | <i>Ossubtus xinguensis</i>          | Unknown        | Xingu, Brazil           |
| <i>Asotana magnifica</i>                                   | INPA-942     | F   | <i>Serrasalmus</i> sp.              | 1987           | Rio Mucajai, Brazil     |
| <i>Braga cichlae</i>                                       | INPA-950     | F   | <i>Cichla temensis</i>              | 1999           | Rio Negro, Brazil       |
| <i>Braga cichlae</i>                                       | INPA-955     | F   | <i>Cichla temensis</i>              | 1986           | Sem Localizacao, Brazil |
| <i>Braga nasuta</i>  | INPA-963     | F   | <i>Cichla temensis</i>              | 1986           | Rio Uatuma, Brazil      |
| <i>Braga patagonica</i>                                    | INPA-959     | F   | <i>Pygocentrus nattereri</i>        | 2002           | Rio Araguaia, Brazil    |
| Gen. nov. et sp. nov.                                      | INPA-13      | F   | <i>Metynnis maculatus</i>           | 1983           | Rio Guaporé, Brazil     |
| <i>Riggia nana</i>   | INPA-1026    | F   | <i>Steatogenys elegans</i>          | 1992           | Toruma, Brazil          |
| <i>Vanamea symmetrica</i>                                  | INPA-1046    | F   | <i>Serrasalmus elongatus</i>        | 2000           | Rio Solimoes, Brazil    |
| Muséum National d'Histoire Naturelle (MNHN), Paris         |              |     |                                     |                |                         |
| <i>Anilocra frontalis</i>                                  | IU-2016-6004 | M   | Unknown                             | 1920           | Brittany, France        |
| <i>Anilocra gigantea</i>                                   | IU-2009-1710 | t.  | <i>Etelis carbunculus</i>           | 2005           | Bourail, New Caledonia  |
| <i>Anilocra longicauda</i>                                 | IU-2016-6005 | F   | <i>Pristipomoides angyrogrammus</i> | 2008           | Dunbea, New Caledonia   |
| <i>Catoessa boscii</i>                                     | IU-2009-3211 | j.  | <i>Carangoides malabaricus</i>      | 2009           | Nanga Pattinam, India   |
| <i>Catoessa boscii</i>                                     | IU-2016-6002 | F   | <i>Carangoides malabaricus</i>      | 2009           | Panangipettai, India    |
| <i>Ceratothoa carinata</i>                                 | IU-2009-3342 | F   | <i>Decapterus macarellus</i>        | 2006           | Noumea, New Caledonia   |

|                                  |               |    |                                |      |                                  |
|----------------------------------|---------------|----|--------------------------------|------|----------------------------------|
| <i>Ceratothoa oestroides</i>     | IU-2013-1875  | F  | <i>Boops Boops</i>             | 2015 | Turkey                           |
| <i>Ceratothoa oxyrrhynchaena</i> | IU-2009-3320  | F  | Unknown                        | 2005 | Iles Chesterfield, New Caledonia |
| <i>Ceratothoa retusa</i>         | IU-2009-3340  | j. | <i>Hemiramphus far</i>         | 2010 | Ilot Canard, New Caledonia       |
| <i>Cymothoa eremita</i>          | IU-2016-6008  | F  | <i>Sagocentrum rubrum</i>      | 2010 | Marche Poissons, New Caledonia   |
| <i>Cymothoa pulchrum</i>         | IU-2009-3353  | M  | Unknown                        | 2003 | Noumea, New Caledonia            |
| <i>Elthusa</i> n. sp.            | IU-2011-8002  | t. | Unknown                        | 2011 | Kouakoue Canyon, New Caledonia   |
| <i>Elthusa</i> n. sp.            | IU-2015-214   | F  | Unknown                        | 2014 | New Britain, Papua New Guinea    |
| <i>Elthusa</i> sp.               | IU-2011-7619  | t. | Unknown                        | 2011 | Au large de Theo, New Caledonia  |
| <i>Elthusa</i> sp.               | IU-2016-6025  | F  | Unknown                        | 2009 | Mozambique                       |
| <i>Elthusa arnoglossi</i>        | IU-2011-7709  | F  | Unknown                        | 2011 | New Caledonia                    |
| <i>Elthusa arnoglossi</i>        | IU-2016-6016  | F  | <i>Arnoglossius</i> sp.        | 2005 | Iles Chesterfield, New Caledonia |
| <i>Elthusa emarginata</i>        | IU-2016-6007  | F  | <i>Panapineus hirtacanthus</i> | 1975 | Ambon Island, Indonesia          |
| <i>Elthusa epinepheli</i>        | IU-2016-6009  | F  | <i>Epinephelus howlandi</i>    | 2006 | Noumea, New Caledonia            |
| <i>Elthusa nierstraszi</i>       | IU-2011-7583  | F  | Lophiidae                      | 2011 | New Caledonia                    |
| <i>Elthusa nierstraszi</i>       | IU-2016-6019  | F  | Lophiidae                      | 2011 | Yate, New Caledonia              |
| <i>Elthusa nierstraszi</i>       | IU-2016-6026  | F  | Lophiidae                      | 2014 | New Guinea                       |
| <i>Elthusa propinqua</i>         | IU-2009-3361  | F  | Unknown                        | 2007 | Iles Salomon, Salomonboa         |
| <i>Elthusa raynaudii</i>         | IU-2016-6011  | t. | Unknown                        | 1972 | Saint Paul, Indian Ocean         |
| <i>Joryma hilsae</i>             | IU-2009-3216  | F  | <i>Sardinella</i> sp.          | 2009 | Muttam, India                    |
| <i>Mothocya taurica</i>          | IU-2013-18751 | F  | <i>Alosa immaculata</i>        | 2015 | Turkey                           |
| <i>Nerocila arres</i>            | IU-2009-1934  | F  | <i>Nemipterus japonicus</i>    | 2011 | Nagapattinam, India              |
| <i>Nerocila orbigny</i>          | IU-2013-18755 | F  | <i>Phycis blennoides</i>       | 2015 | Turkey                           |
| <i>Nerocila orbigny</i>          | IU-2016-6020  | F  | Unknown                        | 1969 | Mediterranean Sea                |
| <i>Nerocila poruvae</i>          | IU-2009-3215  | F  | <i>Thryssa mystax</i>          | 2009 | Vedaranyan, India                |
| <i>Nerocila sigani</i>           | IU-2009-1935  | F  | <i>Siganus aramin</i>          | 2011 | Mudasolodai, India               |
| <i>Norileca indica</i>           | IU-2016-6015  | M  | <i>Selar crumenophthalmus</i>  | 2004 | Sainte Marie, New Caledonia      |
| <i>Olencira praegustator</i>     | IU-2016-6014  | F  | <i>Brevoortia tyzannus</i>     | 2004 | Winyah Bay, South Carolina       |
| <i>Ryukyua circularis</i>        | IU-2009-3358  | F  | <i>Chirocentrus dorab</i>      | 2010 | Marche, New Caledonia            |
| Undetermined                     | IU-2015-263   | M  | Unknown                        | 2014 | New Britain, Papua New Guinea    |
| Undetermined                     | IU-2016-6017  | M  | Unknown                        | 1988 | Iles Chesterfield, New Caledonia |
| Undetermined                     | IU-2016-6018  | F  | Unknown                        | 1988 | Iles Chesterfield, New Caledonia |
| Undetermined                     | IU-2016-6022  | F  | Unknown                        | 2009 | Mozambique                       |
| Undetermined                     | IU-2016-6024  | M  | Unknown                        | 2009 | Mozambique                       |

Natural History Museum (NHM), London

|                                 |               |   |                           |      |                         |
|---------------------------------|---------------|---|---------------------------|------|-------------------------|
| <i>Anilocra capensis</i>        | N-1963:310.12 | F | <i>Priacanthus</i> sp.    | 1952 | West Africa             |
| <i>Anilocra leptostoma</i>      | N-1999:895    | F | <i>Harengula zumasi</i>   | 1952 | Yellow Sea, South Korea |
| <i>Ceratothoa guadichaudii</i>  | N-12          | F | <i>Salmo salar</i>        | 1994 | Chilco, Chile           |
| <i>Ceratothoa imbricata</i>     | N-1981:443.5  | F | <i>Trachurus</i> sp.      | 1981 | Dunedin, New Zealand    |
| <i>Ceratothoa steindachneri</i> | N-2000:182.3  | F | <i>Echiichthys vipera</i> | 1997 | Cornwall, UK            |
| <i>Cymothoa eremita</i>         | N-1980:147.3  | F | <i>Caranx sansum</i>      | 1980 | Jaffna, Sri Lanka       |

|  |               |   |                                      |         |                             |
|--|---------------|---|--------------------------------------|---------|-----------------------------|
| <i>Cymothoa indica</i>                     | N-1983:52.2   | F | <i>Sphyræna jello</i>                | 1983    | Tamil Nadu, India           |
| <i>Glossobius impressus</i>                | N-10          | F | <i>Exocetidae</i> sp.                | 1986    | East Atlantic, Sierra Leone |
| <i>Lironeca</i> sp.                        | N-1989:978.2  | F | Unknown                              | 1989    | Gough Island, S. Atlantic   |
| <i>Elthusa raynaudii</i>                   | N-1986:429.1  | F | Unknown                              | 1986    | Tasmania, Australia         |
| <i>Mothocya halei</i>                      | N-1981:446.15 | F | <i>Hyporamphus melanochir</i>        | 1980    | Adelaide, Australia         |
| <i>Mothocya melanostica</i>                | N-1985:245.2  | F | <i>Kuhlia mugil</i>                  | 1985    | Lord Howe Island, Australia |
| <i>Nerocila bivittata</i>                  | N-1986:428.2  | F | Unknown                              | 1956    | Naples, Italy               |
| <i>Nerocila orbignyi</i>                   | N-1986:435.4  | F | <i>Tilapia galilea</i>               | 1959    | Red Sea, Egypt              |
| <i>Nerocila phaeopleura</i>                | N-1986:424.2  | F | <i>Ilisha melastoma</i>              | 1977    | Kakinada, India             |
| North West University (NWU), Potchefstroom |               |   |                                      |         |                             |
| <i>Anilocra capensis</i>                   | NWU-10S       | F | <i>Cheimerius nufar</i>              | 2015    | South Africa                |
| <i>Mothocya renardi</i>                    | NWU-221       | F | <i>Strongylura leiura</i>            | 2014    | Tamil Nadu, India           |
| <i>Nerocila poruvæ</i>                     | NWU-223       | F | <i>Thryssa mystax</i>                | 2012    | South East India            |
| Universidade Federal do Pará (UFPA), Belém |               |   |                                      |         |                             |
| <i>Artystone minima</i>                    | B-NAN1        | F | <i>Nannostomus beckfordi</i>         | 2015    | Jeju, Brazil                |
| <i>Artystone trysibia</i>                  | B-TRY1        | M | <i>Crenicichla</i> cf. <i>regani</i> | 2015    | Jeju, Brazil                |
| <i>Asotana</i> sp.                         | B-STEFF       | F | <i>Anostomidae</i>                   | 2015    | Ourem, Brazil               |
| <i>Braga patagonica</i>                    | B-T01         | F | <i>Pygocentrus nattereri</i>         | 2016    | Lago do Tucuruí, Brazil     |
| <i>Braga patagonica</i>                    | B-T02         | F | <i>Pygocentrus nattereri</i>         | 2016    | Lago do Tucuruí, Brazil     |
| <i>Braga patagonica</i>                    | B-T03         | F | <i>Pygocentrus nattereri</i>         | 2016    | Lago do Tucuruí, Brazil     |
| <i>Braga patagonica</i>                    | B-T04         | F | <i>Pygocentrus nattereri</i>         | 2016    | Lago do Tucuruí, Brazil     |
| <i>Cymothoa</i> sp.                        | B-CAJ024      | M | <i>Serrasalmus rhombeus</i>          | 2015    | Pará, Brazil                |
| <i>Cymothoa</i> sp.                        | B-PP3B        | F | Unknown                              | 2015    | Pará, Brazil                |
| <i>Ceratothoa</i> sp.                      | B-RO4B2       | F | Unknown                              | 2015    | Pará, Brazil                |
| National Museum of Natural History (USNM)  |               |   |                                      |         |                             |
| <i>Agarna cumulus</i>                      | USNM-1278047  | F | <i>Acanthurus chirurgus</i>          | 1976    | Puerto Rico                 |
| <i>Anilocra</i> sp.                        | USNM-231183   | F | Unknown                              | 1927    | Mekong River, Vietnam       |
| <i>Anilocra</i> sp.                        | USNM-1278052  | F | Unknown                              | 1953    | South Carolina, USA         |
| <i>Anilocra alloceraea</i>                 | USNM-1278048  | F | Unknown                              | 1967    | Singapore                   |
| <i>Anilocra ankistra</i>                   | USNM-1278049  | F | Unknown                              | 1964    | Somalia                     |
| <i>Anilocra capensis</i>                   | USNM-1278050  | F | Unknown                              | 1911    | Namibia                     |
| <i>Anilocra capensis</i>                   | USNM-1278054  | F | Unknown                              | Unknown | Unknown                     |
| <i>Anilocra gigantea</i>                   | USNM-1278051  | F | Unknown                              | 1986    | Hawaii, USA                 |
| <i>Anilocra myripristis</i>                | USNM-1384012  | F | <i>Myripristis jacobus</i>           | 1998    | Cayman Islands              |
| <i>Anilocra physodes</i>                   | USNM-1278055  | F | <i>Sparisoma cretense</i>            | 1993    | Grand Canaria               |
| <i>Anphira branchialis</i>                 | USNM-1278059  | F | <i>Serrasalmus rhombeus</i>          | 1993    | Manaus, Brazil              |
| <i>Artystone trysibia</i>                  | USNM-1278057  | F | Unknown                              | 1977    | Ecuador                     |
| <i>Braga</i> sp.                           | USNM-127083   | F | Unknown                              | 1965    | Rio Lujan, Argentina        |
| <i>Braga patagonica</i>                    | USNM-1278061  | F | <i>Serrasalmus rhombeus</i>          | 1993    | Manaus, Brazil              |
| <i>Ceratothoa</i> sp.                      | USNM-1199006  | F | <i>Trachurus</i> sp.                 | 1983    | Okinawa, Japan              |

---

|                                |              |   |                                |         |                            |
|--------------------------------|--------------|---|--------------------------------|---------|----------------------------|
| <i>Ceratothoa</i> sp.          | USNM-1278068 | F | Elasmobranchii                 | 1985    | Okinawa, Japan             |
| <i>Ceratothoa</i> sp.          | USNM-1278069 | F | Unknown                        | Unknown | Hawaii, USA                |
| <i>Ceratothoa guadichaudii</i> | USNM-1278066 | F | <i>Scomber japonicus</i>       | 1955    | Galapagos, Ecuador         |
| <i>Ceratothoa guadichaudii</i> | USNM-1278067 | F | <i>Oncorhynchus kisutch</i>    | 1985    | Valparaiso, Chile          |
| <i>Cymothoa</i> sp.            | USNM-1393520 | F | <i>Orthopristis ruber</i>      | 1999    | Venezuela                  |
| <i>Cymothoa bychowskyi</i>     | USNM-1198988 | F | <i>Fistularia commarosini</i>  | 1985    | Okinawa, Japan             |
| <i>Cymothoa eremita</i>        | USNM-1278065 | F | Unknown                        | 1991    | Jeddah, Saudi Arabia       |
| <i>Cymothoa excisa</i>         | USNM-1278064 | F | <i>Ocyurus chrysurus</i>       | 1993    | Tobago                     |
| <i>Cymothoa exigua</i>         | USNM-1383047 | F | <i>Lutjanus colorado</i>       | 1990    | Costa Rica                 |
| <i>Cymothoa oestrum</i>        | USNM-1278062 | F | <i>Caranx crysos</i>           | 2003    | El Salvador                |
| <i>Elthusa</i> sp.             | USNM-1106195 | F | Unknown                        | 1990    | Eastern Pacific Ocean      |
| <i>Elthusa raynaudii</i>       | USNM-1278070 | F | <i>Lutjanus kasmiri</i>        | 1997    | Guam, Pacific Ocean        |
| <i>Elthusa vulgaris</i>        | USNM-1278071 | F | <i>Synodus luciiceps</i>       | 1977    | California, USA            |
| <i>Glossobius</i> sp.          | USNM-1278072 | F | <i>Hemiramphus far</i>         | 1985    | Okinawa, Japan             |
| <i>Joryma sawah</i>            | USNM-1278073 | F | Unknown                        | 1993    | Kerala, India              |
| <i>Livoneca redmanii</i>       | USNM-1286837 | F | Unknown                        | 2012    | Maryland, USA              |
| <i>Mothocya</i> sp.            | USNM-1278075 | F | <i>Psettias sebas</i>          | 1986    | Niger Delta, Niger         |
| <i>Mothocya omidaptria</i>     | USNM-1278074 | F | Unknown                        | 1984    | Curacao, Caribbean Sea     |
| <i>Nerocila</i> sp.            | USNM-1278076 | F | <i>Rastelliger branchysoma</i> | 1986    | Penang, Malaysia           |
| <i>Olencira praegustator</i>   | USNM-1286928 | F | Unknown                        | 2013    | Maryland, USA              |
| <i>Ourozeuktes monocanthii</i> | USNM-42602   | F | Unknown                        | Unknown | New South Wales, Australia |
| <i>Ryukyua circularis</i>      | USNM-1278078 | F | Unknown                        | 1984    | Phuket, Thailand           |
| <i>Telotha henselii</i>        | USNM-144050  | F | Loricariidae                   | 1965    | Xingu, Brazil              |
| <i>Vanamea symmetrica</i>      | USNM-250452  | F | Unknown                        | 1986    | Manaus, Brazil             |
| Undetermined                   | USNM-1278056 | F | <i>Charax</i> sp.              | 1980    | Rio Santiago, Peru         |

---

## DNA Extraction

DNA extractions from MNHN and NHM specimens were conducted at The University of Manchester's (UoM) ancient biomolecules facility, and those from USNM were extracted at the Paleogenomics Laboratory of the University of California, Santa Cruz (UCSC). Samples from INPA and UFPA were extracted at UFPA, in a UV-irradiated fume cupboard located in an isolated genomics laboratory. We adhered to strict aDNA protocols throughout, including: use of sterile reagents and consumables, surfaces were cleaned with 10% bleach and rinsed with 70% ethanol between specimens, and tools were wiped

with DNA-Away (ThermoFisher) and UV-irradiated. In total, we attempted to extract DNA from 81 of our collected samples and extracted nine samples per batch, including one extraction blank.

DNA was extracted from MNHN, NHM, and USNM specimens following the column-based protocol of Dabney et al. (2013a), with some modifications based on advice from Dr. Mattias Meyer (pers. comm., 17th January 2017). These included the use of a pre-assembled large volume column (described below) as opposed to the custom assembly described, and use of Buffer PB for binding as its capacity is similar to the buffer in Dabney et al. (2013a) but saves considerable time not requiring to be mixed. This method is designed to recover ultra-short DNA fragments, which are typical of ancient DNA, from complex matrices like teeth and bones and has been used, with success, to extract DNA from other arthropods (Tin et al., 2014; Cotoras et al., 2017). We used a Blood & Tissue DNA Kit (Qiagen) for INPA and UFPA specimens, as these had to be conducted in Brazil where some reagents and consumables were not available and could not be imported. For both protocols, we first dried and weighed pereopods to the nearest whole milligram, before grinding them using a pestle and mortar. The Qiagen protocol was then followed with an overnight lysis. For INPA specimens, and B-NAN1, B-TRY1, B-STEFF from UFPA (see Table 4.1), which we knew had previously been fixed in 4% formaldehyde solution, during lysis we heated the lysate to 96°C for 10 minutes to break DNA-DNA and DNA-protein crosslinks (Campos and Gilbert, 2012). For all samples, we performed two separate elution steps, rather than one, to ensure a maximal amount of DNA was recovered from the silica membrane, each time applying 20 $\mu$ L of 10mM Tris-HCl pH 8.0, incubating at room temperature for 10 minutes, and then centrifuging for one minute at maximum speed. Finally, for the formalin specimens, we combined the two eluates and treated the DNA with NEBNext FFPE Repair Mix (New England Biolabs), which repairs damaged bases, single-strand nicks, and gaps, following the manufacturer's instructions.

After treatment, DNA was purified with 2 X HighPrep PCR beads (MagBio), and eluted back in 40 $\mu$ L Tris-HCl pH 8.0.

For the Dabney et al. (2013a) protocol, we transferred ground pereopods to a 50mL tube, containing 1 mL of extraction buffer (0.45 M EDTA, 0.25 mg/mL Proteinase K, pH 8.0), and incubated the samples in a rotator at room temperature for 24 hours. After overnight lysis, we pelleted any remaining debris in a centrifuge and combined the supernatant with 13mL of PB buffer (Qiagen) in a new 50mL tube. This solution, containing the binding buffer and extraction supernatant, was then poured into the extender assembly of a High Pure Viral Nucleic Acid Large Volume Kit (Roche). We spun the binding apparatus in a centrifuge at 3,500rpm for four minutes, rotated the assembly 90°, and spun for an additional two minutes. The column was then removed and centrifuged for one minute at 6,000rpm in a standard 2mL collection tube to dry the silica membrane. We performed two wash steps, each time adding 750 $\mu$ L of PE buffer (Qiagen) to the column and centrifuging at 6,000rpm for one minute, and followed these washes with another drying step at maximum speed for two minutes. Finally, we eluted DNA using the same two step method described above, and quantified concentrations using a Qubit 3.0 fluorometer, high sensitivity assay (ThermoFisher).

## **Library Preparation and Sequencing**

We converted extractions to illumina libraries for 64 specimens at the UoM ancient biomolecules facility using two different protocols. For some specimens up to three libraries were prepared to compensate for poor preservation because potentially few unique DNA molecules might have been recovered per library (Cotoras et al., 2017). The first library preparation method used the Ultralow Input Library Kit (Qiagen), which produces libraries by ligation of full-length, combinatorially dual-indexed, TruSeq (illumina) adapters. Due to the small molecule length of ancient DNA, we did not perform a

fragmentation step, but otherwise followed the manufacturer's protocol exactly. However, we limited preparation of libraries with this kit to DNA extracts with relatively larger fragments. This was because use of full-length adapters for ancient DNA is limited by the difference in size between adapter-dimers (complimentary binding of 5' and 3' adapter oligos during ligation) and the DNA fragments to be converted into sequencing library. It is imperative to remove adapter-dimers before sequencing on illumina instruments to avoid flow cell overclustering and this is normally achieved with bead-based size-selection. Bead clean-ups have a precision threshold, so if the difference in size between dimers and DNA-library is between 50bp to 150bp, many of the target library fragments will also be removed, resulting in low library yields.

Due to the fragment size restriction of the Ultralow kit, we also prepared sequencing libraries based on the blunt-end ligation protocol of Meyer and Kircher (2010), including improvements described in Fortes and Paijmans (2015), and our own modifications. First, we designed a set of 8bp, P5 and P7 primer indices using the python scripts from Meyer and Kircher (2010), available at <https://bioinf.eva.mpg.de/multiplex/>. We then ordered a subset which could be appropriately multiplexed, accounting for illumina's two channel imaging and ensuring a minimum Hamming distance between indices of at least 3bp. The library protocol then proceeded with the following steps: blunt end repair, adapter ligation, qPCR quantification, and indexing PCR. As with DNA extractions and Ultralow kit library preparations, libraries were constructed in batches of eight plus one negative control. Twenty-five microlitres of DNA extract was added to 25 $\mu$ L of end repair master mix containing, in the final reaction, Buffer Tango (1X), dNTPs (100 $\mu$ M each), ATP (1mM), T4 polynucleotide kinase (0.5 U/ $\mu$ L), and T4 DNA polymerase (0.1 U/ $\mu$ L). We incubated the reaction in a thermo-cycler at 25°C for 15 minutes, followed by five minutes at 12°C. We then used a MinElute PCR Purification Kit (Qiagen) with two washes to clean the reaction, before resuspending the end-repaired DNA in 20 $\mu$ L of

10mM Tris-HCl pH 8.0.

Next, we ligated incomplete P5 and P7 adapters, which after a subsequent adapter fill-in step are equivalent to illumina sequencing primers. Small library fragments can be retained with this protocol because the incomplete adapter configuration yields dimers of approximately 50bp and MinElute columns only recover fragments >80bp. The 50 $\mu$ L ligation reaction contained T4 DNA ligase buffer (1X), PEG-4000 (5%), T4 DNA ligase (0.125 U/ $\mu$ L), 0.2 $\mu$ M of each adapter, and 20 $\mu$ L of end-repaired DNA. Ligation reactions were incubated at 22°C for 30 minutes and then cleaned with MinElute columns as before. We then continued with adapter fill-in by combining the eluate with 20 $\mu$ L of master mix comprised of ThermoPol reaction buffer (1X), dNTPs (250 $\mu$ M each), and large fragment Bst polymerase (0.3 U/ $\mu$ L). Bst polymerase was used because, following the fill-in incubation for 20 minutes at 37°C, the enzyme can be inactivated by heating to 80°C for 20 minutes, thus avoiding an additional clean-up step.

We quantified library concentrations with qPCR on a Roche LightCycler 480, to determine the minimum required number of cycles for indexing PCR and, therefore, reduce library over-amplification. The reactions comprised KAPA SYBR FAST qPCR Master Mix (1X), 0.5 $\mu$ M of each IS7 and IS8 primer, 1 $\mu$ L of library diluted 1:10, and ddH<sub>2</sub>O for a total volume of 20 $\mu$ L. Cycling conditions included enzyme activation at 95°C for three minutes, and 40 cycles of 95°C for 15 seconds (denaturation) plus 60°C for 45 seconds (annealing and extension). We used KAPA HiFi Uracil+ Master Mix for indexing PCR since the polymerase is tolerant to uracil-containing DNA — a feature of ancient DNA molecules (Campos and Gilbert, 2012; Dabney et al., 2013b). The master mix (1X in final volume) was combined in 25 $\mu$ L reactions with 200nM of each indexed P5 and P7 primer, 5 $\mu$ L of adapter-ligated library, and 6.5 $\mu$ L of ddH<sub>2</sub>O. After initial denaturation at 98°C for 45 seconds, we used the minimum number of cycles for optimal amplification of each sample (98°C for 15 seconds, 60°C for 30 seconds, 72°C for 30 seconds), followed

by extension at 72°C for one minute. After PCR we cleaned the libraries with a 1X bead clean up and eluted the libraries in 20 $\mu$ L Tris-HCl pH 8.0.

Finally, we measured all library concentrations — from both library preparation protocols — with a Qubit high sensitivity assay, and scored mean fragment lengths on a TapeStation 4200 (Agilent) using HS-D1000 tapes and reagents. We then calculated the molar concentration of each library with the following formula, where 660 g/mol is the molar mass of a single base pair DNA molecule:

$$\text{Molarity} = \left( \frac{\text{concentration (ng/}\mu\text{L)}}{660 \text{ g/mol} \times \text{fragment length}} \right) \times 10^6$$

Forty-four libraries passed quality control, and these were pooled at equimolar concentrations and sequenced across two lanes of a HiSeq 2500 for 200 cycles (2x100bp) using TruSeq SBS V3 chemistry, by Macrogen Inc., South Korea. Library negative controls all yielded DNA concentrations < 100pM when measured with qPCR and, therefore, were not included in sequencing pools.

## Data Processing and Read Mapping

We processed our demultiplexed reads using a custom pipeline optimised for ancient DNA. First, we used `SeqPrep2` (St. John, 2011) to remove adapter contamination, low quality reads, and reads shorter than 25bp since sequences these lengths are difficult to unambiguously map (Schubert et al., 2012). We also made a merged read dataset with a minimum quality score cutoff of 10 for mismatches in the overlap ('-q 10'). Next, we used `PRINSEQ-lite 0.20.04` (Schmieder and Edwards, 2011) to complexity filter both merged and paired-end read sets using the dust score algorithm with a threshold of seven — scores above this can be considered low-complexity (Camacho et al., 2009). Merged and paired read sets from multiple libraries for the same species were then concatenated before mapping. We mapped our complexity filtered read sets to the mitogenome of

*Ceratothoa italica* from this study (Chapter 3), and first hard-masked the long non-coding region that is present in cymothoid mitogenomes, located between 12S and NAD1 (see Chapter 3). This region contains repetitive features, including inversions, which may have caused misalignments. We used `bbmap` 38.95 (Bushnell, 2019) for mapping with the settings 'slow', 'k=10', 'maxindel=200', 'minratio=0.1'.

For each sample, we then used `samtools` 1.8 (Li et al., 2009) to convert the resulting paired and merged `.sam` files to `.bam` format, combine these into a single `.bam` file, remove duplicate reads, add read groups, and to calculate coverage statistics. `GATK` 3.8 (McKenna et al., 2010) was then used to perform a local realignment around indels with the `RealignerTargetCreator` and `IndelRealigner` modules, to reduce the number of mismatching bases, which might erroneously be called as SNPs. We used `mapDamage` 2.0.8 (Ginolhac et al., 2011) to assess damage patterns of our mapped reads and to rescale the per base quality scores accounting for possible damage and misincorporations (option 'rescale'). FASTQ formatted consensus sequences of the rescaled `.bam` files were called in a custom python script (Sibley, 2018: <https://github.com/MullinsLab/simple-consensus-per-read-group>) with a frequency threshold of 0.6 and a minimum depth of 10 for a basecall to be made. This threshold was chosen as a compromise between introducing many completely degenerate bases (N's) and false base-calls (Cotoras et al., 2017). Finally, we used `seqkit` 0.10.2 (Shen et al., 2016) to convert the FASTQ consensus sequences to FASTA format and to produce base composition statistics.

## Phylogenetic Estimation

We extracted Protein Coding Genes (PCGs) and rRNAs from consensus sequences with less than 33% ambiguous bases, by alignment to individual genes of *Ceratothoa italica*, using `mafft` v7.310 (Kato and Standley, 2013) on default settings. We then added

sequences for *Anilocra physodes* and *Cinusa tetrodontis* from this study (Chapter 3), and publicly available sequences for *Cymothoa indica* (MH396438) and *Ichthyoxenos japonensis* (MF419233), as well as two cirolanid outgroup taxa: *Bathynomus* sp. (KU057374) and *Eurydice pulchra* (GU130253). We also made a COI only alignment, in which we combined all unique species sequences available from Genbank, downloaded using the `rentrez` package (Winter, 2017) and the search term: `'("Cymothoidae"[Organism] OR ("Cymothoidae"[Organism] OR "Cymothoidae" [All Fields])) AND (("500"[SLEN] : "750"[SLEN]) AND "COI"[All Fields])'`. Each gene alignment was then realigned with `mafft v7.310` as before, and poorly aligned sites were trimmed with `trimAl v1.2` (Capella-Gutiérrez et al., 2009) using the settings `'-gt 0.9'`, `'-cons 60'`, and `'-st 0.001'`. Due to the high number of ambiguous bases present in our consensus sequences, we also checked and edited each alignment by eye using `AliView 1.18` (Larsson, 2014) to ensure there were no spurious indels or stop codons, before concatenating the gene alignments into a supermatrix with `FASconCAT-G` (Kück and Longo, 2014).

We estimated phylogeny with Maximum Likelihood (ML) in `IQ-TREE 1.6.11` (Nguyen et al., 2014) while concurrently assessing the best model fit for partitions using `ModelFinder` (Kalyaanamoorthy et al., 2017). We evaluated combinations of merged partitions starting with separate rRNAs and codon positions in PCGs, and included rate-free models in our search (option `'-m MFP+MERGE'`), with linked branch lengths (`'-spp'`), and the 'greedy' search algorithm (Lanfear et al., 2016). Branch supports were assessed with 1000 (`'-bb 1000'`) ultrafast bootstrap replicates (Hoang et al., 2017), including an optimisation step using nearest neighbour interchange (`'bnni'`). The final trees were summarised as the majority-rule consensus.

## DNA Yield and Metagenomic Analysis

As well as recovering mitogenome sequences from liquid preserved museum specimens, we were interested in assessing the effects of specimen age and mass of sampled tissue on DNA yield, and on the metagenomic content of sequenced DNA. To assess factors affecting DNA yield, we only included samples from MNHN, NHM, and USNM, which had all been extracted using the same protocol. We excluded samples for which tissue could not be weighed at the time of extraction, and those where recovered DNA concentrations were above 100 ng/ $\mu$ L (two samples — USNM-1286928 and IU-2009-1710 — had concentrations >600ng/ $\mu$ L and 120ng/ $\mu$ L, respectively). We log-transformed our dependent variable (concentration) after inspecting QQplots, and modelled concentration against the continuous covariates mass and age, the institute from where the specimen was collected as a three-level factor, and the interaction between institute and age to account for the fact MNHN had obtained more recent specimens:

$$\log(\text{Concentration}) \sim \text{Institute} + \text{Mass} + \text{Age} + \text{Institute} : \text{Age}$$

We used the R package RRPP (Collyer and Adams, 2018) to perform Analysis of Covariance (ANCOVA) with the Randomised Residual Permutation Procedure (RRPP). One thousand permutations were used to generate empirical sampling distributions for significance testing, and from which effect sizes were estimated as standard deviates (z-scores). To visualise the influence of each predictor we plotted scatterplots for continuous variables and a boxplot for institution with `ggplot2` 3.2.0 (Wickham, 2016).

For metagenomic analyses of our merged reads, we wrote a custom R script to emulate the BLAST based analyses implemented in MEGAN (Huson et al., 2007), by incorporating `blastn` 2.6.0 (Camacho et al., 2009), and the packages `taxonomizr` 0.5.3 (Sherrill-Mix, 2019) and `ggplot2`. First, we only included merged read sets and, for each specimen,

filtered out reads less than 150bp in length. We then used `blastn` to align reads against the whole NCBI non-redundant database using a word size of 11 to account for smaller fragment length, allowing a maximum number of five hits from the same species, and returning a maximum of five hits (options `'-max_hsp 5'`, `'-max_target_seqs 10'`, and `'-word_size 11'`). The results were filtered to include hits with an e-value less than 0.0001, and an identity of at least 70%. For each hit passing this filter, for each read we returned its complete taxonomy using `taxonomizr`, and where a single read produced matches to several taxa, assigned that read to the least common ancestor (function `'condenseTaxa'`). Finally, we calculated, and plotted, the relative proportions of reads assigned to taxonomic groups of interest: the class containing our target species, as well as groups representing potential sources of contamination, including: bacteria, fishes, fungi, and mammals.

## 4.3 Results

### Sequencing and Mapping

We successfully converted DNA extracts into sequencing libraries for 44 specimens. Between them, we generated 922,646,760 raw, paired-end sequencing reads (Table 4.2). Thirty-four specimens passed our cut-off of less than 33% ambiguous bases — above this and alignments became problematic. Of the remaining samples, the average number percentage of ambiguous bases was 11%, and these samples all had high coverage with an average of 58.6X. Five samples had coverage greater than 100X, and three had less than 700 ambiguous sites. Comparing the percentage of reads mapped against the raw reads generated per sample, the mean was 0.07% with the lowest being 0.003% (USNM-1383047) and the highest 0.39% (1999:895). This conversion of sequenced mapped reads is comparable if not slightly lower than similar studies (Cotoras et al., 2017). NHM and INPA specimens produced the lowest DNA yields and also the shortest

merged read lengths.

Table 4.2: Read, filtering, and mapping statistics per sample. Avg. Length is average merged read length. -/N bases and -/N % refer to missing or ambiguous bases.

| ID          | Raw      | Filtered |         | Mapping      |             |          |           |
|-------------|----------|----------|---------|--------------|-------------|----------|-----------|
|             |          | Merged   | Paired  | Reads Mapped | Avg. Length | Coverage | -/N bases |
| NWU-10S     | 27039644 | 6759911  | 3759249 | 8109         | 125.4       | 52.00    | 1761      |
| INPA-13     | 15091516 | 3772879  | 183605  | 13934        | 86          | 68.00    | 1811      |
| INPA-02     | 14792700 | 3698175  | 150366  | 18070        | 69.8        | 76.00    | 1780      |
| INPA-08     | 28675888 | 7168972  | 111185  | 25472        | 72.9        | 118.00   | 1849      |
| INPA-12     | 15716812 | 3929203  | 125567  | 15729        | 60.5        | 64.00    | 1778      |
| IU20093211  | 6781416  | 1695354  | 107807  | 7773         | 142         | 34.00    | 1201      |
| IU20093340  | 19087924 | 4771981  | 1946003 | 10637        | 111.1       | 56.00    | 252       |
| IU20093342  | 5050740  | 1262685  | 459031  | 4686         | 100.4       | 25.00    | 643       |
| IU20093320  | 8844032  | 2211008  | 573234  | 5007         | 97          | 20.00    | 1399      |
| IU20166005  | 15000252 | 3750063  | 3048549 | 8979         | 76.6        | 30.00    | 1768      |
| IU20091710  | 14193212 | 3548303  | 3063654 | 3385         | 104.6       | 14.00    | 2862      |
| IU20166007  | 1104896  | 276224   | 387774  | 416          | 135.5       | 7.00     | 10712     |
| IU20166008  | 11986072 | 2996518  | 309360  | 6208         | 98.4        | 39.00    | 1825      |
| IU20166009  | 24039708 | 6009927  | 706487  | 8214         | 104.4       | 39.00    | 1062      |
| IU20093353  | 6233052  | 1558263  | 750651  | 7918         | 87.5        | 30.00    | 1302      |
| IU20166011  | 28729464 | 7182366  | 251453  | 19543        | 103.4       | 102.00   | 1949      |
| IU201318751 | 5481336  | 1370334  | 186562  | 2025         | 106.3       | 8.00     | 7010      |
| IU20093358  | 16883608 | 4220902  | 529790  | 1754         | 107.6       | 8.00     | 6876      |
| IU20131875  | 65408456 | 16352114 | 4018432 | 23706        | 119.1       | 108.00   | 188       |
| IU20166026  | 1693212  | 423303   | 179558  | 889          | 110.5       | 3.00     | 12239     |
| IU2015214   | 66759976 | 16689994 | 4276467 | 19910        | 98.4        | 90.00    | 1615      |
| IU20117583  | 21674948 | 5418737  | 1826833 | 9693         | 108         | 40.00    | 1574      |
| IU20166024  | 39208584 | 9802146  | 3453791 | 12919        | 127.4       | 52.00    | 1266      |
| 1986:428    | 30376296 | 7594074  | 114368  | 25865        | 121         | 121.00   | 1749      |
| N-10        | 18150924 | 4537731  | 1828142 | 14278        | 48.2        | 99.00    | 280       |
| 2000:18     | 18999256 | 4749814  | 484591  | 28274        | 47.4        | 96.00    | 1024      |
| 1999:895    | 4683560  | 1170890  | 6062    | 18005        | 60.2        | 63.00    | 1400      |
| B-NAN1      | 24515016 | 6128754  | 1979613 | 10290        | 122.4       | 50.00    | 1664      |
| US280497    | 20277632 | 5069408  | 1331371 | 5089         | 156         | 23.00    | 1661      |
| US1278056   | 5517584  | 1379396  | 291414  | 6602         | 116.6       | 49.00    | 5145      |
| US1198988   | 6287768  | 1571942  | 393119  | 3181         | 133.5       | 16.00    | 3951      |
| US1393520   | 25585388 | 6396347  | 905360  | 2595         | 123.6       | 15.00    | 3257      |

---

|           |          |          |         |       |       |        |       |
|-----------|----------|----------|---------|-------|-------|--------|-------|
| US1278064 | 64702216 | 16175554 | 1981221 | 9129  | 100.2 | 42.00  | 792   |
| US1278065 | 35103872 | 8775968  | 454735  | 8483  | 130.8 | 29.00  | 1740  |
| US1383047 | 19077488 | 4769372  | 449783  | 591   | 151   | 3.00   | 11369 |
| US1106195 | 1441436  | 360359   | 863922  | 1129  | 144.1 | 25.00  | 11625 |
| US1278070 | 18838852 | 4709713  | 1697900 | 8808  | 118.4 | 60.00  | 2021  |
| US1278071 | 13936768 | 3484192  | 1810906 | 2513  | 157.6 | 14.00  | 10189 |
| US1286837 | 5035672  | 1258918  | 2316547 | 564   | 124   | 3.00   | 9669  |
| US1278074 | 5701248  | 1425312  | 386988  | 657   | 142.3 | 4.00   | 10893 |
| US144050  | 3546920  | 886730   | 308074  | 8788  | 144.7 | 64.00  | 4743  |
| US42602   | 1098644  | 274661   | 396228  | 310   | 117.9 | 8.00   | 8825  |
| B-STEFF   | 86791948 | 21697987 | 4186906 | 15372 | 145.2 | 86.00  | 1454  |
| B-T01     | 53500824 | 13375206 | 4858696 | 20437 | 107.6 | 123.00 | 1322  |

---

## DNA Yield

Forty samples had a concentration below 5ng/ $\mu$ L and 33 had less than 1ng/ $\mu$ L. The mean concentration was 12.04ng/ $\mu$ L with a standard deviation of 17. MNHN samples produced the highest concentrations (mean = 22.85) compared with NHM (1.58) and USNM (4.97). NHM and USNM had comparable average sample ages of 36 and 40 years, and similar mass of starting material for extraction at 10.04mg and 12.84mg, respectively. In contrast, MNHN tissue samples averaged a much lower mass at 7.77mg, but overall the samples were much younger with a mean age of 13 years. Only the institution from where specimens were sampled had a significant effect on DNA yield (z-score = 2.05, df = 2, p < 0.01; Figure 4.1), with NHM clearly producing the lowest concentrations. Despite younger samples from MNHN having higher concentrations, there was no significant interaction overall between institution and specimen age (z-score = 0.55, df = 1, p = 0.76). Age, itself, showed the smallest effect size of our predictors (z-score = 0.22).

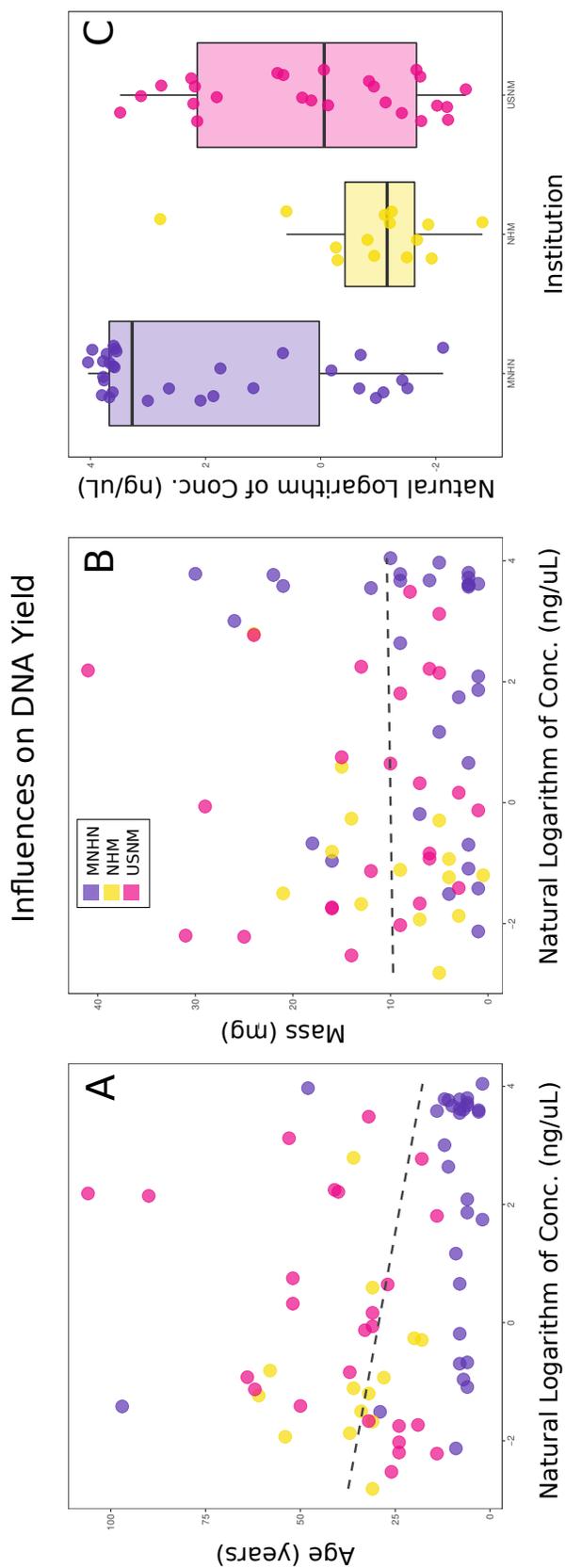


Figure 4.1: Influences of age (A), mass (B), and institution from where tissue samples were collected (C) on DNA yield. Dashed lines in A and B are ordinary least squares lines of best fit for each variable added to aid visualisation. MNHN = Muséum National d’Histoire Naturelle, Paris; NHM = Natural History Museum, London; USNM = National Museum of Natural History, Washington D.C..

## Metagenomic Analysis

Reads assigned to Malacostraca — the largest crustacean class of which isopods are a member — accounted for an average of 31.65% of the total reads per sample, with a standard deviation of 8.54 (Figure 4.2). Of the taxonomic groups we sought to assign, bacteria counted for 8.7%, fish 14.8%, fungi 10.07%, and mammals 7.64% of subsampled reads. Specimens from NHM generally had a higher proportion of bacterial and fungal reads compared to other institutions and fewer malacostracan reads, which correlates well with the fact that DNA yields and library quality were lower for NHM specimens on the whole. The same was not observed for INPA specimens despite the poor quality of the libraries — likely because they also had fewer merged reads greater than 150bp in length to be assigned. Fish reads, presumably from hosts, were present in a few samples in larger proportions than Malacostraca.

## Phylogenetic Reconstruction

Near-complete mitogenomes, for the first time, appear to corroborate the hypothesis of an external-attachment origin for cymothoids (Figure 4.3). We recover two lineages of *Anilocra* as successive sister-taxa to other cymothoids. *Anilocra leptostoma* is separate to these, within a dubious catch-all clade containing several supposedly unrelated, short-branch taxa. The only other external species included in our whole mitogenome tree, *Nerocila bivittata*, is also nested within this short-branch clade, which supports a secondary adaptation to skin-attachment. We also find two separate mouth-attaching lineages: a *Catoessa* plus *Cymothoa* clade, and a *Ceratothoa* plus *Glossobius* and *Cinusa* clade. There is also some evidence of two freshwater colonisations in South America; a main clade from which *Anphira xinguensis*, *Asotana magnifica*, and *Riggia nana* are absent.

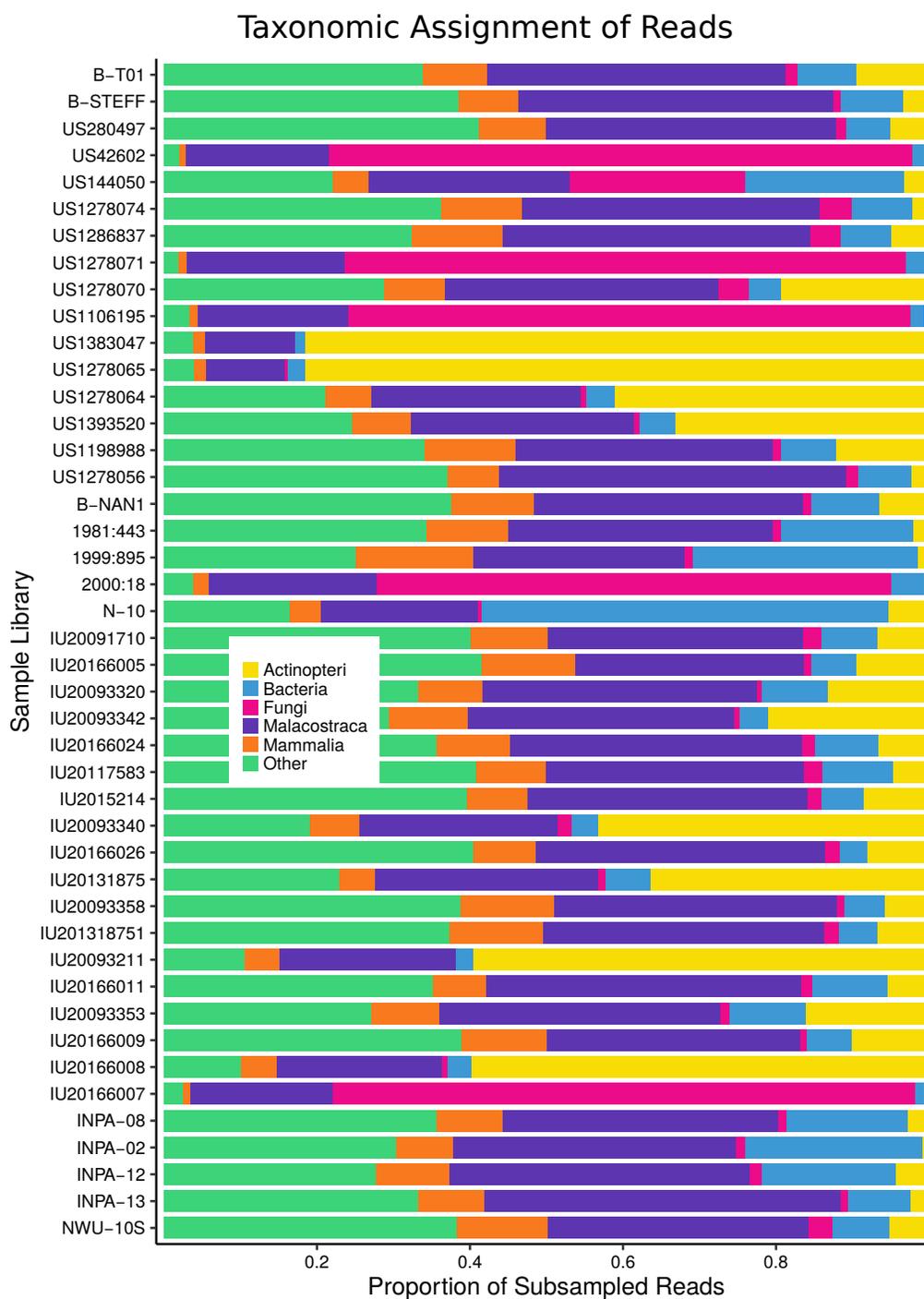


Figure 4.2: Per specimen proportion of subsampled, quality, and complexity-filtered reads assigned to different taxonomic groups.

In contrast, our COI phylogeny resolves all South American freshwater taxa in a single clade, sister to all other cymothoids (Figure 4.4). *Elthusa* — an exclusively gill-attaching genus — is sister to the remaining taxa, but is polyphyletic ('Undertmined-01' also likely belongs to *Elthusa*). We also find most other gill-attaching taxa form two clades, one corresponding to the genus *Mothocya*, and the other comprised of *Norileca* and *Ryukyua*. The *Ceratothoa* clade is still separated from the other mouth-attaching clade, *Cymothoa*, and is monophyletic, though with *Glossobius impressus* and *Ichthyoxenos japonensis* nested. *Anilocra* is polyphyletic due to *Anilocra leptostoma* being nested with the South American taxa, but the main *Anilocra* clade is recovered in a terminal position. *Nerocila*, the other skin-attaching genus, are sister to an *Anilocra* plus *Mothocya* (external- and gill-attaching) clade.

## 4.4 Discussion

We successfully recovered mitochondrial sequences from cymothoid specimens accessioned in museums up to 55 years ago. Partial mitogenomes were constructed for 21 species that do not have current molecular data, including for eight freshwater species from South America: only two gene sequences for two species were previously available. We found that most specimens were highly degraded with average fragment lengths of 100bp, and would, therefore, not have been easily amplified by conventional PCR targeting. Short-read, shotgun sequencing, which requires small insert libraries anyway, proved to be effective at recovering high-coverage sequences of mitogenomic origin since these are naturally present in high copy number.

Our phylogenies establish that colonisation of freshwater was likely a single event, with some uncertainty as to whether it occurred early or late in cymothoid evolutionary history. The whole mitogenome tree does recover a polyphyletic South American clade

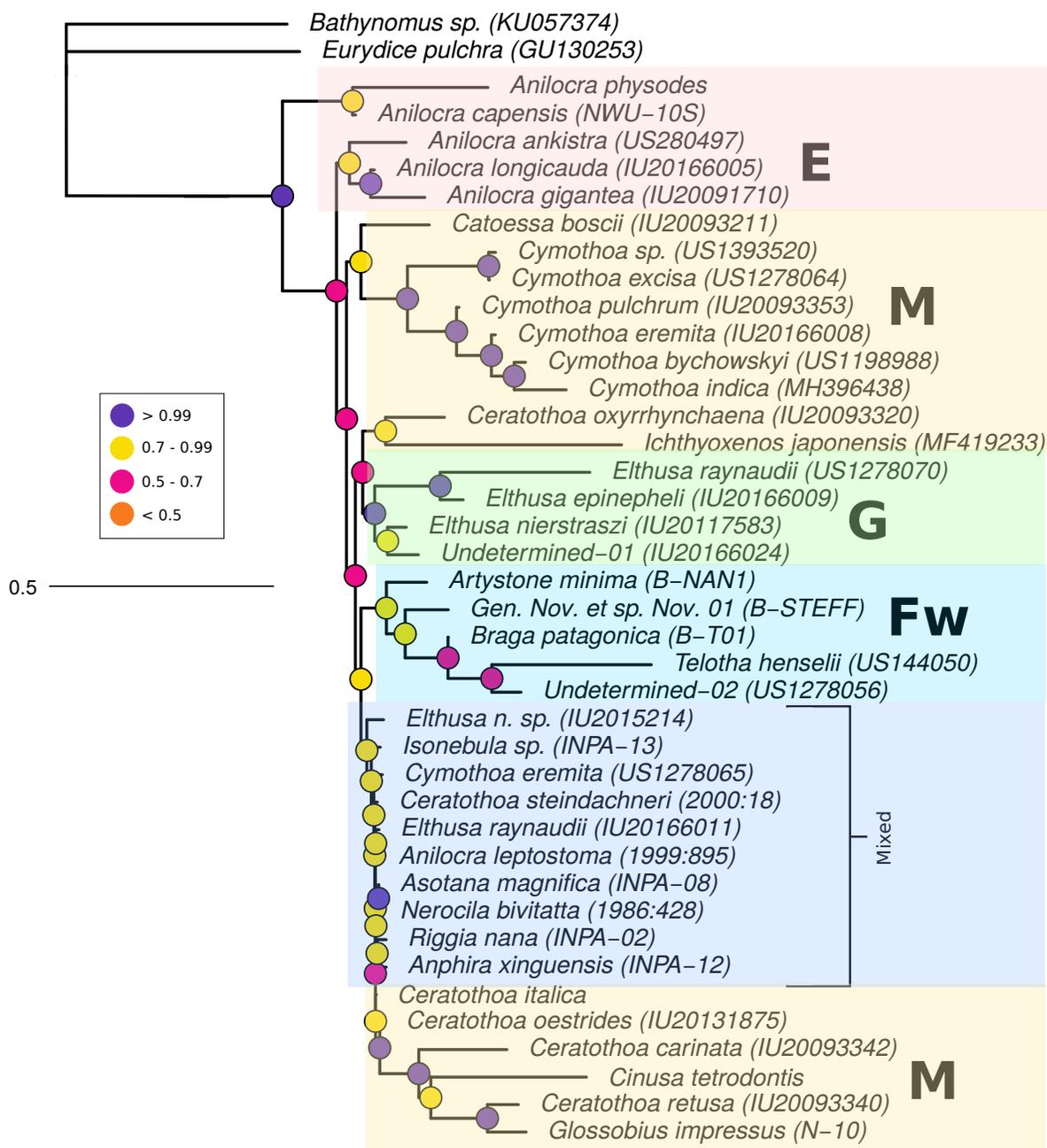


Figure 4.3: Cymothoid whole mitogenome phylogeny. Key refers to bootstrap support values for coloured nodes. E = External, Fw = Freshwater, G = Gill, M = Mouth.

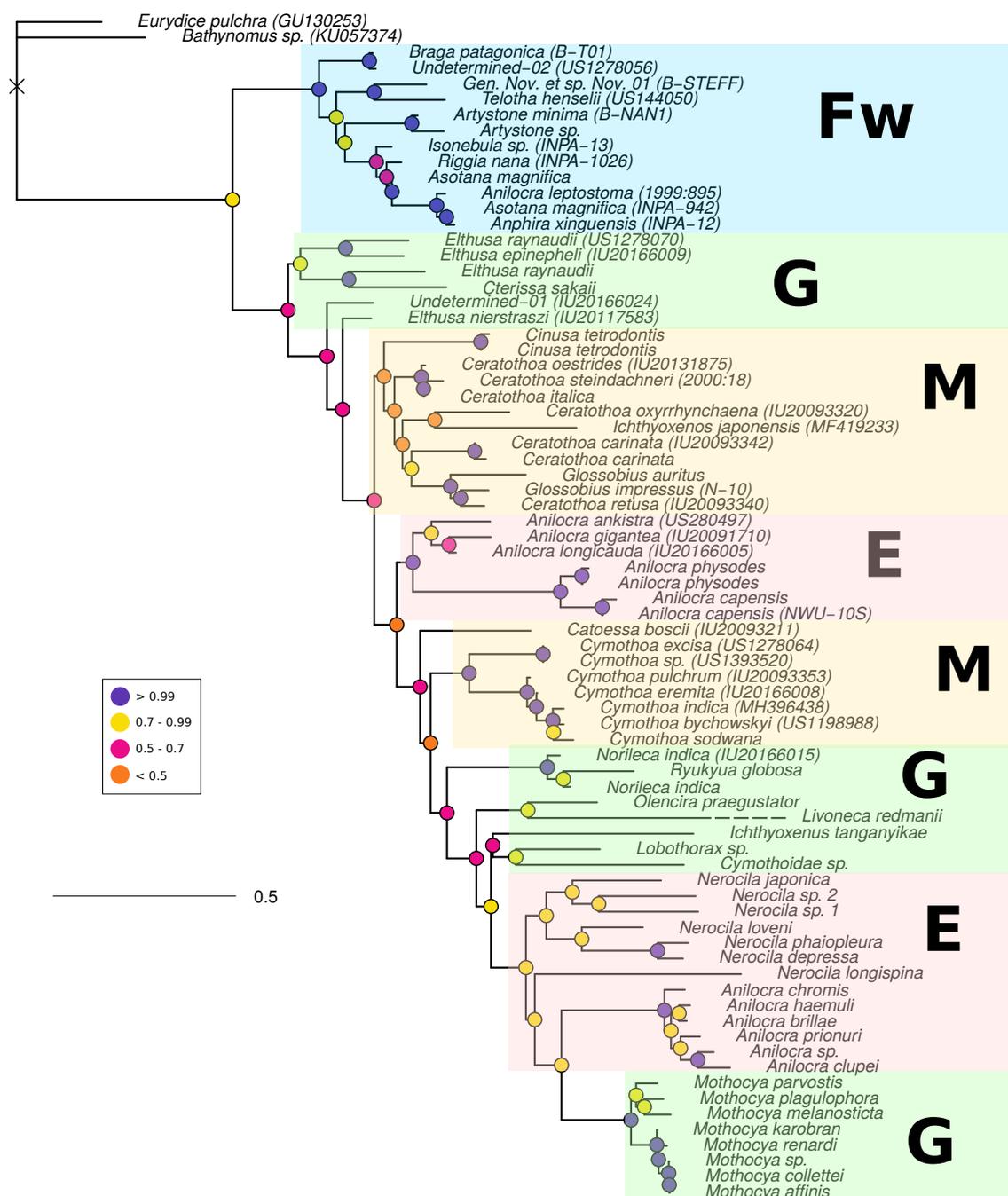


Figure 4.4: Cymothoid COI phylogeny. Key refers to bootstrap support values for coloured nodes. Most recent common ancestor of cymothoid and outgroup taxa is denoted by an 'X'. E = External, Fw = Freshwater, G = Gill, M = Mouth

but the position of the three missing species within the short-branch clade cannot be considered reliable. Despite there being more data in the whole genome dataset, the COI alignment had proportionally less ambiguous bases. In contrast, our COI phylogeny provides strong support for a single origin for freshwater colonisation in South America — expanding upon the findings of Hata et al. (2017). Furthermore, *Asotana* and *Braga* are not closely related within the freshwater clade, in either of our phylogenies which also appears to contradict Brusca's (1981) proposal of two freshwater colonisations, one including *Asotana* and *Braga*, and the other predominantly comprised of the burrowing species. With improved data and expanded sampling to include other freshwater species from Africa and Asia, a molecular dating analysis could help to resolve the timing and direction of shifts between marine and freshwater habitats.

We observe a similarly dynamic pattern of attachment as in previous molecular studies, but cannot establish what is the plesiomorphic condition. It is interesting that we find two lineages of *Anilocra* species to be successive sister-taxa to the remaining cymothoids in our whole mitogenome phylogeny: to date externally-attaching taxa have always been recovered in a terminal position (Ketmaier et al., 2008; Hata et al., 2017). In both phylogenies, *Anilocra* is polyphyletic, which has previously been recognised based on the large genetic distances between Mediterranean and Caribbean species (Welicky et al., 2019). Here we find a more complex biogeographic pattern, with one clade containing species from the Caribbean and Japan (the *A. chromis* clade) and another with species from the Indian Ocean and the Mediterranean (the *A. physodes* clade). It would be interesting to discover the timing of divergence between species in the former clade to understand how they could be distributed either side of the Americas: perhaps it is a very recent colonisation, post building of the Panama Canal, or dispersal occurred before the formation of the Isthmus of Panama 5 – 15 MYA (Montes et al., 2015). All of the most terminal taxa in our COI tree are gill- and skin-attaching species.

The most parsimonious explanation given this tree is that skin-attachment has arisen twice, with two reversals to gill-attachment in *Norileca* and *Mothocya*. Clearly cymothoids have adapted to mouth-attachment more than once since in both trees we find two clades, each corresponding to the genera *Ceratothoa* and *Cymothoa*. Nested within the *Ceratothoa* clade we recover *Glossobius* spp. which were once considered the same genus (Richardson, 1905), and only separated based on length and shape of pereonites one and six (Martin et al., 2015). It is likely these taxa and the status of *Cinusa tetrodontis*, which also appears within this clade, need to re-examined.

As in other studies, we find that specimen age is not necessarily related to the amount of DNA recovered (Burrell et al., 2015). We also achieved DNA yields consistent with other work on small, ethanol-preserved invertebrates, averaging 420ng of DNA (Tin et al., 2014; Cotoras et al., 2017). However, we did experience a low conversion rate between samples extracted and those producing sufficient coverage and quality to be mapped and assembled. The largest drop out was after extraction. The primary protocol we used was designed for bones and teeth, using relatively concentrated EDTA to release DNA bound to hydroxyapatite molecules by decalcification — a feature we thought might have analogous effects with chitinous, crustacean exoskeleton (Dabney et al., 2013a). In hindsight most DNA from our specimens is likely to have been preserved in cells and the addition of a surfactant might have increased yields (Cotoras et al., 2017). Additionally, the mass of material we sampled was exceptionally small. In isolation, mass is not necessarily a limiting factor, as museums often require minimal or non-destructive sampling of specimens, and many previous studies have had equally little starting material or even extracted directly from preservative (Shokralla et al., 2010). However, we suppose that cell-containing tissue as a proportion of mass is particularly small for cymothoid pereopods since they have such a thickened cuticle. This could also explain the wide variability in the number of reads mapping, and in the number of

malacostracan reads in our libraries.

We could not find any other study that attempted to process a number of specimens equal to that here, and perhaps with a more selective sampling approach and better knowledge of preservation history we could have produced data for more specimens. Certainly, institution is important when considering use of archived specimens for sequencing. The low yield and small read lengths of specimens from NHM, we assume, is because of their use of industrial methylated spirits, rather than ethanol, as a preservative. The Principal Curator in Charge of Invertebrates, Dr. Laura Hughes, also informed us after we had extracted the samples that specimens were likely to have once been preserved in formalin (pers. comm., 27th November 2018). Records of preservation condition were generally poor, but we did recover mitogenomes for several specimens which we knew had been formalin-fixed by adjusting lysis steps and including a post-extraction, enzymatic treatment. Indeed, it is feasible to obtain nuclear genome sequences from specimens over 100 years old (Ruane and Austin, 2017). Specimens from NHM and INPA showed a far higher percentage of malacostracan reads than we expected given the quality of the samples, and what we knew about preservation. However, we did choose to assign merged reads with a length of 150bp to increase accuracy, and these samples did have fewer merged reads of this length compared with other samples. Therefore, we can conclude that if ‘good’ reads (i.e. passing quality and complexity filtering, and longer than 150bp) were present, they were more likely to be of malacostracan origin. We adhered to strict aDNA protocols during all laboratory procedures and mammalian reads are consistent across samples, suggesting that these levels of stringency were effective in reducing human contamination — a major risk when working with degraded specimens (Dabney et al., 2013a; Llamas et al., 2017).

## 4.5 Conclusion

We recovered partial mitogenomes from museum-preserved cymothoid specimens, including nine representatives of the South American freshwater taxa. Our results show that colonisation of South American rivers probably has a single origin. In our COI analysis, we also recover these freshwater taxa as sister to other isopods — a surprising relationship not previously considered. In contrast, our whole mitogenome tree finds *Anilocra* as sister to other isopods — also a novel finding in molecular phylogenies of cymothoids, and one which supports the original formulation of cymothoid evolution by Brusca (1981). However, in both trees we find, and in common with previous analyses, that there have been multiple transitions between parasitic modes, and a linear evolutionary pathway from externally-attaching, to gill and mouth attachment is not supported. Likewise, we find that taxonomic concepts for genera warrant further examination, and revisions to include molecular systematics data. Some genera (e.g. *Cymothoa* and *Mothocyca*) seem to be upheld by our phylogenies, while others (e.g. *Anilocra*, *Ichthyoxenos*, and *Nerocila*) are not. Ancient DNA analysis of museum specimens, ‘museomics’, is a potentially powerful approach for phylogenetic studies of scarce, or hard to sample taxa. For cymothoids, we would recommend increasing the mass of starting material, and potentially to use other tissues, such as pleopods, as long as important morphological characters were not destroyed.

## **Conclusions**

This work has examined several aspects of cymothoid isopod evolution, including functional traits, the position of Cymothoidae within Isopoda, and the relationships between cymothoid species. Two important findings emerge: mitogenome sequences are not suitable for reliably recovering phylogenetic relationships between isopod families, and current taxonomic concepts are not reflective of the complex diversification processes underlying cymothoid evolution. With our data, we can confidently say the pattern by which different parasitic strategies have evolved is not linear, with multiple transitions between attachment modes. This supports the findings of the very first molecular phylogeny of cymothoids, which had only six species (Ketmaier et al., 2008). Furthermore, it appears that the South American freshwater species transitioned to riverine habitats in a single event and diversified thereafter, contrary to previous hypotheses (Brusca, 1981). Although outstanding questions remain, such as identification of the ancestral cymothoid condition, we have succeeded in meeting the objectives of this thesis.

In Chapter 3, where we aimed to assess the position of cymothoids within Isopoda, we chanced upon the truly remarkable possibility that all isopods, not just the oniscids, possess an atypical mitogenome architecture. We suggest that, if this is case, preferential use of origin of replication copies could explain differences in compositional strand bias which so negatively impact phylogenetic inference. We could not absolutely establish the ubiquity of this mitogenome conformation with the sampling and other methods used (and our aim was not to directly test the presence of this structure), but characterisation of mitogenomes across the group should be a priority for future research as it is likely central to the story of isopod evolution. Compositional strand bias could not explain all long-branch effects in our analyses, where, supposedly unrelated, species from very different environments but with the same negative strand bias clustered together. This suggests that positive selection may also be influencing mitogenome base composition and evolutionary rates. Isopods are noteworthy in their ecological diversity, occupying niches from

the deep ocean to temperate forests. It could be that atypical mitogenome architectures are the very thing that made isopod diversification possible, maximising the potential of crucial mitogenome function to adapt to new environments. By understanding the factors that maintain these mitogenome configurations, their interaction with the large-scale gene rearrangements we observe, and the timing of their origin, these features may even prove to be useful phylogenetic markers in their own right. Sequencing across the dimer junctions and enzyme-based probes on their own, it seems, might miss architectural signatures in some species, so new assays will need to be developed. We suggest future work needs to incorporate long-reads from other platforms, such as Pacific Biosciences SMRT sequencing (<https://www.pacb.com/smrt-science/smrt-sequencing/>), combined with conformation capture sequencing (Belton et al., 2012), and technologies like ATAC-seq (Buenrostro et al., 2015) to assess higher-order genome organisation and the formation of isopod mitogenomes during embryogenesis.

Regardless of the taxonomic extent of atypical isopod mitogenomes, we showed phylogeny reconstructions based on these data are untrustworthy, and expanding the availability of nuclear genomes will be key to resolution of inter-familial relationships. This will be a particular challenge when we consider the fact that isopods are amongst the least studied arthropods compared with their diversity. Since we must discount evidence from mitogenome phylogenies above the family level, the current hypothesis of a single origin of terrestrialisation in isopods still stands. If this is the case, it would be interesting to understand the constraints preventing multiple transitions to land, and the contrasts with deep-sea colonisations, which have been numerous (Lins et al., 2012). If, as we suspect, all isopods maintain a dimerised mitogenome structure, resolution of isopod phylogeny could answer whether the major ecological shifts we observe are associated with mitochondrial structural variation.

Below the family level, mitogenomes appear to be useful markers for phylogenetics, and in Chapter 4 we assembled partial mitogenomes for further cymothoid species using museum specimens. The high copy number of mitochondria, coupled with the short read lengths of illumina platforms, meant we were able to produce high coverage mitogenomes from degraded DNA using a shotgun sequencing approach. Combining just the COI region with other previously published sequences, we have expanded the availability of sequence data for cymothoids to over 17% of currently known species. It is clear in both our mitochondrial Cytochrome Oxidase subunit I (COI) and whole mitogenome trees that previous comments are well justified, where they suggest many morphological features used in cymothoid taxonomy are homoplastic, and thus evolutionary relationships are likely to be obscured by the arrangement of taxa in the current system. Supporting this we also find, in Chapter 2, that ecology of parasitic strategy is the primary driver of attachment morphology, with relatively less related species sharing morphological characteristics. Future taxonomic work needs to focus more on systematics and apply molecular phylogenetics methods, which incorporate more objective evolutionary models. Many genera also need to be re-evaluated. From our analyses, and recent taxonomic literature, the genera *Anilocra* and *Elthusa* should be high priorities (Welicky and Smit, 2019). We propose that *Anilocra* likely represents three evolutionary lineages with distinct distributions (Mediterranean and East Atlantic, Indian Ocean and Western Pacific, and Eastern Pacific and Caribbean), that may, or may not, have arisen from a common ancestor: ecological convergence of morphological traits providing an explanation for why *Anilocra* species have thus far been lumped together.

Our study was entirely shaped by sampling strategy and the availability of specimens: restrictions which have limited the application of molecular data to cymothoid phylogenetics, which future studies must address. In particular, we were confronted with a compromise between specimen age and amount of tissue (limited to a few pereopods),

and broad taxonomic sampling to better represent familial diversity. With limited budget, museum specimens produced lower data quality and reliability than would fresh tissues, but the diversity of specimens in our analyses was high compared with other cymothoid phylogenies. Our study also provides further evidence that ancient DNA methods work well for small, liquid-preserved arthropods, though protocols could be further optimised. More extracts, libraries, and sequencing of our specimens could also have delivered increased data confidence but would have incurred costs beyond our means. Regardless, if the cymothoid species-tree, as opposed to relationships between mitochondrial lineages, is to be resolved future studies will need to incorporate genome-wide, single-copy markers of phylogenetic utility. Methods for the recovery of markers such as Ultra Conserved Elements (UCEs) and whole exomes, perform much better with high molecular weight DNA, which requires larger quantities of fresh material.

That is not to say that museum specimens should be completely discounted for sourcing suitable genetic material. In our study, Muséum National d'Histoire Naturelle, Paris (MNHN) provided well-preserved and relatively young specimens because the museum continues to engage in collection expeditions focussed on under-represented taxa. Further to our molecular work using specimens from the EXBODI expedition (<https://expeditions.mnhn.fr/campaign/exbodi>), we identified three MNHN specimens of *Elthusa parva* previously only known, and described, from a single specimen collected in 1905. We have been able to add host data for *E. parva*, ecological information, and to build species distribution models for what is obviously a rare species — all of which are the subject of a forthcoming publication. As a community we should actively promote and fund museums, with their expertise, to lead specimen collection, as well as curation, research and education work. Fortunately, most museums also now host genomics facilities and are increasingly connecting new and old collection work with genomics projects — a pertinent example being the involvement of the Natural History Museum, London (NHM) in

the The Darwin Tree of Life Project (<https://expeditions.mnhn.fr/campaign/exbodi>).

Originally, when we conceived this project, our ultimate ambition was to develop cymothoid parasitism as a model for understanding co-evolutionary processes. As a first step we were most interested in uncovering macroevolutionary patterns, including: the extent to which cymothoids co-speciated with hosts; how prevalent host-switching might have been as a speciation mechanism; which combination of host or parasite traits might favour intrahost speciation; the directionality of evolutionary change between parasitic strategies; the factors influencing host range (the number of suitable host species). These lofty ambitions were abruptly tempered by the realisation that we do not yet understand, or even have sufficient samples, to concretely establish boundaries for any cymothoid taxon, let alone the whole family. This work patently needs to be done and broader collaboration between groups would help: taxonomists need to be less protective of specimens and ecologists more so. Money helps. We had several grant rejections for proposals to resolve cymothoid systematic uncertainty based on the work being too systematic.

At the very least we have taken the first steps to modernising cymothoid phylogenetic research, both in terms of the data collected, and the models and computational methods applied, and hopefully that will be continued by others alongside the systematics ground-work. And be reminded on your next woodland walk, that crustaceans with marvellously strange mitogenomes, whose cousins bite the tongues of fish, litter the undergrowth.

## Bibliography

- Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(2), DOI: 10.1093/nar/gkq291.
- Adams, D. C. and Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4):393–399.
- Adlard, R. and Lester, R. (1995). The life-cycle and biology of *Anilocra pomacentri* (Isopoda: Cymothoidae), an ectoparasitic isopod of the coral-reef fish, *Chromis nitida* (Perciformes: Pomacentridae). *Australian Journal of Zoology*, 43(3):271–281.
- Araujo, S. B. L., Braga, M. P., Brooks, D. R., Agosta, S. J., Hoberg, E. P., von Hartenthal, F. W., and Boeger, W. A. (2015). Understanding host-switching by ecological fitting. *PLoS ONE*, 10:139–225, DOI: 10.1371/journal.pone.0139225.
- Aznar, F. J., Hernández-Orts, J. S., and Raga, J. A. (2018). Morphology, performance and attachment function in *Corynosoma* spp. (Acanthocephala). *Parasites & Vectors*, 11:633, DOI: 10.1186/s13071-018-3165-1.
- Bakenhaster, M. D., McBride, R. S., and Price, W. W. (2006). Life history of *Glossobius hemiramphi* (Isopoda: Cymothoidae): development, reproduction, and symbiosis with its host *Hemiramphus brasiliensis* (Pisces: Hemiramphidae). *Journal of Crustacean Biology*, 26(3):283–294.

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5):455–477.
- Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*, 58(3):268–276.
- Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, 21(2):163–193.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J., Middendorf, M., and Stadler, P. F. (2013). MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2):313–319.
- Bollback, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, 19(7):1171–1180.
- Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243.
- Botero-Castro, F., Tilak, M.-K., Justy, F., Catzeflis, F., Delsuc, F., and Douzery, E. J. (2018). In cold blood: compositional bias and positive selection drive the high evolutionary rate of vampire bats mitochondrial genomes. *Genome Biology and Evolution*, 10(9):2218–2239.
- Bowman, T. E. (1960). Description and notes on the biology of *Lironeca puhi*, n. sp. (Isopoda: Cymothoidae), parasite of the Hawaiian moray eel, *Gymnothorax eurostus* (Abbott). *Crustaceana*, 1:84–91.

- Boyko, C., Bruce, N., Hadfield, K., Merrin, K., Ota, Y., Poore, G., Taiti, S., Schotte, M., and Wilson, G. E. (2019). World marine, freshwater and terrestrial isopod crustaceans database. Cymothoidae Leach, 1818. <http://www.marinespecies.org/aphia.php?p=taxdetails&id=118274>.
- Brandt, A. and Poore, G. C. (2003). Higher classification of the flabelliferan and related Isopoda based on a reappraisal of relationships. *Invertebrate Systematics*, 17(6):893, DOI: 10.1071/is02032.
- Bruce, N. L. (1986). Revision of the isopod crustacean genus *Mothocya* Costa, in Hope, 1851 (Cymothoidae: Flabellifera), parasitic on marine fishes. *Journal of Natural History*, 20(5):1089–1192.
- Bruce, N. L. (1987). Australian species of *Nerocila* Leach, 1818, and *Creniola* n.gen. (Isopoda: Cymothoidae), crustacean parasites of marine fishes. *Records of the Australian Museum*, 39(6):355–412, DOI: 10.3853/j.0067-1975.39.1987.174.
- Bruce, N. L. (1990). The genera *Catoessa*, *Elthusa*, *Enispa*, *Ichthyoxenus*, *Idusa*, *Livoneca* and *Norileca* n.gen. (Isopoda: Cymothoidae), crustacean parasites of marine fishes, with descriptions of Eastern Australian species. *Records of the Australian Museum*, 42(3):247–300, DOI: 10.3853/j.0067-1975.42.1990.118.
- Brusca, R. C. (1978). Studies on the cymothoid fish symbionts of the Eastern Pacific (Isopoda: Cymothoidae) I. Biology of *Nerocila californica*. *Crustaceana*, 34(2):141–154.
- Brusca, R. C. (1981). A monograph on the Isopoda Cymothoidae (Crustacea) of the Eastern Pacific. *Zoological Journal of the Linnean Society*, 73(2):117–199, DOI: 10.1111/j.1096-3642.1981.tb01592.x.
- Brusca, R. C. and Gilligan, M. R. (1983). Tongue replacement in a marine fish (*Lutjanus guttatus*) by a parasitic isopod (Crustacea: Isopoda). *Copeia*, 3:813–816.

- Brusca, R. C. and Wilson, G. D. (1991). A phylogenetic analysis of the Isopoda with some classificatory recommendations. *Memoirs of Queensland Museum*, 31:143–204.
- Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109(1):21–29.
- Bunkley-Williams, L. and Williams Jr., E. H. (1981). Nine new species of *Anilocra* (Crustacea: Isopoda: Cymothoidae) external parasites of West Indian coral reef fishes. *Proceedings of the Biological Society of Washington*, 94(4):1005–1047.
- Bunkley-Williams, L. and Williams Jr., E. H. (1998). Isopods associated with fishes: a synopsis and corrections. *Journal of Parasitology*, 84:893–896.
- Burrell, A. S., Disotell, T. R., and Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, 79:35–44.
- Bush, S. E., Sohn, E., and Clayton, D. H. (2006). Ecomorphology of parasite attachment: experiments with feather lice. *Journal of Parasitology*, 92:25–31, DOI: 10.1645/GE-612R.1.
- Bushnell, B. (2019). BBtools. <https://sourceforge.net/projects/bbmap/>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). Blast+: architecture and applications. *BMC Bioinformatics*, 10(1):421.
- Campos, P. F. and Gilbert, T. M. (2012). DNA extraction from formalin-fixed material. In Shapiro, B. and Hofreiter, M., editors, *Ancient DNA*, volume 840, pages 81–85. Humana Press, DOI: [https://doi.org/10.1007/978-1-61779-516-9\\_11](https://doi.org/10.1007/978-1-61779-516-9_11).

- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.
- Carrassón, M. and Cribb, T. (2014). Benign effect of the fish parasitic isopod *Ceratothoa* cf. *imbricata* on *Selenotoca multifasciata* (Scatophagidae) from Australia. *Diseases of Aquatic Organisms*, 110(3):173–180.
- Castellana, S., Vicario, S., and Saccone, C. (2011). Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biology and Evolution*, 3:1067–1079.
- Castresana, J. (2002). Gblocks: selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Version 0.91 b. Copyrighted by J. Castresana, EMBL*.
- Cenzi de Ré, F., Loreto, E. L., and Robe, L. J. (2010). Gene and species trees reveal mitochondrial and nuclear discordance in the *Drosophila cardini* group (Diptera: Drosophilidae). *Invertebrate Biology*, 129(4):353–367.
- Chandler, C. H., Badawi, M., Moumen, B., Grève, P., and Cordaux, R. (2015). Multiple conserved heteroplasmic sites in tRNA genes in the mitochondrial genomes of terrestrial isopods (Oniscidea). *G3: Genes, Genomes, Genetics*, 5(7):1317–1322.
- Charif, D. and Lobry, J. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In Bastolla, U., Porto, M., Roman, H., and Vendruscolo, M., editors, *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, Biological and Medical Physics, Biomedical Engineering, pages 207–232. Springer Verlag, New York. ISBN : 978-3-540-35305-8.

- Collyer, M. L. and Adams, D. C. (2018). RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. *Methods in Ecology and Evolution*, 9(7):1772–1779.
- Cotoras, D., Murray, G., Kapp, J., Gillespie, R., Griswold, C., Simison, W., Green, R., and Shapiro, B. (2017). Ancient DNA resolves the history of *Tetragnatha* (Araneae: Tetragnathidae) spiders on Rapa Nui. *Genes*, 8(12):403.
- Crampton-Platt, A., Douglas, W. Y., Zhou, X., and Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, 5(1):15.
- Cummins, C. A. and McInerney, J. O. (2011). A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology*, 60(6):833–844.
- Curole, J. P. and Kocher, T. D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution*, 14(10):394–398.
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013a). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39):15758–15763.
- Dabney, J., Meyer, M., and Pääbo, S. (2013b). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, 5(7):a012567.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8):772.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray, Albemarle Street, London.

- Dayhoff, M., Schwartz, R., and Orcutt, B. (1978). A model of evolutionary change in proteins. In Dayhoff, M., editor, *Atlas of Protein Sequence and Structure*, volume 5, pages 345–352. National Biomedical Research Foundation, Silver Spring.
- de Lima, J. T., Chellappa, S., and Thatcher, V. (2005). *Livoneca redmanni* Leach (Isopoda: Cymothoidae) and *Rocinelasignata* Schioedte & Meinert (Isopoda: Aegidae), ectoparasites of *Scomberomorus brasiliensis* Collette, Russo & Zavala-camin (Osteichthyes: Scombridae) in Rio Grande do Norte, Brazil. *Revista Brasileira de Zoologia*, 22(4):1104–1108.
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4):e18–e18.
- Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F., and Jetz, W. (2008). Homage to Linnaeus: how many parasites? How many hosts? *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11482–11489.
- Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, 20(9):525–527.
- Doublet, V., Helleu, Q., Raimond, R., Souty-Grosset, C., and Marcadé, I. (2013). Inverted repeats and genome architecture conversions of terrestrial isopods mitochondrial DNA. *Journal of Molecular Evolution*, 77(3):107–118.
- Doublet, V., Raimond, R., Grandjean, F., Lafitte, A., Souty-Grosset, C., and Marcadé, I. (2012). Widespread atypical mitochondrial DNA structure in isopods (Crustacea: Peracarida) related to a constitutive heteroplasmy in terrestrial species. *Genome*, 55(3):234–244.

- Doublet, V., Ubrig, E., Alioua, A., Bouchon, D., Marcadé, I., and Maréchal-Drouard, L. (2015). Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean *Armadillidium vulgare*. *RNA Biology*, 12(10):1159–1168.
- Dreyer, H. and Wägele, J.-W. (2001). Parasites of crustaceans (Isopoda: Bopyriidae) evolved from fish parasites: molecular and morphological evidence. *Zoology*, 103(3/4):157–178.
- Dreyer, H. and Wägele, J.-W. (2002). The Scutocoxifera tax. nov. and the information content of nuclear ssu rDNA sequences for reconstruction of isopod phylogeny (Peracarida: Isopoda). *Journal of Crustacean Biology*, 22(2):217–234.
- Duchêne, S., Archer, F. I., Vilstrup, J., Caballero, S., and Morin, P. A. (2011). Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS one*, 6(11):e27138.
- Dym, C. L. and Williams, H. E. (2010). Stress and displacement estimates for arches. *Journal of Structural Engineering*, 137(1):49–58.
- Engelstädter, J. and Fortuna, N. Z. (2019). The dynamics of preferential host switching: host phylogeny as a key predictor of parasite distribution. *Evolution*.
- Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G., and Pisani, D. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Current Biology*, 27(24):3864–3870.
- Fogelman, R. and Grutter, A. (2008). Mancae of the parasitic cymothoid isopod, *Anilocra apogonae*: early life history, host-specificity, and effect on growth and survival of preferred young cardinal fishes. *Coral Reefs*, 27(3):685.

- Fogelman, R. M., Kuris, A. M., and Grutter, A. S. (2009). Parasitic castration of a vertebrate: effect of the cymothoid isopod, *Anilocra apogonae*, on the five-lined cardinalfish, *Cheilodipterus quinquelineatus*. *International Journal for Parasitology*, 39(5):577–583.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit 1 from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 26(5):294–299.
- Fortes, G. G. and Paijmans, J. L. (2015). Analysis of whole mitogenomes from ancient samples. In Kroneis, T., editor, *Whole Genome Amplification. Methods in Molecular Biology*, volume 1347, pages 179–195. Human Press, New York.
- Foster, P. G. (2004). Modeling compositional heterogeneity. *Systematic Biology*, 53(3):485–495.
- Foster, P. G. and Hickey, D. A. (1999). Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, 48(3):284–290.
- Frandsen, P. B., Calcott, B., Mayer, C., and Lanfear, R. (2015). Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. *BMC Evolutionary Biology*, 15(1):13.
- Fryer, G. (1968). A new parasitic isopod of the family Cymothoidae from clupeid fishes of Lake Tanganyika — a further Lake Tanganyika enigma. *Journal of Zoology*, 156(1):35–43.
- Fussey, G. (1984). The distribution of the two forms of the woodlouse *Trichoniscus pusillus* Brandt (Isopoda: Oniscoidea) in the British Isles: a reassessment of geographic parthenogenesis. *Biological journal of the Linnean Society*, 22(4):309–321.

- George, R. Y. (1972). Biphasic moulting in isopod Crustacea and the finding of an unusual mode of moulting in the Antarctic genus *Glyptonotus*. *Journal of Natural History*, 6:651–656, DOI: 10.1080/00222937200770591.
- Gillett, C. P., Crampton-Platt, A., Timmermans, M. J., Jordal, B. H., Emerson, B. C., and Vogler, A. P. (2014). Bulk *de novo* mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, 31(8):2223–2237.
- Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., and Orlando, L. (2011). mapdamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, 27(15):2153–2155.
- Goldman, N. (1993). Statistical tests of models of DNA substitution. *Journal of molecular evolution*, 36(2):182–198.
- Gorb, S. N. (2008). Biological attachment devices: exploring nature's diversity for biomimetics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366:1557–1574, DOI: 10.1098/rsta.2007.2172.
- Grandjean, F., Tan, M. H., Gan, H. M., Lee, Y. P., Kawai, T., Distefano, R. J., Blaha, M., Roles, A. J., and Austin, C. M. (2017). Rapid recovery of nuclear and mitochondrial genes by genome skimming from northern hemisphere freshwater crayfish. *Zoologica Scripta*, 46(6):718–728.
- Grutter, A. S. (2002). Cleaning symbioses from the parasites' perspective. *Parasitology*, 124(7):65–81.
- Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704.

- Guschanski, K., Krause, J., Sawyer, S., Valente, L. M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z. T., et al. (2013). Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology*, 62(4):539–554.
- Hadfield, K. A., Bruce, N. L., and Smit, N. J. (2010). Redescription of the monotypic genus *Cinusa* Schioedte and Meinert, 1884 (Isopoda: Cymothoidae), a buccal-cavity isopod from South Africa. *Zootaxa*, 2437:51–68.
- Hadfield, K. A., Bruce, N. L., and Smit, N. J. (2013). Review of the fish-parasitic genus *Cymothoa* Fabricius, 1793 (Isopoda: Cymothoidae: Crustacea) from the southwestern Indian Ocean, including a new species from South Africa. *Zootaxa*, 3640:152–176, DOI: 10.11646/zootaxa.3640.2.2.
- Hadfield, K. A., Bruce, N. L., and Smit, N. J. (2014). Review of the fish parasitic genus *Ceratothoa* Dana, 1852 (Crustacea: Isopoda: Cymothoidae) from South Africa, including the description of two new species. *ZooKeys*, 400:1–42, DOI: 10.3897/zookeys.400.6878.
- Hadfield, K. A., Bruce, N. L., and Smit, N. J. (2016a). Redescription of poorly known species of *Ceratothoa* Dana, 1852 (Crustacea: Isopoda: Cymothoidae), based on original type material. *ZooKeys*, 592:39–91.
- Hadfield, K. A., Bruce, N. L., and Smit, N. J. (2016b). *Elthusia nierstraszi* nom. n., the replacement name for *Elthusia parva* (Nierstrasz, 1915), a junior secondary homonym of *Elthusia parva* (Richardson, 1910) (Isopoda: Cymothoidae). *ZooKeys*, 619:167–170, DOI: 10.3897/zookeys.619.10143.
- Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads — a baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13):e129–e129.

- Hassanin, A., Leger, N., and Deutsch, J. (2005). Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Systematic Biology*, 54(2):277–298.
- Hata, H., Sogabe, A., Tada, S., Nishimoto, R., Nakano, R., Kohya, N., Takeshima, H., and Kawanishi, R. (2017). Molecular phylogeny of obligate fish parasites of the family Cymothoidae (Isopoda: Crustacea): evolution of the attachment mode to host fish and the habitat shift from saline water to freshwater. *Marine Biology*, 164(5):105.
- Heath, T. A., Hedtke, S. M., and Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3):239–257.
- Hernandez, D., François, P., Farinelli, L., Østerås, M., and Schrenzel, J. (2008). *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*, 18(5):802–809.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2):518–522.
- Horton, T. (2000). *Ceratothoa steindachneri* (Isopoda: Cymothoidae) new to British waters with a key to north-east Atlantic and Mediterranean *Ceratothoa*. *Journal of the Marine Biological Association of the United Kingdom*, 80(6):1041–1052.
- Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C., et al. (2019). World Register of Marine Species (WoRMS). <http://www.marinespecies.org>.
- Horton, T. and Okamura, B. (2001). Cymothoid isopod parasites in aquaculture: a review and case study of a Turkish sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus auratus*) farm. *Diseases of Aquatic Organisms*, 46(3):181–188.

- Hua, C. J., Li, W. X., Zhang, D., Zou, H., Li, M., Jakovlić, I., Wu, S. G., and Wang, G. T. (2018). Basal position of two new complete mitochondrial genomes of parasitic Cymothoida (Crustacea: Isopoda) challenges the monophyly of the suborder and phylogeny of the entire order. *Parasites & vectors*, 11(1):628.
- Hua, C. J., Zhang, D., Zou, H., Wu, S. G., Wang, G. T., Braicovich, P. E., and Li, W. X. (2017). Effects of the parasitic isopod *Ichthyoxenus japonensis* Richardson, 1913 (Isopoda: Cymothoidae) on the growth and gonad development of the goldfish *Carassius auratus* Linnaeus, 1758. *The Journal of Crustacean Biology*, 37(3):356–358.
- Hung, C.-M., Lin, R.-C., Chu, J.-H., Yeh, C.-F., Yao, C.-J., and Li, S.-H. (2013). The *de novo* assembly of mitochondrial genomes of the extinct passenger pigeon (*Ectopistes migratorius*) with next generation sequencing. *PLoS One*, 8(2):e56301.
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3):377–386.
- Jones, C. M., Miller, T. L., Grutter, A. S., and Cribb, T. H. (2008). Natatory-stage cymothoid isopods: description, molecular identification and evolution of attachment. *International Journal for Parasitology*, 38(3–4):477–491.
- Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., and Stadler, P. F. (2011). Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Research*, 40(7):2833–2845.
- Kaltz, O. and Shykoff, J. A. (1998). Local adaptation in host–parasite systems. *Heredity*, 81(4):361–370, DOI: 10.1046/j.1365-2540.1998.00435.x.

- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587.
- Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.
- Kayış, Ş. and Ceylan, Y. (2011). First report of *Nerocila orbigny* (Crustacea: Isopoda: Cymothoidae) on *Solea solea* (Teleostei: Soleidae) from Turkish Sea. *Turkish Journal of Fisheries and Aquatic Sciences*, 11(1):167–169.
- Kern, M., McGeehan, J. E., Streeter, S. D., Martin, R. N., Besser, K., Elias, L., Eborall, W., Malyon, G. P., Payne, C. M., Himmel, M. E., et al. (2013). Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences*, 110(25):10189–10194.
- Ketmaier, V., Joyce, D., Horton, T., and Mariani, S. (2008). A molecular phylogenetic framework for the evolution of parasitic strategies in cymothoid isopods (Crustacea). *Journal of Zoological Systematics and Evolutionary Research*, 46(1):19–23.
- Kilpert, F., Held, C., and Podsiadlowski, L. (2012). Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. *Molecular Phylogenetics and Evolution*, 64(1):106–117.
- Kilpert, F. and Podsiadlowski, L. (2006). The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea: Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics*, 7(1):241.

- Kilpert, F. and Podsiadlowski, L. (2010). The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 5(1):36–44.
- Klingenberg, C. P. (2016). Size, shape, and form: concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, 226(3):113–137.
- Knapp, M. and Hofreiter, M. (2010). Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes*, 1(2):227–243.
- Kück, P. and Longo, G. C. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, 11(1):81.
- Kück, P. and Struck, T. H. (2014). BaCoCa – A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Molecular Phylogenetics and Evolution*, 70:94–98.
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). Partition-Finder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3):772–773.
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278.
- Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC evolutionary biology*, 7(1):S4.
- Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology*, 62(4):611–615.

- Laslett, D. and Canbäck, B. (2007). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2):172–175.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362(6422):709–715.
- Linnaeus, C. (1758). *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Holmiae : Impensis Direct*, 10(ii):824.
- Lins, L. S., Ho, S. Y., and Lo, N. (2017). An evolutionary timescale for terrestrial isopods and a lack of molecular support for the monophyly of Oniscidea (Crustacea: Isopoda). *Organisms Diversity & Evolution*, 17(4):813–820.
- Lins, L. S., Ho, S. Y., Wilson, G. D., and Lo, N. (2012). Evidence for Permo-Triassic colonization of the deep sea by isopods. *Biology Letters*, 8(6):979–982.
- Liu, Y., Song, F., Jiang, P., Wilson, J.-J., Cai, W., and Li, H. (2018). Compositional heterogeneity in true bug mitochondrial phylogenomics. *Molecular Phylogenetics and Evolution*, 118:135–144.
- Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A., and Haak, W. (2017). From the field to the laboratory: controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research*, 3(1):1–14.

- Llewellyn, J. (1956). The host-specificity, micro-ecology, adhesive attitudes, and comparative morphology of some trematode gill parasites. *Journal of the Marine Biological Association of the United Kingdom*, 35(1):113–127.
- Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D., and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature Communications*, 10(1):2295.
- Mandeng, F. D. M., Bilong, C. F. B., Pariselle, A., Vanhove, M. P., Nyom, A. R. B., and Agnès, J.-F. (2015). A phylogeny of *Cichlidogyrus* spp. (Monogenea, Dactylogyridea) clarifies a host-switch between fish families and reveals an adaptive component to attachment organ morphology of this parasite genus. *Parasites & Vectors*, 8(1):582.
- Marcadé, I., Cordaux, R., Doublet, V., Debenest, C., Bouchon, D., and Raimond, R. (2007). Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). *Journal of Molecular Evolution*, 65(6):651–659.
- Marín, A. and Xia, X. (2008). GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: new substitution models incorporating strand bias. *Journal of Theoretical Biology*, 253(3):508–513.
- Marks, R. E., Juanes, F., Hare, J. A., and Conover, D. O. (1996). Occurrence and effect of the parasitic isopod, *Lironeca ovalis* (Isopoda: Cymothoidae), on young-of-the-year bluefish, *Pomatomus saltatrix* (Pisces: Pomatomidae). *Canadian Journal of Fisheries and Aquatic Sciences*, 53(9):2052–2057.
- Martin, M. B., Bruce, N. L., and Nowak, B. F. (2013). Redescription of *Ceratothoa carinata* (Bianconi, 1869) and *Ceratothoa oxyrrhynchaena* Koelbel, 1878 (Crustacea:

- Isopoda: Cymothoidae), buccal-attaching fish parasites new to Australia. *Zootaxa*, 3683:395–410, DOI: 10.11646/zootaxa.3683.4.4.
- Martin, M. B., Bruce, N. L., and Nowak, B. F. (2015). Review of the buccal-attaching fish parasite genus *Glossobius* Schioedte & Meinert, 1883 (Crustacea: Isopoda: Cymothoidae). *Zootaxa*, 3973(2):337–350.
- Maxwell, J. G. H. (1982). Infestation of the jack mackerel, *Trachurus declivis* (Jenyns), with the cymothoid isopod, *Ceratothoa imbricatus* (Fabricus), in south eastern Australian waters. *Journal of Fish Biology*, 20(3):341–349.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303.
- Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6):pdb-prot5448.
- Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J., Valencia, V., Ayala, C., Pérez-Angel, L., Rodríguez-Parra, L., Ramirez, V., et al. (2015). Middle Miocene closure of the Central American seaway. *Science*, 348(6231):226–229.
- Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 49(4):718–726.
- Moran, R., Morgan, C., and O’Connell, M. (2015). A guide to phylogenetic reconstruction using heterogeneous models — a case study from the root of the placental mammal tree. *Computation*, 3(2):177–196.

- Morgan, C. C., Creevey, C. J., and O'Connell, M. J. (2014). Mitochondrial data are not suitable for resolving placental mammal phylogeny. *Mammalian genome*, 25(11-12):636–647.
- Morton, B. (1974). Host specificity and position on the host in *Nerocila phaeopleura* Bleeker (Isopoda: Cymothoidae). *Crustaceana*, 26(2):143–148.
- Nagasawa, K. (2017). *Mothocya collettei* Bruce, 1986 (Isopoda: Cymothoidae), a marine fish parasite new to Japan. *Crustaceana*, 90(5):613–616.
- Nagler, C. and Haug, J. T. (2016). Functional morphology of parasitic isopods: understanding morphological adaptations of attachment and feeding structures in *Nerocila* as a pre-requisite for reconstructing the evolution of Cymothoidae. *PeerJ*, 4:21–88.
- Nagler, C., Hyžný, M., and Haug, J. T. (2017). 168 million years old "marine lice" and the evolution of parasitism within isopods. *BMC Evolutionary Biology*, 17(1):76.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1):268–274.
- Nielsen, E. E., Morgan, J., Maher, S., Edson, J., Gauthier, M., Pepperell, J., Holmes, B., Bennett, M., and Ovenden, J. (2017). Extracting DNA from 'jaws': high yield and quality from archived tiger shark (*Galeocerdo cuvier*) skeletal material. *Molecular Ecology Resources*, 17(3):431–442.
- Olsen, A. M. (2017). Feeding ecology is the primary driver of beak shape diversification in waterfowl. *Functional Ecology*, 31(10):1985–1995, DOI: 10.1111/1365-2435.12890.
- Östlund-Nilsson, S., Curtis, L., Nilsson, G. E., and Grutter, A. S. (2005). Parasitic isopod *Anilocra apogonae*, a drag for the cardinal fish *Cheilodipterus quinquelineatus*. *Marine Ecology Progress Series*, 287:209–216.

- Outomuro, D. and Johansson, F. (2017). A potential pitfall in studies of biological shape: does size matter? *Journal of Animal Ecology*, 86(6):1447–1457, DOI: 10.1111/1365-2656.12732.
- Palozzi, J. M., Jeedigunta, S. P., and Hurd, T. R. (2018). Mitochondrial DNA purifying selection in mammals and invertebrates. *Journal of Molecular Biology*, 430(24):4834–4848.
- Paradis, E. and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3):526–528.
- Pawluk, R. J., Ciampoli, M., and Mariani, S. (2015). Host size constrains growth patterns in both female and male *Ceratothoa italica*, a mouth-dwelling isopod. *Marine and Freshwater Research*, 66(4):381–384.
- Peccoud, J., Chebbi, M. A., Cormier, A., Moumen, B., Gilbert, C., Marcadé, I., Chandler, C., and Cordaux, R. (2017). Untangling heteroplasmy, structure, and evolution of an atypical mitochondrial genome by PacBio sequencing. *Genetics*, 207(1):269–280.
- Perna, N. T. and Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution*, 41(3):353–358.
- Philippe, H., Poustka, A. J., Chiodin, M., Hoff, K. J., Dessimoz, C., Tomiczek, B., Schiffer, P. H., Müller, S., Domman, D., Horn, M., et al. (2019). Mitigating anticipated effects of systematic errors supports sister-group relationship between Xenacoelomorpha and Ambulacraria. *Current Biology*, 29(11):1818–1826.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2018). *nlme: Linear and Nonlinear Mixed Effects Models*, <https://CRAN.R-project.org/package=nlme>. R package version 3.1-137.

- Poore, G. C. and Bruce, N. L. (2012). Global diversity of marine isopods (except Asellota and crustacean symbionts). *PLoS One*, 7(8):e43529.
- Poulin, R. (2007). Investing in attachment: evolution of anchoring structures in acanthocephalan parasites. *Biological Journal of the Linnean Society*, 90(4):637–645.
- Poulin, R. (2009). Interspecific allometry of morphological traits among trematode parasites: selection and constraints. *Biological Journal of the Linnean Society*, 96(3):533–540, DOI: 10.1111/j.1095-8312.2008.01163.x.
- Poulin, R., Krasnov, B. R., Mouillot, D., and Thieltges, D. W. (2011). The comparative ecology and biogeography of parasites. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366:2379–2390, DOI: 10.1098/rstb.2011.0048.
- Poulin, R. and Morand, S. (2000). The diversity of parasites. *The Quarterly Review of Biology*, 75(3):277–293.
- Quick, J. (2017). One-pot ligation protocol for oxford nanopore libraries. DOI: 10.17504/protocols.io.k9acz2e.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Raimond, R., Marcadé, I., Bouchon, D., Rigaud, T., Bossy, J.-P., and Souty-Grosset, C. (1999). Organization of the large mitochondrial genome in the isopod *armadillidium vulgare*. *Genetics*, 151(1):203–210.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5):901–904.

- Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, 1(4):319–329, DOI: 10.1111/j.2041-210X.2010.00044.x.
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223, DOI: 10.1111/j.2041-210X.2011.00169.x.
- Richardson, H. (1905). A monograph on the isopods of North America. *Bulletin of the United States National Museum*, (54):1–727.
- Richter, S. and Scholtz, G. (2001). Phylogenetic analysis of the Malacostraca (Crustacea). *Journal of Zoological Systematics and Evolutionary Research*, 39(3):113–136.
- Ripma, L. A., Simpson, M. G., and Hasenstab-Lehman, K. (2014). Geneious! Simplified genome skimming methods for phylogenetic systematic studies: a case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences*, 2(12):1400062.
- Rizzi, E., Lari, M., Gigli, E., de Bellis, G., and Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution*, 44(1):21.
- Roche, D. G., Strong, L. E., and Binning, S. A. (2013). Prevalence of the parasitic cymothoid isopod *Anilocra nemipteri* on its fish host at Lizard Island, Great Barrier Reef. *Australian Journal of Zoology*, 60(5):330–333.
- Rodríguez-González, A., Sarabeev, V., and Balbuena, J. A. (2017). Evolutionary morphology in shape and size of haptoral anchors in 14 *Ligophorus* spp. (Monogenea: Dactylogyridae). *PLoS ONE*, 12(5):e0178367.
- Rohland, N. and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols*, 2(7):1756.

- Rohlf, F. J. (2015). The tps series of software. *Hystrix the Italian Journal of Mammology*, 26(1):9–12.
- Rohlf, F. J. and Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1):40–59.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and 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.
- Rosenberg, M. S. and Kumar, S. (2003). Taxon sampling, bioinformatics, and phylogenomics. *Systematic Biology*, 52(1):119.
- Rota, J., Malm, T., Chazot, N., Peña, C., and Wahlberg, N. (2018). A simple method for data partitioning based on relative evolutionary rates. *PeerJ*, 6:e5498.
- Ruane, S. and Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, 17(5):1003–1008.
- Saclier, N., François, C. M., Konecny-Dupré, L., Lartillot, N., Guéguen, L., Duret, L., Malard, F., Douady, C. J., and Lefébure, T. (2018). Life history traits impact the nuclear rate of substitution but not the mitochondrial rate in isopods. *Molecular Biology and Evolution*, 35(12):2900–2912.
- Sala-Bozano, M., van Oosterhout, C., and Mariani, S. (2012). Impact of a mouth parasite in a marine fish differs between geographical areas. *Biological Journal of the Linnean Society*, 105(4):842–852.

- Sanderson, M., Wojciechowski, M., Hu, J.-M., Khan, T. S., and Brady, S. (2000). Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution*, 17(5):782–797.
- Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, 19(1):101–109.
- Saravanakumar, A., Balasubramanian, T., Raja, K., and Trilles, J.-P. (2012). A massive infestation of sea snakes by cymothoid isopods. *Parasitology Research*, 110(6):2529–2531.
- Schioedte, J. and Meinert, F. (1881). *Symbolae ad monographium cymothoarum crustaceorum isopodum familiae*. II. Anilocridae. *Naturhistorisk Tidsskrift*, 13(1-166):1–10.
- Schioedte, J. and Meinert, F. (1883). *Symbolae ad monographium cymothoarum crustaceorum isopodem familiae*. III. Saophridae. IV. Ceratothoinae. *Naturhistorisk Tidsskrift*, 13(281-378):11–16.
- Schmidt, C. (2008). Phylogeny of the terrestrial Isopoda (Oniscidea): a review. *Arthropod Systematics & Phylogeny*, 66(2):191–226.
- Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–864.
- Schram, F. R. (1970). Isopod from the Pennsylvanian of Illinois. *Science*, 169(3948):854–855.
- Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A., Willerslev, E., Krogh, A., and Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, 13(1):178.

- Seibt, K. M., Schmidt, T., and Heitkam, T. (2018). FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. *Bioinformatics*, 34(20):3575–3577.
- Sfendourakis, S. and Taiti, S. (2015). Patterns of taxonomic diversity among terrestrial isopods. *ZooKeys*, 515:13–25, DOI: 10.3897/zookeys.515.9332.
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, 11(10):e0163962.
- Shen, Y., Kou, Q., Zhong, Z., Li, X., He, L., He, S., and Gan, X. (2017). The first complete mitogenome of the South China deep-sea giant isopod *Bathynomus* sp. (Crustacea: Isopoda: Cirolanidae) allows insights into the early mitogenomic evolution of isopods. *Ecology and evolution*, 7(6):1869–1881.
- Sherrill-Mix, S. (2019). taxonomizr: functions to work with NCBI accessions and taxonomy. <https://CRAN.R-project.org/package=taxonomizr>.
- Shimodaira, H. and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16(8):1114–1114.
- Shokralla, S., Singer, G. A., and Hajibabaei, M. (2010). Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *Biotechniques*, 48(3):305–306.
- Sibley, T. (2018). simple-consensus-per-read-group. <https://github.com/MullinsLab/simple-consensus-per-read-group>.
- Smit, N. J., Bruce, N. L., and Hadfield, K. A. (2014). Global diversity of fish parasitic isopod crustaceans of the family Cymothoidae. *International Journal for Parasitology: Parasites and Wildlife*, 3(2):188–197, DOI: 10.1016/j.ijppaw.2014.03.004.

- Sota, T. and Vogler, A. P. (2001). Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Systematic Biology*, 50(1):39–59.
- Spears, T., DeBry, R. W., Abele, L. G., and Chodyla, K. (2005). Peracarid monophyly and interordinal phylogeny inferred from nuclear small-subunit ribosomal DNA sequences (Crustacea: Malacostraca: Peracarida). *Proceedings of the Biological Society of Washington*, 118(1):117–158.
- Sproul, J. S. and Maddison, D. R. (2017). Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*, 17(6):1183–1201.
- St. John, J. (2011). Seqprep. <https://github.com/jstjohn/SeqPrep>.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313.
- Susko, E. and Roger, A. J. (2007). On reduced amino acid alphabets for phylogenetic inference. *Molecular Biology and Evolution*, 24(9):2139–2150.
- Symonds, M. R. E. and Blomberg, S. P. (2014). A primer on Phylogenetic Generalised Least Squares. In L. Z. Garamszegi, editor, *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*, chapter 5, pages 105–130. Springer-Verlag, Berlin, DOI: 10.1007/978-3-662-43550-2.
- Taberner, R. (1998). *Isonebula acanthopleon* sp. n. (Isopoda: Cymothoidae) nuevo ectoparásito de peces curimátidos. *Revista Brasileira de Zoologia*, 15(2):297–305, DOI: 10.1590/S0101-81751998000200004.
- Talavera, G. and Vila, R. (2011). What is the phylogenetic signal limit from mitogenomes? the reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evolutionary Biology*, 11(1):315.

- Tanabe, A. (2008). MrBayes5d. <http://fifthdimension.jp>.
- Tarrío, R., Rodríguez-Trelles, F., and Ayala, F. J. (2001). Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Molecular Biology and Evolution*, 18(8):1464–1473.
- Tavares-Dias, M., Araújo, C. S. O., Barros, M. S., and Viana, G. M. (2014). New hosts and distribution records of *Braga patagonica*, a parasite (Cymothoidae) of fishes from the amazon. *Brazilian Journal of Aquatic Science and Technology*, 18(1):91–97, DOI: 10.14210/bjast.v18n1.p91-97.
- Thatcher, V. (1988). *Asotana magnifica* n. sp. (Isopoda: Cymothoidae) an unusual parasite (commensal?) of the buccal cavities of piranhas (*Serrasalmus* sp.) from Roraima, Brazil. *Amazoniana*, 10:239–248.
- Thatcher, V. (2006). *Amazon Fish Parasites*. Pensoft, Sofia.
- Thatcher, V. E. (2002). *Anphira guianensis* sp. nov. (Isopoda, Cymothoidae) from the gills of *Acnodon oligacanthus* eigenmann (Pisces, Serrasalminidae) of French Guiana. *Revista Brasileira de Zoologia*, 19:53–59.
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2):178–192.
- Timmermans, M. J., Barton, C., Haran, J., Ahrens, D., Culverwell, C. L., Ollikainen, A., Dodsworth, S., Foster, P. G., Bocak, L., and Vogler, A. P. (2015). Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Genome Biology and Evolution*, 8(1):161–175.

- Tin, M. M.-Y., Economo, E. P., and Mikheyev, A. S. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One*, 9(5):e96793.
- Trilles, J.-P. (2007). *Olencira praegustator* (Crustacea: Isopoda: Cymothoidae) parasitic on *Brevoortia* species (Pisces: Clupeidae) from the southeastern coasts of North America: review and re-description. *Marine Biology Research*, 3(5):296–311.
- Trilles, J.-P. and Öktener, A. (2004). *Livoneca sinuata* (Crustacea: Isopoda: Cymothoidae) on *Loligo vulgaris* from Turkey, and unusual cymothoid associations. *Diseases of Aquatic Organisms*, 61(3):235–240.
- Uribe, J. E., Irisarri, I., Templado, J., and Zardoya, R. (2019). New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Molecular Phylogenetics and Evolution*, 133:12–23.
- van der Wal, S., Smit, N. J., and Hadfield, K. A. (2017). Redescription and molecular characterisation of the fish parasitic isopod *Norileca indica* (Milne Edwards, 1840) (Crustacea: Isopoda: Cymothoidae) with a key to the genus. *African Zoology*, 52(3):163–175.
- van der Wal, S., Smit, N. J., and Hadfield, K. A. (2019). Review of the fish parasitic genus *Elthusa* Schioedte & Meinert, 1884 (Crustacea: Isopoda: Cymothoidae) from South Africa, including the description of three new species. *ZooKeys*, 841:1.
- van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, 1:1–30.
- Varvarigos, P. (2003). Parasitic isopods (suborder Flabellifera) affecting the farmed marine fish in Greece, with special reference to *Ceratothoa oestroides* (family Cymothoidae). [http://www.vetcare.gr/ARTPRES/Pathogenic\\_isopoda.htm](http://www.vetcare.gr/ARTPRES/Pathogenic_isopoda.htm).

- Vignon, M., Pariselle, A., and Vanhove, M. P. (2011). Modularity in attachment organs of African *Cichlidogyrus* (Platyhelminthes: Monogenea: Ancyrocephalidae) reflects phylogeny rather than host specificity or geographic distribution. *Biological Journal of the Linnean Society*, 102(3):694–706, DOI: 10.1111/j.1095-8312.2010.01607.x.
- Vignon, M. and Sasal, P. (2010). The use of geometric morphometrics in understanding shape variability of sclerotized haptor structures of monogeneans (Platyhelminthes) with insights into biogeographic variability. *Parasitology International*, 59(2):183–191.
- Wägele, J. W. (1989). Evolution und phylogenetisches System der Isopoda. *Zoologica*, 140:1–262.
- Wall, A. R., Campo, D., and Wetzer, R. (2014). Genetic utility of natural history museum specimens: endangered fairy shrimp (Branchiopoda: Anostraca). *ZooKeys*, 457:1.
- Wei, S.-J., Shi, M., Chen, X.-X., Sharkey, M. J., van Achterberg, C., Ye, G.-Y., and He, J.-H. (2010). New views on strand asymmetry in insect mitochondrial genomes. *PLoS One*, 5(9):e12708.
- Welicky, R., Parkyn, D., and Sikkel, P. (2018). Host-dependent differences in measures of condition associated with *Anilocra* spp. parasitism in two coral reef fishes. *Environmental Biology of Fishes*, 101(8):1223–1234.
- Welicky, R. L., Hadfield, K. A., Sikkel, P. C., and Smit, N. J. (2017). Molecular assessment of three species of *Anilocra* (Isopoda: Cymothoidae) ectoparasites from Caribbean coral reef fishes, with the description of *Anilocra brillae* sp. n. *ZooKeys*, 663:21–43.
- Welicky, R. L., Malherbe, W., Hadfield, K. A., and Smit, N. J. (2019). Understanding growth relationships of African cymothoid fish parasitic isopods using specimens from museum and field collections. *International Journal for Parasitology: Parasites and Wildlife*, 8:182–187.

- Welicky, R. L. and Sikkel, P. C. (2014). Variation in occurrence of the fish-parasitic cymothoid isopod, *Anilocra haemuli*, infecting french grunt (*Haemulon flavolineatum*) in the north-eastern Caribbean. *Marine and Freshwater Research*, 65(11):1018–1026.
- Welicky, R. L. and Smit, N. J. (2019). Redescription and molecular characterisation of the fish ectoparasite *Anilocra capensis* Leach, 1818 (Isopoda: Cymothoidae), with description of six new species of *Anilocra* Leach, 1818 from Africa. *Parasites and Vectors*, 12(1):387, DOI: 10.1186/s13071-019-3578-5.
- Wetzer, R. (2002). Mitochondrial genes and isopod phylogeny (Peracarida: Isopoda). *Journal of Crustacean Biology*, 22(1):1–14.
- Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., and Wagner, L. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Research*, 31(1):28–33.
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, 13(6):e1005595, DOI: 10.1371/journal.pcbi.1005595.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- Williams, E. H., Bunkley-Williams, L., and Ebert, D. (2010). An accidental attachment of *Elthusa raynaudii* (Isopoda: Cymothoidae) in *Etmopterus* sp. (Squaliformes: Etmopteridae). *Acta Parasitologica*, 55(1):99–101.
- Williams Jr., E. H. and Bunkley-Williams, L. (1994). Four cases of unusual crustacean-fish associations and comments on parasitic processes. *Journal of Aquatic Animal Health*, 6(3):202–208.

- Wilson, G. D. (2007). Global diversity of Isopod crustaceans (Crustacea: Isopoda) in freshwater. *Hydrobiologia*, 595(1):231–240, DOI: 10.1007/s10750-007-9019-z.
- Wilson, G. D. (2009). The phylogenetic position of the Isopoda in the Peracarida (Crustacea: Malacostraca). *Arthropod Systematics & Phylogeny*, 67(2):159–198.
- Winter, D. J. (2017). rentrez: an R package for the NCBI eUtils API. *The R Journal*, 9(2):520–526.
- Xu, W., Jameson, D., Tang, B., and Higgs, P. G. (2006). The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *Journal of Molecular Evolution*, 63(3):375–392.
- Yamano, H., Yamauchi, T., and Hosoya, K. (2011). A new host record of *Ichthyoxenus amurensis* (Crustacea: Isopoda: Cymothoidae) from the amur bitterling *Rhodeus sericeus* (Cypriniformes: Cyprinidae). *Limnology*, 12(1):103–106.
- Yamauchi, T. (2009). Deep-sea cymothoid isopods (Crustacea: Isopoda: Cymothoidae) of Pacific coast of Northern Honshu, Japan. *National Museum of Nature and Science Monographs*, 39:467–481.
- Yu, J., An, J., Li, Y., and Boyko, C. B. (2018). The first complete mitochondrial genome of a parasitic isopod supports Epicaridea Latreille, 1825 as a suborder and reveals the less conservative genome of isopods. *Systematic Parasitology*, 95(5):465–478.
- Zardoya, R. and Meyer, A. (2001). Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rDNA sequence data due to an adverse phylogenetic signal/noise ratio. *Systematics Association Special Volume*, 61:135–155.
- Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B. G., and Besnard, G. (2015). Museomics illuminate the history of an extinct, paleoendemic plant lineage

- (Hesperelaea: Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, 117(1):44–57.
- Zimmermann, J., Hajibabaei, M., Blackburn, D. C., Hanken, J., Cantin, E., Posfai, J., and Evans, T. C. (2008). DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology*, 5(1):18.
- Zou, H., Jakovlić, I., Zhang, D., Chen, R., Mahboob, S., Al-Ghanim, K. A., Al-Misned, F., Li, W.-X., and Wang, G.-T. (2018). The complete mitochondrial genome of *Cymothoa indica* has a highly rearranged gene order and clusters at the very base of the Isopoda clade. *PLoS one*, 13(9):e0203089.
- Zuguang, G., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). *circlize* implements and enhances circular visualization in R. *Bioinformatics*, 30(19):2811–2812, DOI: doi:10.1093/bioinformatics/btu393.

# Appendix

## ML Partitioned (nucleotides)

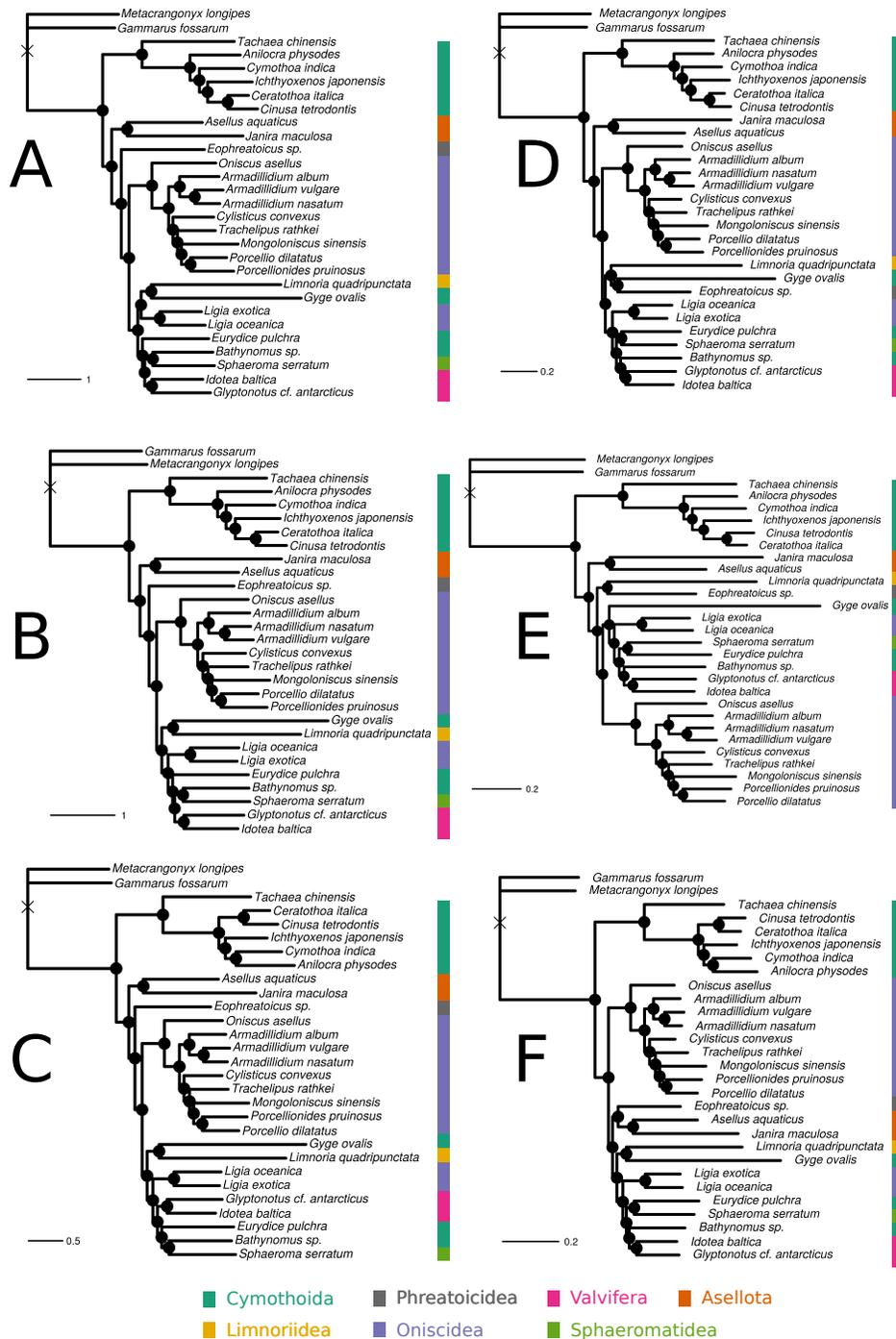


Figure 1: ML partitioned trees (nucleotides). A = PCG-123, B = PCG-123+12S+16S, C = PCG-123+12S+16S+18S, D = PCG-3RY, E = PCG-3RY+12S+16S, F = PCG-3RY+12S+16S+18S. Black circles are bootstrap support values >0.9, grey <0.9 & >0.7, and white <0.7.

## BI Partitioned (nucleotides)

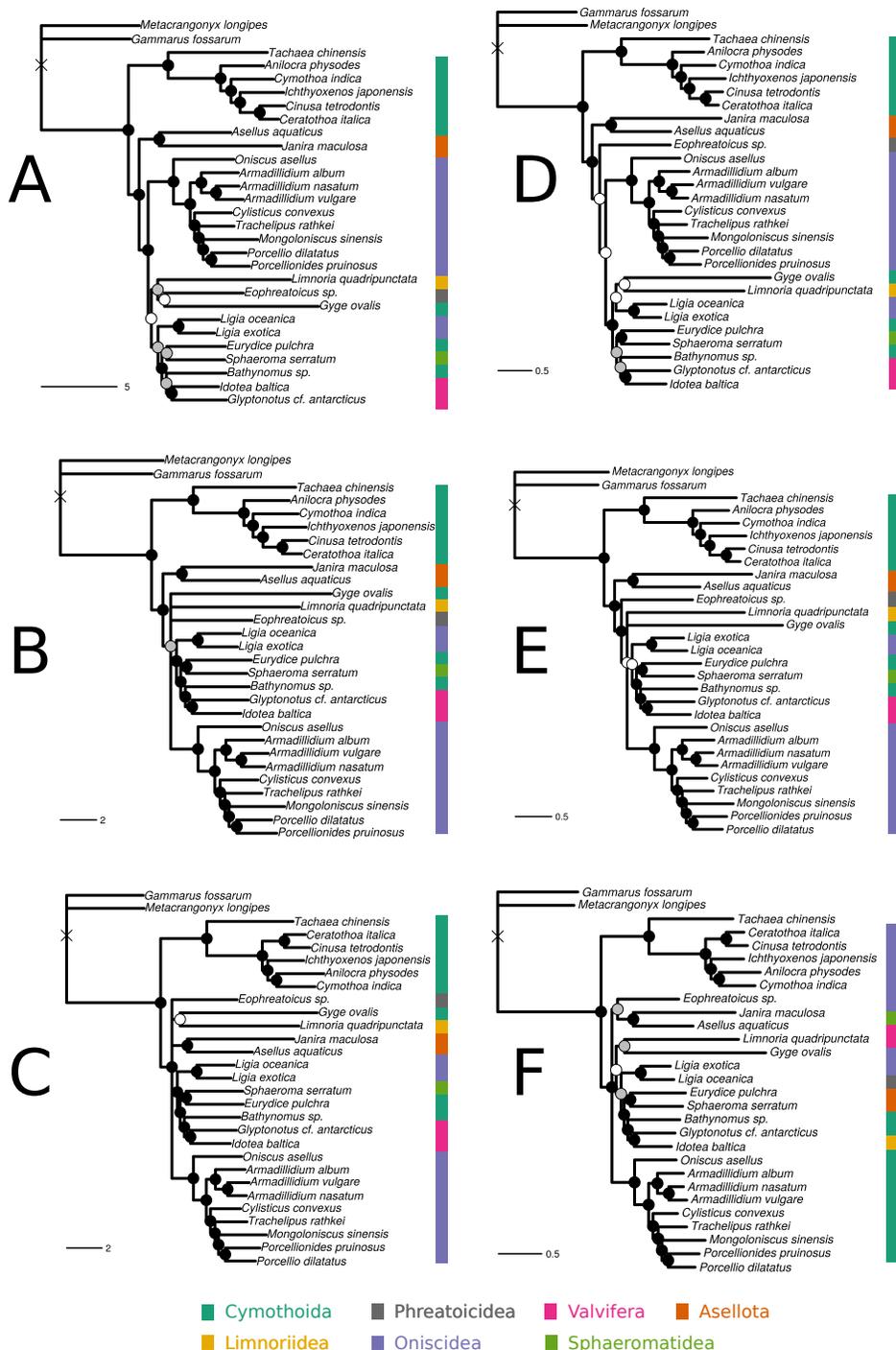


Figure 2: BI partitioned trees (nucleotides). A = PCG-123, B = PCG-123+12S+16S, C = PCG-123+12S+16S+18S, D = PCG-3RY, E = PCG-3RY+12S+16S, F = PCG-3RY+12S+16S+18S. Black circles are Bayesian Posterior Probabilities (BPP) > 0.9, grey is BPP < 0.9 & > 0.7, and is white BPP < 0.7.

## Partitioned (amino acids)

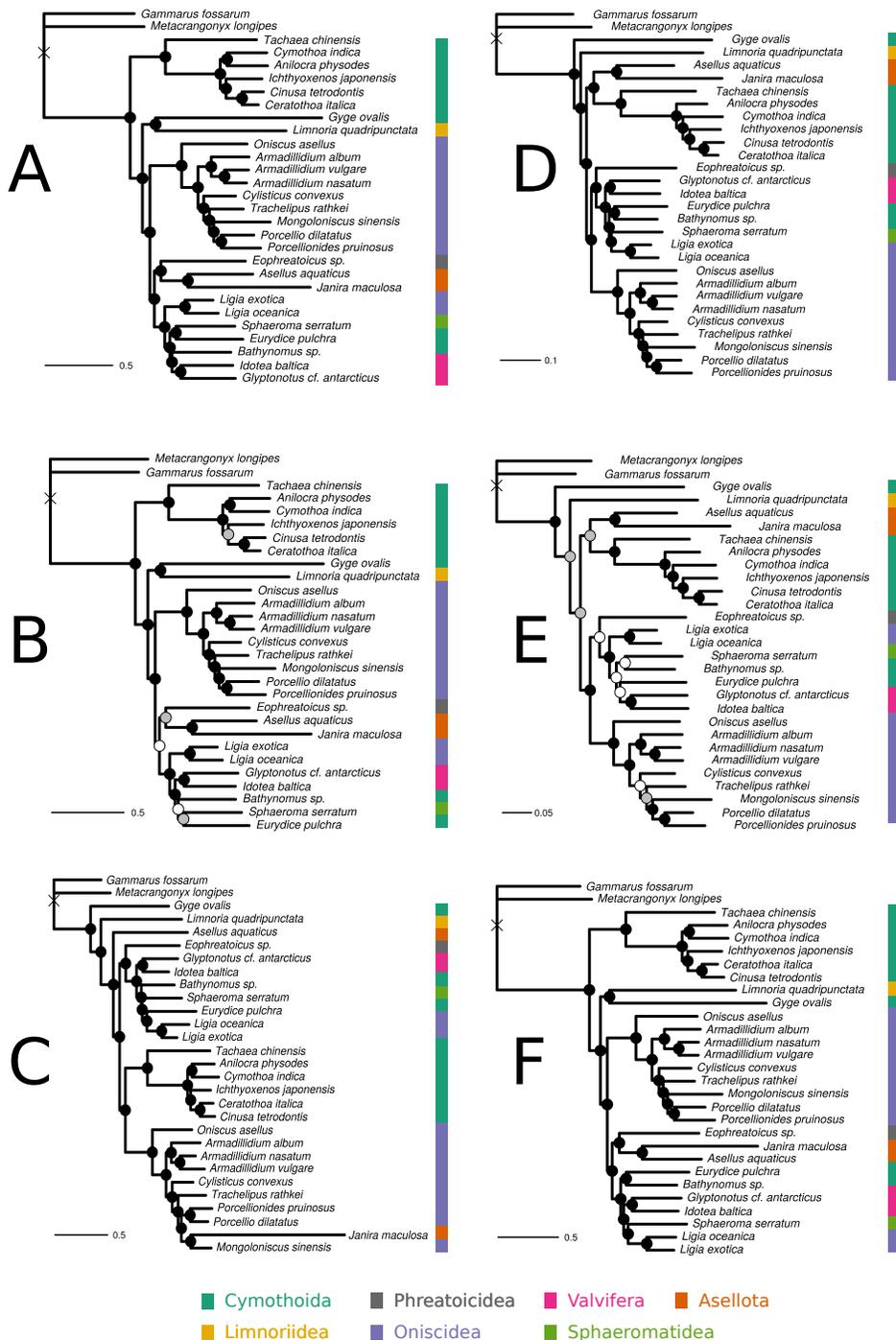


Figure 3: Amino Acid Partitioned Trees. A = PCG-AA (ML), B = PCG-AA (BI), C = PCG-AA-R, D = PCG-DAY (ML), E = PCG-DAY (BI), F = PCG-DAY-R. Black circles are bootstrap supports (for ML analyses) or Bayesian Posterior Probabilities (BPP) >0.9, grey is BPP <0.9 & >0.7, and is white BPP <0.7.

# CATGTR (site-heterogeneous)

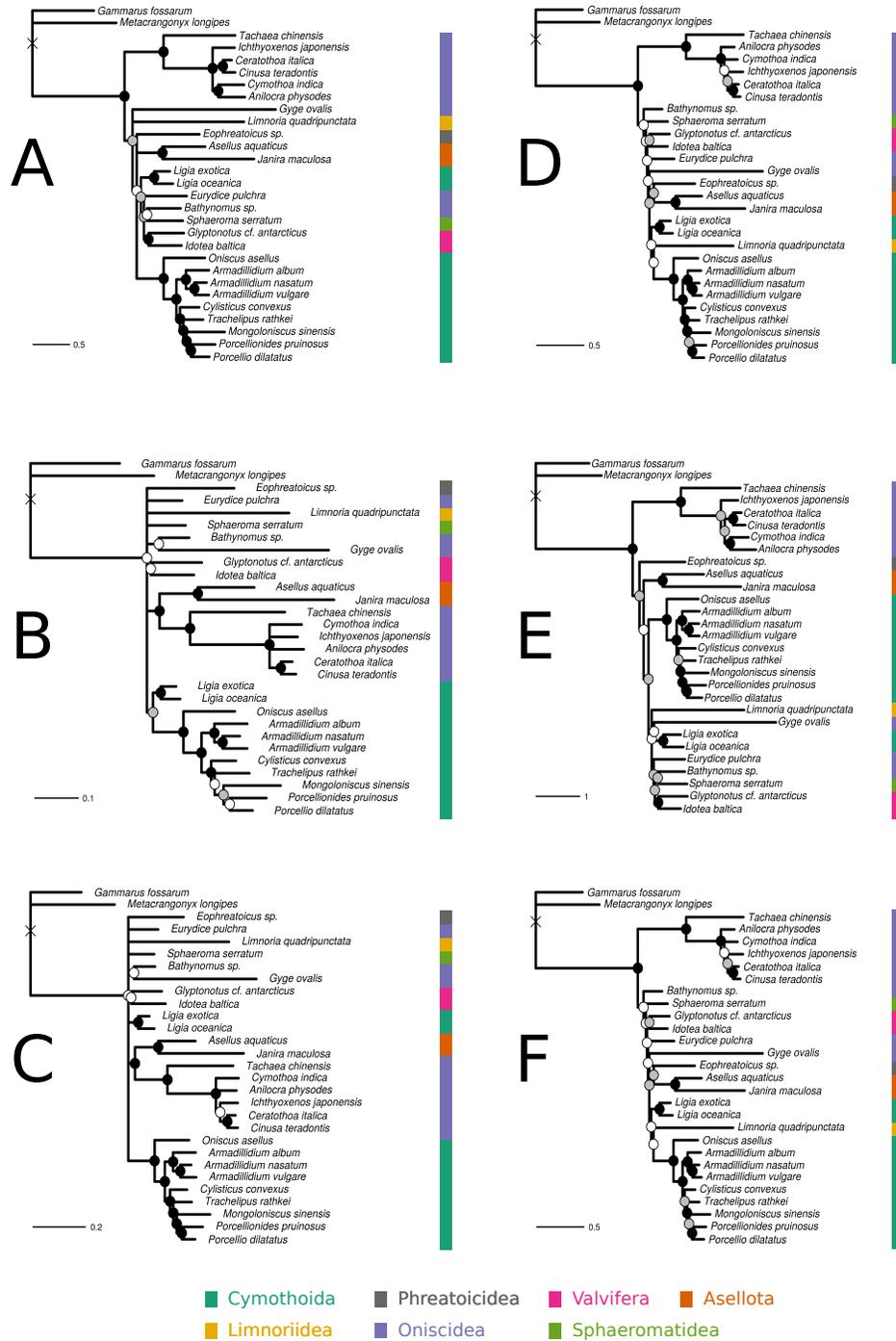


Figure 4: Site-heterogeneous Trees. A = PCG-AA, B = PCG-HOMO, C = PCG-DAY, D = PCG-123, E = PCG-12R, F = PCG-12X. Black circles are Bayesian Posterior Probabilities (BPP) >0.9, grey is BPP <0.9 & >0.7, and is white BPP <0.7.