Limited transmission of microbial species among coral reef fishes from the Great Barrier Reef, Australia

Vincenzo A. Costa[1], David R. Bellwood[2], Jonathon C.O. Mifsud[1], Jemma L. Geoghegan[3,4], Erin Harvey[1], Edward C. Holmes[1]*

[1]Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

[2]Research Hub for Coral Reef Ecosystem Functions, College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia.

[3]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand.

[4]Institute of Environmental Science and Research, Kenepuru, Porirua, 5022, New Zealand.

*Corresponding author:

Prof. Edward C. Holmes,

Sydney Institute for Infectious Diseases, School of Medical Sciences,

The University of Sydney, Sydney, NSW 2006, Australia.

Email: edward.holmes@sydney.edu.au

Abstract

Reef fishes account for one-third of all extant marine fishes and exhibit enormous biodiversity within a highly interactive ecosystem. Yet relatively little is known about the diversity and evolution of microbial species (bacteria, viruses, and eukaryotes) associated with reef fish, even though this may provide valuable insights into the factors that shape microbial communities within vertebrate hosts as well as the extent and pattern of cross-species transmission. Through metatranscriptomic sequencing we characterised the viruses, bacteria, and single-celled eukaryotes from 128 reef fish species inhabiting Lizard Island and Orpheus Island on the Great Barrier Reef, Australia. We assessed whether microbial communities differed between islands that are separated by approximately 450 kilometres, and to what extent viruses were able to emerge in new hosts. Notably, despite strong ecological interactions in the reef environment, and the presence of the same families and subfamilies of viruses and bacteria on both islands, there was minimal evidence for the cross-species transmission of individual microorganisms among fish species. An exception was the high prevalence of the bacterial pathogen *Photobacterium damselae* among apparently healthy cardinalfishes from both islands, indicating that these fish species are natural reservoirs within the reef system. Overall, these data suggest that reef fishes have microbial-host associations that arose prior to the formation of the Great Barrier Reef, leading to strong host barriers to cross-species microbial transmission even within a highly interactive and species-rich environment.

Introduction

Symbiotic interactions are ubiquitous in nature and play an important role in animal, plant, and microbial evolution. In reef systems, the interdependence of corals, fishes, and their microbial symbionts supports fundamental ecological processes, and their dissociation may cause devastating declines in species abundance and reef functioning (1). Healthy tropical coral reefs are characterised by remarkable biodiversity and exceptionally complex interactions. This provides an ideal forum for investigating the evolutionary and ecological factors shaping microbial communities, especially among teleost fishes. Teleost fishes are the most speciose group of vertebrates on coral reefs and rank among the most phylogenetically and ecologically diverse group of vertebrates, accounting for one-third of all currently described marine fishes (2, 3). Reef fishes exhibit exceptional dispersal capabilities. Most have geographic ranges spanning thousands of kilometres, with some species spanning approximately two thirds of the global tropics (4). This level of interconnectivity is also exhibited on individual reefs, with highly complex food webs. Reef fishes display diverse trophic guilds (e.g., carnivores, mobile invertivores, omnivores, planktivores, sessile invertivores, herbivores/detritivores), occur in a range of habitats (e.g., coral, sand, rubble, caves), and commonly live in exceptionally close proximity (5-7).

This high degree of interaction is exemplified in the cryptobenthic reef fishes: behaviourally cryptic species with adult body sizes of approximately 5 cm or less that typically occupy the benthic zone (6). Cryptobenthic reef fishes engage with larger reef fishes through extensive predator-prey interactions, and their frequent consumption is an integral component of coral reef food webs via the transfer of energy from microscopic prey to large predators (8). For example, the dwarf goby (*Eviota sigillata*) has a maximum lifespan of just 59 days and experiences mortality rates of 7.8% per day (9). In addition to predation, fishes interact through cleaning (including the removal of blood-sucking parasites) (10) and extensive food

webs of coprophagy (consuming the faeces of other fishes) (11). The potential for microbial transmission on reefs is therefore considerable.

Although reef fish co-exist in a highly diverse and interactive ecosystem, relatively little is known about whether and how microbial composition (i.e., of viruses, bacteria, eukaryotes) differs across fish groups within and among communities. Studies of microbial ecology in teleosts have largely focused on animals utilised in aquaculture and in laboratory model species, with very few investigations of wild ecosystems (12-15). Fish-microbe interactions are highly beneficial for fish nutrition and immunity and are shaped by host and environmental factors including trophic level, age, water quality, and host phylogeny (12, 16-20).

The extent of phylogenetic divergence between animal species has an important impact on virus ecology and evolution, particularly the likelihood of successful cross-species virus transmission, and it is therefore a key determinant of infectious disease emergence (21-25). The “phylogenetic distance” theory posits that microorganisms are more likely to be transmitted between closely related species that have conserved cellular properties, such as cell receptors (26). This idea is supported by numerous studies across a broad spectrum of host taxa (e.g., vertebrates, invertebrates) and pathogen groups including virus-host interactions in reef fishes (16, 22). For example, recent work has shown that reef fishes from a spatially restricted (100 square metre) community from Orpheus Island in the Great Barrier Reef (GBR), Australia, harbour diverse viral assemblages that are highly host-specific despite ample opportunity for cross-species transmission (16).

With approximately 2500 coral reefs and 900 islands, fishes of the GBR constitute a natural model system to investigate spatial patterns of microbial evolution and diversity in vertebrate hosts. Molecular and fossil evidence indicates that the majority of reef fish families originated during the Paleocene and Eocene, approximately 66 to 50 million years ago (Ma).

Subsequently, a notable acceleration in lineage diversification took place during the Oligocene and Miocene (34–5.3 Ma), with this shift occurring in the Indo-Australia Archipelago (IAA) (i.e., the IAA biodiversity hotspot) (27). During the Miocene (23-5.3 Ma), the reciprocal diversification of fish and coral species likely led to the development of the functional reef ecosystems that are observable today. By the early stages of the Pleistocene (5.3-0 Ma), almost all reef fish taxonomic groups were established and began their settlement on reefs across all tropical oceans (27), with the formation of the GBR fish communities likely occurring within the last 10,000 years following the stabilization of sea level to its current height ~6000-8000 years ago (5, 28). Whether and how these colonization events have shaped microbial evolution and diversity is unknown. For example, many reef fishes—particularly cryptobenthics that have limited dispersal—exhibit strong site fidelity, maintaining the same community composition year-round (29). The seemingly consistent community composition might therefore lead to the generation of distinct microbial communities in different geographic areas, that will be most pronounced for rapidly evolving RNA viruses. Conversely, it is possible that the high ecological similarities among disjunct reef locations might result in broadly similar microbial compositions among fish communities.

We characterised the total assemblage of viruses, bacteria, and single-celled eukaryotes from 128 reef fish species spread across Lizard Island and Orpheus Island in the GBR. This study comprised 28 reef fish families, making it one of the largest investigations of microbial diversity and evolution in reef fish undertaken to date. Using metatranscriptomics, we aimed to determine the relationship between host community diversity and microbial diversity and identify whether these communities differ between fish species from two islands separated by approximately 450 kilometres. We also aimed to determine the impact of host ecology on microbial diversity and evolution and to identify whether particular fish groups are potential

reservoirs for important viral or bacterial pathogens. This is of particular importance given the high utilisation of reef fish in aquaculture as well as the ongoing threats of biodiversity loss on coral reefs from anthropogenic climate change, itself a significant contributor to the emergence of infectious diseases (30).

Results

Composition of sequence reads

We sequenced a total of 10.7 billion RNA reads, including 4.7 billion reads newly generated from Lizard Island fishes. The remaining reads (Orpheus Island) are available on NCBI Sequence Read Archive (SRA) under BioProject PRJNA841039 (16). These data were generated from a total of 140 sequencing libraries, representing 128 reef fish species and 28 families (Figure 1) (mean 54,617,146 reads per library). Fish RNA accounted for 93% of the total reads, followed by RNA associated with cnidarians (5.7%), bacteria (0.31%), single-celled eukaryotes (0.29%), arthropods (0.23%), platyhelminths (0.18%), molluscs (0.04%), poriferans (0.03%), annelids (0.02%), nematodes (0.01%), fungi (0.009%), and viruses (0.009%) (Supplementary Table 1).

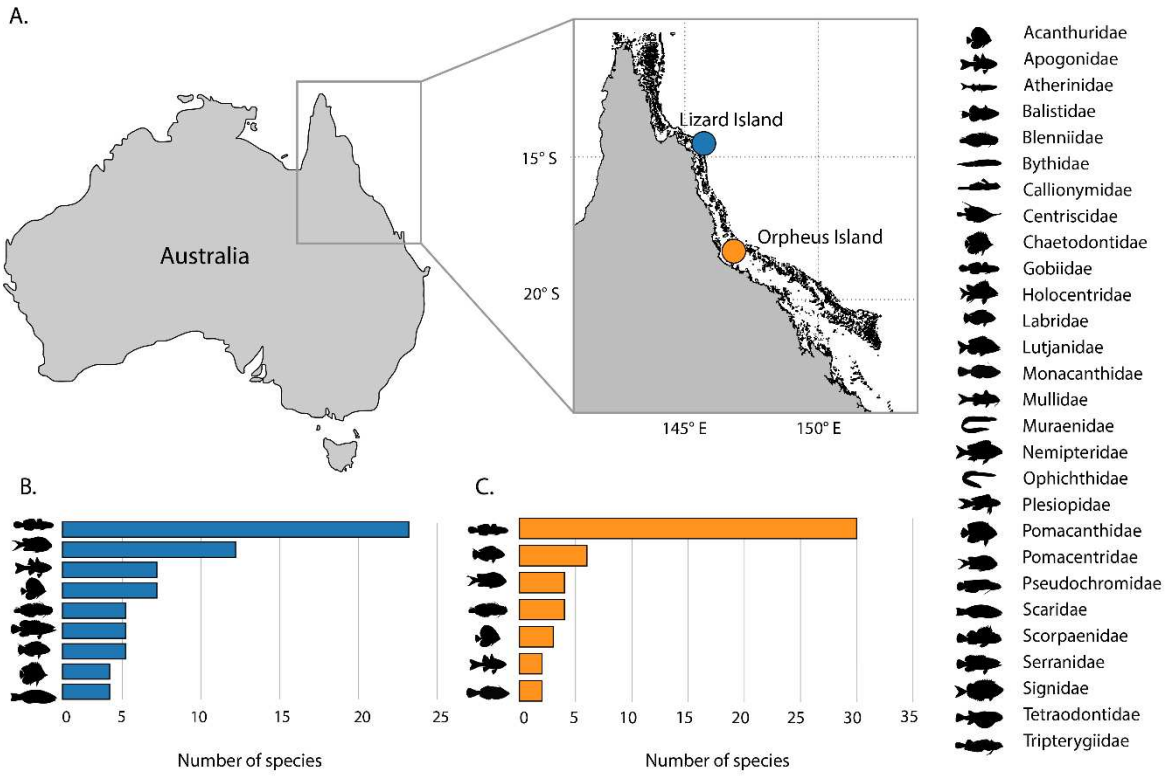


Figure 1. Sampling locations and taxonomic diversity of reef fish. (A) Location of Lizard Island and Orpheus Island in the GBR. (B) Taxonomy of samples collected from Lizard Island. (C) Taxonomy of samples collected from Orpheus Island. All other families with less than one species sampled are omitted from panels B and C.

Diversity and abundance of the reef fish virome

We identified sequences representing 64 vertebrate-associated viruses (i.e., those likely infecting fish tissues), including 27 newly discovered from Lizard Island fishes (Supplementary Table 2). The *Astroviridae* comprised 32.4% of vertebrate-associated viral reads, followed by the *Iridoviridae* (21.8%), *Picornaviridae* (18.6%), *Chuviridae* (15.9%), *Parvoviridae* (6.8%), *Hantaviridae* (2.9%) with all other groups representing <1% of the total viral reads: *Flaviviridae*, *Orthomyxoviridae (order Articulavirales)*, *Poxviridae*,

Paramyxoviridae, *Reoviridae*, *Circoviridae*, *Coronaviridae*, *Hepeviridae*, *Rhabdoviridae*, and *Caliciviridae* (Figure 2).

As well as vertebrate-associated viruses, we identified 194 viruses that were likely infecting porifera, arthropods, molluscs, fungi, plants, and microbial eukaryotes (e.g., dinoflagellates), including 100 that were newly discovered at Lizard Island. Because these viruses were likely associated with fish diet and not infecting the fish themselves, we assume that their presence does not reflect key aspects of fish biology (i.e., immunity or receptor binding) and hence effectively serve as a "negative control" in comparison to the vertebrate-associated viruses. We refer to this group as "non-vertebrate" viruses. The most abundant viral groups in this category were the *Flaviviridae* (24% of non-vertebrate viral reads), unclassified picornaviruses (i.e. "picorna-like" viruses) (22.5%) *Narnaviridae* (21.9%), *Nodaviridae* (8.9%), *Hepeviridae* (6.2%), *Partitiviridae* (4%), *Solemoviridae* (3.2%), *Negevirus* (2.1%) and *Totiviridae* (2%), with all other groups comprising <1%: *Reovirales*, *Weivirus*, *Qinviridae*, *Iflaviridae*, *Dicistroviridae*, *Rhabdoviridae*, *Picobirnaviridae*, *Quenyavirus*, *Bunyavirales*, *Chuviridae*, and *Tombusviridae*.

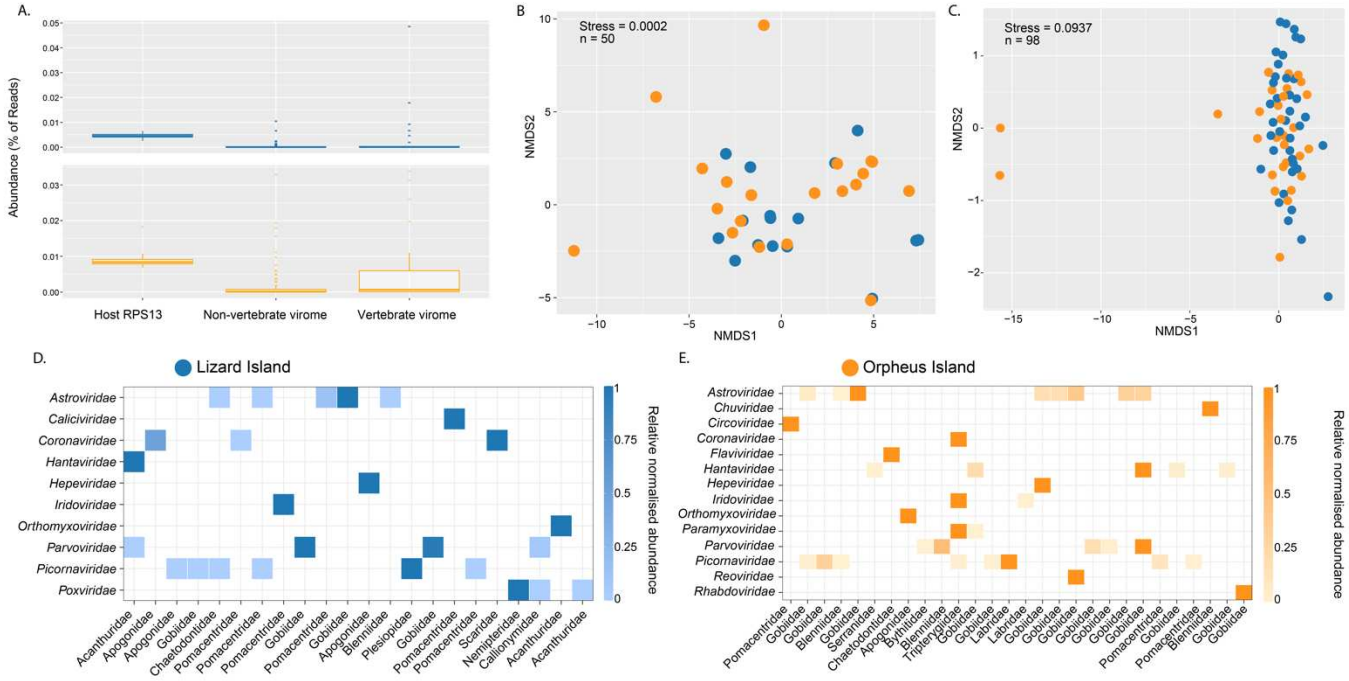
Spatial comparisons of the reef fish virome

To determine whether viral composition (i.e., virus families or subfamilies) differed between Orpheus and Lizard islands, we analysed beta diversity using permutational multivariate analysis of variance (PERMANOVA) with the Bray–Curtis dissimilarity matrix. This revealed no significant difference in vertebrate-associated viral communities on both islands ($F = 1.42$, $p = 0.06$), with overlapping viromes at the viral family/subfamily level (Figure 2b). Accordingly, both islands contained viruses assigned to the *Picornaviridae*, *Astroviridae*, *Parvoviridae*, *Hantaviridae*, *Orthomyxoviridae*, *Coronaviridae*, *Hepeviridae* and *Iridoviridae* (Figure 2d-e). A similar pattern was observed when analysing the non-vertebrate virome (F =

169 1.53, $p = 0.07$) (Figure 2c) with both islands dominated by unclassified picornaviruses,

170 *Narnaviridae*, *Totiviridae*, *Partitiviridae*, and *Nodaviridae*.

171



172

173 **Figure 2. (**A) Abundance of host gene marker (RPS13), non-vertebrate virome and vertebrate

174 virome. (B) Non-metric multi-dimensional scaling (NMDS) plot (Bray-Curtis dissimilarity

175 matrix) for vertebrate-associated viruses for each fish species. (C) NMDS plot for non-

176 vertebrate-associated viruses for each fish species. (D-E) Relative normalised abundance of

177 vertebrate-associated viral families from each island.

178

179 **Evolutionary history and biogeographical patterns of vertebrate-associated viruses**

180 While our analysis revealed similarities at the level of virus family/subfamily, almost all of

181 the viruses identified exhibited levels of genetic divergence that reflected long-term virus-

182 host associations, rather than recent cross-species transmission within each reef ecosystem.

183 The only instance of the same virus being shared between fish species was the presence of

184 highly similar astroviruses (~96 similarity across the entire genome) in gobies from Orpheus
185 Island (see ref. 16 for full description of these viruses). No viruses were shared among the
186 fishes sampled from Lizard Island.

187 We now describe the phylogenetic relationships of the shared viral groups in turn. We focus
188 primarily on the relationships of viruses between both islands as well as those newly
189 discovered at Lizard Island.

190 ***Positive-sense single-stranded RNA viruses (+ssRNA): Picornaviridae, Astroviridae,***
191 ***Hepeviridae, Caliciviridae and, Coronaviridae***

192 The *Picornaviridae* were the most common viral group in our data set with 14 viruses: six
193 from Lizard Island and eight from Orpheus Island. We discovered a group of four novel
194 viruses that formed a distinct clade with the newly formed genus *Danipivirus*, represented by
195 a single virus that is commonly detected in model zebrafishes (31) (Figure 2). In this group it
196 was notable that we identified two relatively closely related viruses (70.6% sequence
197 similarity across the entire polyprotein) in damselfishes (Pomacentridae) from both islands. A
198 more distantly related virus was identified in *Chaetodon baronessa* (Chaetodontidae; 48-49%
199 RdRp similarity with both damselfish picornaviruses) suggesting that these viruses
200 diversified within reef fish, although on an unknown time scale. Evidence of reef
201 diversification was also identified in Pomacentrus nagasakiensis picornavirus and Blenniella
202 picornavirus from Orpheus Island that exhibited 66.1% similarity and were related to
203 fipiviruses found only at Lizard Island (Figure 3).

204 We identified 12 astroviruses, including five novel viruses from Lizard Island. These were
205 spread across three major clades: clade I, represented exclusively by fish viruses including
206 five goby viruses from Orpheus Island (see above); clade II, similarly represented by fish
207 viruses including those from both Lizard and Orpheus Island; and clade III that fell sister to

the genera *Mamastrovirus*, found in mammals and *Avastrovirus*, exclusively in birds (Figure 3). Notably, both Plagiotremus tapeinosoma astrovirus (Lizard Island) and Blenniella astrovirus (Orpheus Island) fell within clade III, exhibiting 75.8% sequence similarity in the RdRp gene. Similarly, both Eviota astrovirus and Chaetodon baronessa astrovirus clustered together within clade II, while the other reef fish viruses were more divergent.

Among other +ssRNA viruses, we identified two hepeviruses – Istigobius decoratus hepevirus (Orpheus Island) and Fowleria viaulae hepevirus (Lizard Island) – that grouped with other fish hepeviruses in phylogenetic trees, as well as a novel calicivirus in *Pomacentrus brachialis* that similarly fell within a broad group of fish and amphibian caliciviruses (Figure 3).

Both letoviruses (*Coronaviridae*, subfamily *Letovirinae*) —Chromis atripectoralis letovirus and Scarus psittacus letovirus—from Lizard Island formed a basal clade with Microhyla letovirus and three other viruses identified in African cichlids (32, 33). Together, this group likely formed a novel genus within the *Letovirinae* that includes diverse ectothermic hosts such as amphibians, jawless and ray-finned fishes (32).

Negative-sense single-stranded RNA viruses (-ssRNA): Hantaviridae* and *Articulavirales

A notable observation from our previous sampling of reef fish was the detection of hantaviruses in four different goby species at Orpheus Island. In contrast, at Lizard Island we only detected one hantavirus in a surgeonfish (*Acanthurus nigrofuscus*) that was sister to Wenling red spikefish hantavirus (NCBI/GenBank accession: AVM87662.1) and was highly divergent to those found at Orpheus Island (~29% RdRp similarity). Overall, these viruses fell within the subfamily *Actantavirinae*, that exclusively infects ray-finned fishes. Similarly, we detected two viruses with high levels of divergence (40% similarity) from both islands that were related to quaranjaviruses (order *Articulavirales*) (Figure 3).

232 ***DNA viruses***

233 We identified reef fish from three lineages within the *Parvoviridae*: the *Parvovirinae*,
234 *Chapparvovirus*, and *Ichthamaparvovirus* groups. Reef fish *Parvovirinae* fell as a sister
235 lineage to mammalian and avian viruses (e.g. *Dependoparvovirus*, *Aveparvovirus*), again
236 with high sequence divergence between islands. For example, the closest relatives among
237 both islands were Acanthurus nigrofuscus parvovirus (Lizard Island) and Dinematichthys
238 parvovirus (Orpheus Island) that shared 62% NS1 gene sequence similarity (Figure 3).
239 Notably, we identified a novel chapparvovirus in *Cryptocentrus strigilliceps* that was related
240 to tilapia parvovirus—a pathogen in farmed Tilapia in China—making it the second fish virus
241 identified in this group (34). Among the genus *Ichthamaparovirus*, both pleurosicya
242 icthamaparvovirus (Lizard Island) and Luposicya lupus ichthamaparvovirus (Orpheus Island)
243 shared a common ancestor (57.9% NSI similarity) and formed a distinct clade with other fish-
244 infecting parvoviruses (16, 33, 35).

245 Similar patterns were observed in reef fish *Betairidovirinae* (*Iridoviridae*), between both
246 Chrysiptera rollandi iridovirus (Lizard Island) and Enneapterygius tutuilae iridovirus
247 (Orpheus Island). These viruses exhibited 82.2% similarity in the conserved major capsid
248 protein (MCP) and clustered with erythrocytic necrosis virus (ENV).

249 Poxviruses were only identified at Lizard Island, with phylogenetic analysis placing them
250 with other fish viruses—Salmon gill pox virus and carp edema virus—within the subfamily
251 *Chordopoxvirinae*. Notably, Zebrasoma veliferum poxvirus and Scolopsis bilineata poxvirus
252 shared a common ancestor (80.3% similarity), strongly suggestive of an origin in reef fish
253 (Figure 3).

254

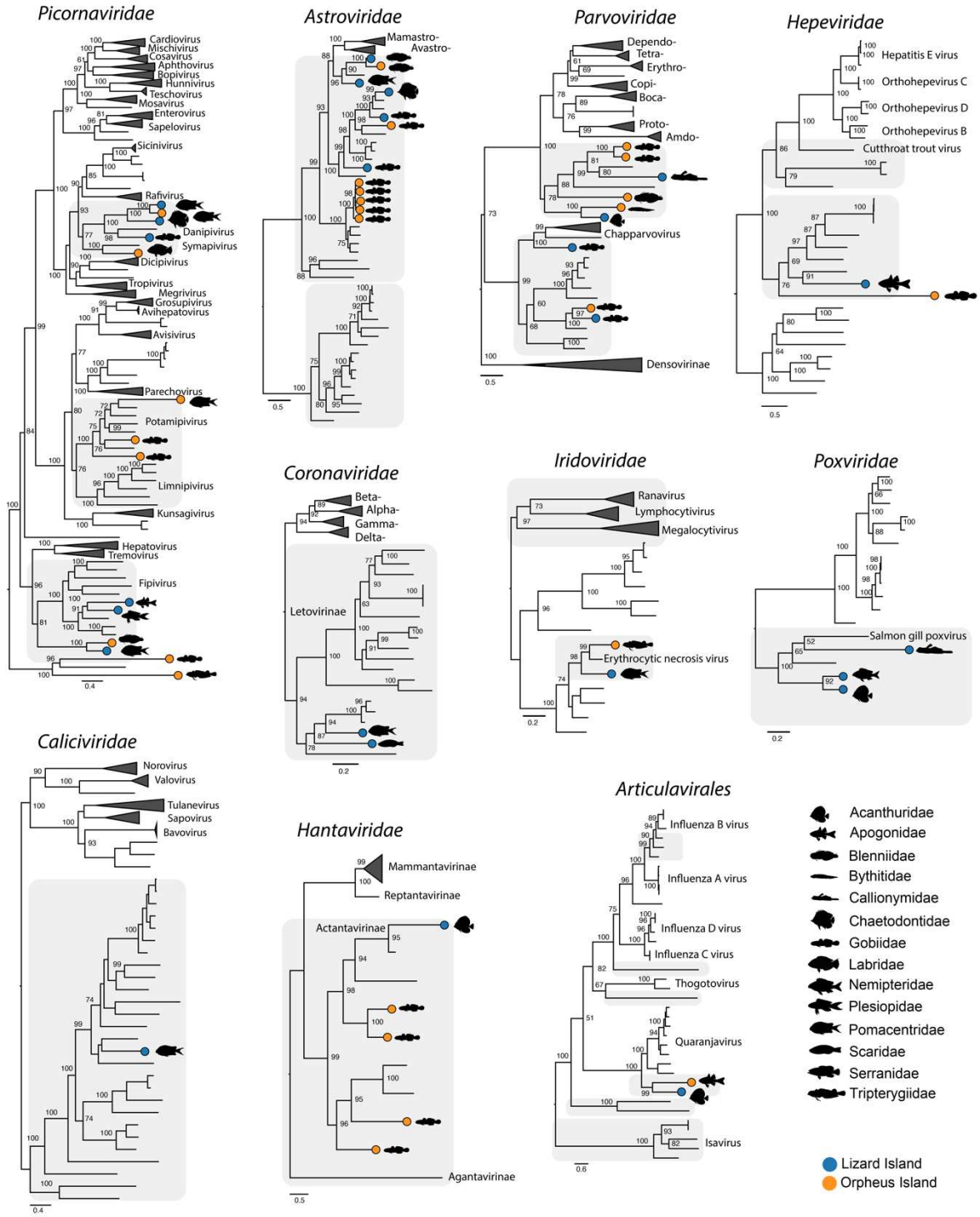
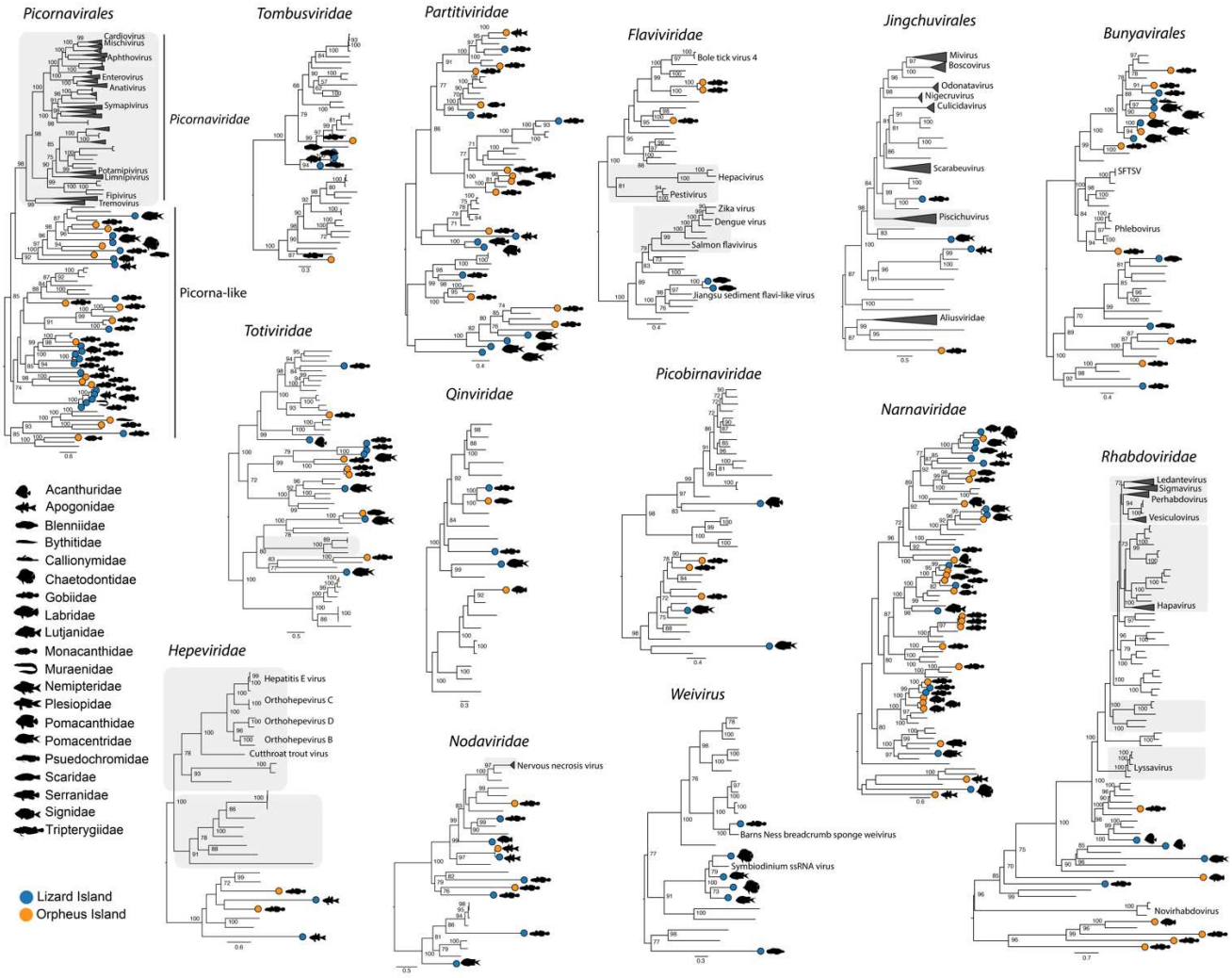


Figure 3. Phylogenetic analysis reveals relationships among reef fish viruses from both islands. Phylogenies were estimated using the RdRp gene for RNA viruses (*Picornaviridae*, *Astroviridae*, *Hepeviridae*, *Hantaviridae, Caliciviridae*, *Coronaviridae*, *Articulavirales*), NS1 gene for parvoviruses and DNA polymerase for iridoviruses and poxviruses. Coloured circles on branch tips represent viruses identified in this study. The scale bar represents the number

261 of amino acid substitutions per site. Shaded branches represent fish viruses. Trees were
262 midpoint rooted for clarity only.

263 **Close phylogenetic relationships of non-vertebrate-associated viruses**

264 A common observation in our data set was that genetically diverse reef fish exhibited very
265 similar assemblages of non-vertebrate-associated viruses (Figure 2). This pattern sits in
266 marked contrast to the vertebrate-associated viruses, that were rarely transmitted among
267 species. In particular, we identified the same virus (i.e., 98-99% similarity) in multiple host
268 species within the following groups: unclassified picornaviruses, *Tombusviridae*, *Totiviridae*,
269 *Nodaviridae*, *Narnaviridae*, *Bunyavirales* and (Figure 4). As expected, most of these cases of
270 virus sharing occurred within each island. For example, we identified the same tombusvirus
271 in two blennies and one wrasse from Lizard Island as well as the same narnavirus in three
272 gobies from Orpheus Island (Figure 4). A notable exception was the presence of two
273 nodaviruses identified in *Fowleria vaiulae* from both islands that exhibited 98% similarity in
274 the RdRp gene. While nodaviruses are capable of infecting fish species (e.g. nervous necrosis
275 virus) we detected reads associated with decapods in both libraries. These reads, in
276 combination with the phylogenetic positions of these viruses (i.e., highly divergent from
277 nervous necrosis virus and more related to crustacean viruses) strongly implies that these
278 viruses are of dietary origin.



279

280 **Figure 4.** Close phylogenetic relationships among non-vertebrate-associated viruses sampled
281 from reef fish. Phylogenies were estimated using amino acid sequences of the RdRp gene.
282 Coloured circles on branch tips represent viruses identified in this study. The scale bar
283 represents the number of amino acid substitutions per site. Shaded branches represent
284 vertebrate-associated viruses. Trees were midpoint rooted for clarity only.

285

286 **Diversity and abundance of bacteria in reef fish assemblages**

287 After the removal of cyanobacteria, the phylum *Proteobacteria* accounted for 73.5% of the
288 total sequence reads from bacteria, followed by *Firmicutes* (9%), *Actinobacteria* (8%),
289 *Bacteroidetes* (2.9%), *Spirochaetes* (2.3%), *Fusobacteria* (1.7%), with all other phyla each

representing <1% (Figure 5a). At the family level, the *Vibrionaceae* (42.7%) and *Enterobacteriaceae* (11.6%) were present at the highest frequencies followed by the *Endozoicomonadaceae* (5.9%), *Comamonadaceae* (5.9%), *Shewanellaceae* (5.8%), *Propionibacteriaceae* (5%), *Clostridiaceae* (4.9%), *Micrococcaceae* (2.3%), *Fusobacteriaceae* (2%), *Pseudomonadaceae* (1.3%), and *Mycoplasmataceae* (1.3%) (Figure 5).

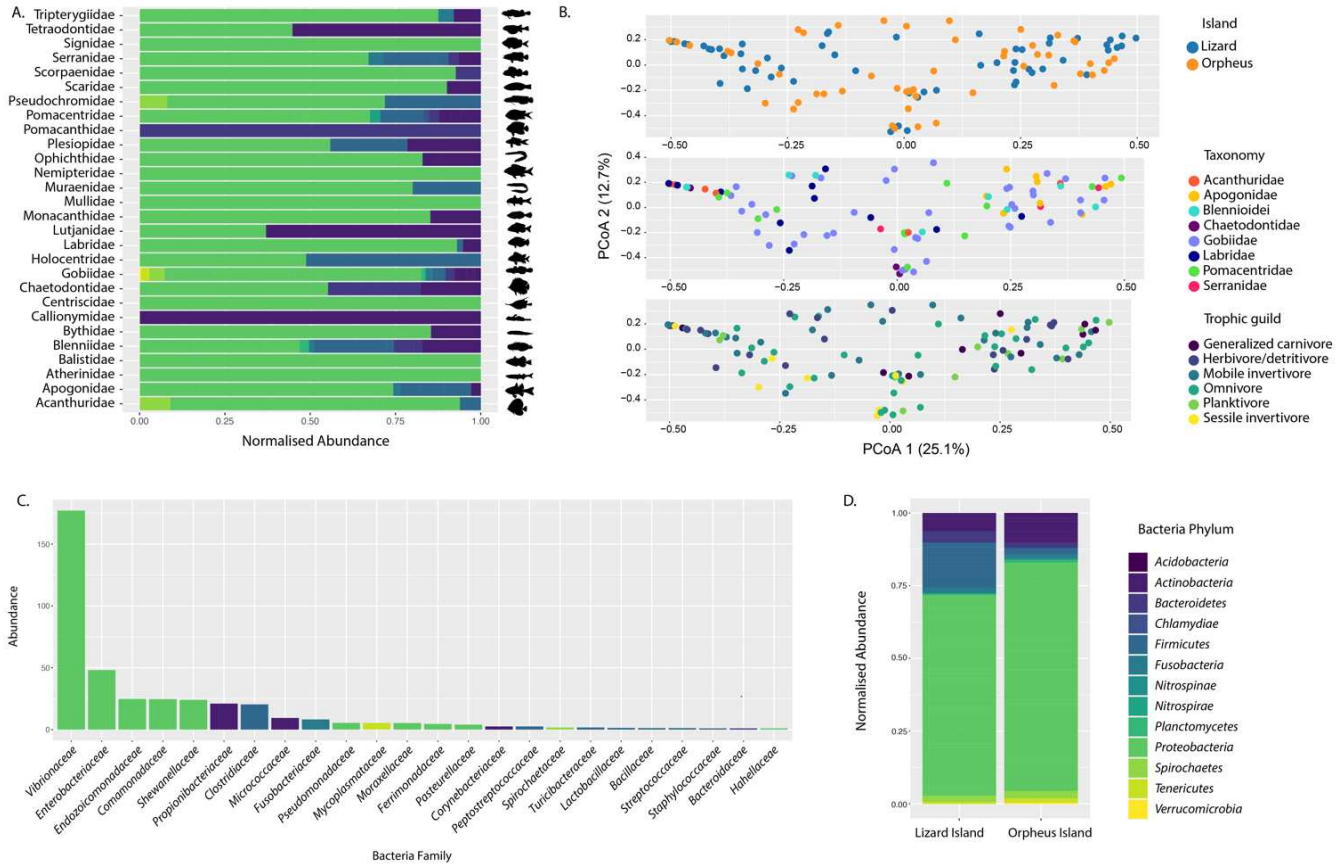
To explore patterns of bacterial diversity between both islands and among reef fish groups, we performed a principal coordinate analysis (PCoA). This revealed no partitioning according to reef location with largely overlapping bacterial communities between the two islands ($F = 1.711$, $R^2 = 0.015$, $p = 0.082$) (Figure 6). PERMANOVA revealed significant differences in bacterial composition between fish taxonomic groups ($F = 2.42$, $R^2 = 0.153$, $p = 0.001$). However, there was overlap among all fish species, particularly the gobies, which may be explained by their high involvement in coral reef food webs (5, 6) (Figure 5b).

Despite the high degree of overlap in bacterial families among the reef fishes, there was limited evidence for the sharing of individual bacterial species between fish species. Notably, however, we identified *Photobacterium damselae*—an important pathogen in aquaculture—in 88% of the cardinalfish species examined from both islands (Figure 6). We also detected other potentially pathogenic *Vibrionaceae*, including four that were related to those within the *Vibrio harveyi* clade: *V. parahaemolyticus*, *V. campbellii*, and *V. owensii*. (Supplementary Figure 1). In addition, we identified *V. fortis*—an opportunistic pathogen of coral—in the surgeonfish, *Ctenochaetus binotatus* (36) (Supplementary Figure 1).

To assess the impact of fish ecology on bacterial composition, we grouped fish species into six trophic guilds: carnivores, mobile invertivores, omnivores, planktivores, sessile invertivores, herbivores/detritivores (7). While we detected significant differences in bacterial

315 composition between these groups ($F = 1.485$, $R^2 = 0.074$, $p = 0.035$), there was similarly high
316 overlap with no clear partitioning according to host trophic guild (Figure 5b).

317



318

319 **Figure 5. (**A) Normalised abundance of bacterial phyla for each reef fish family. (B)
320 Principal coordinate analysis (PCoA) plots of bacterial communities for island, host
321 taxonomy and trophic guild using Bray-Curtis dissimilarity matrix. (C) Abundance of
322 bacterial families coloured by phylum. (D) Normalised abundance of bacterial phyla for each
323 island.

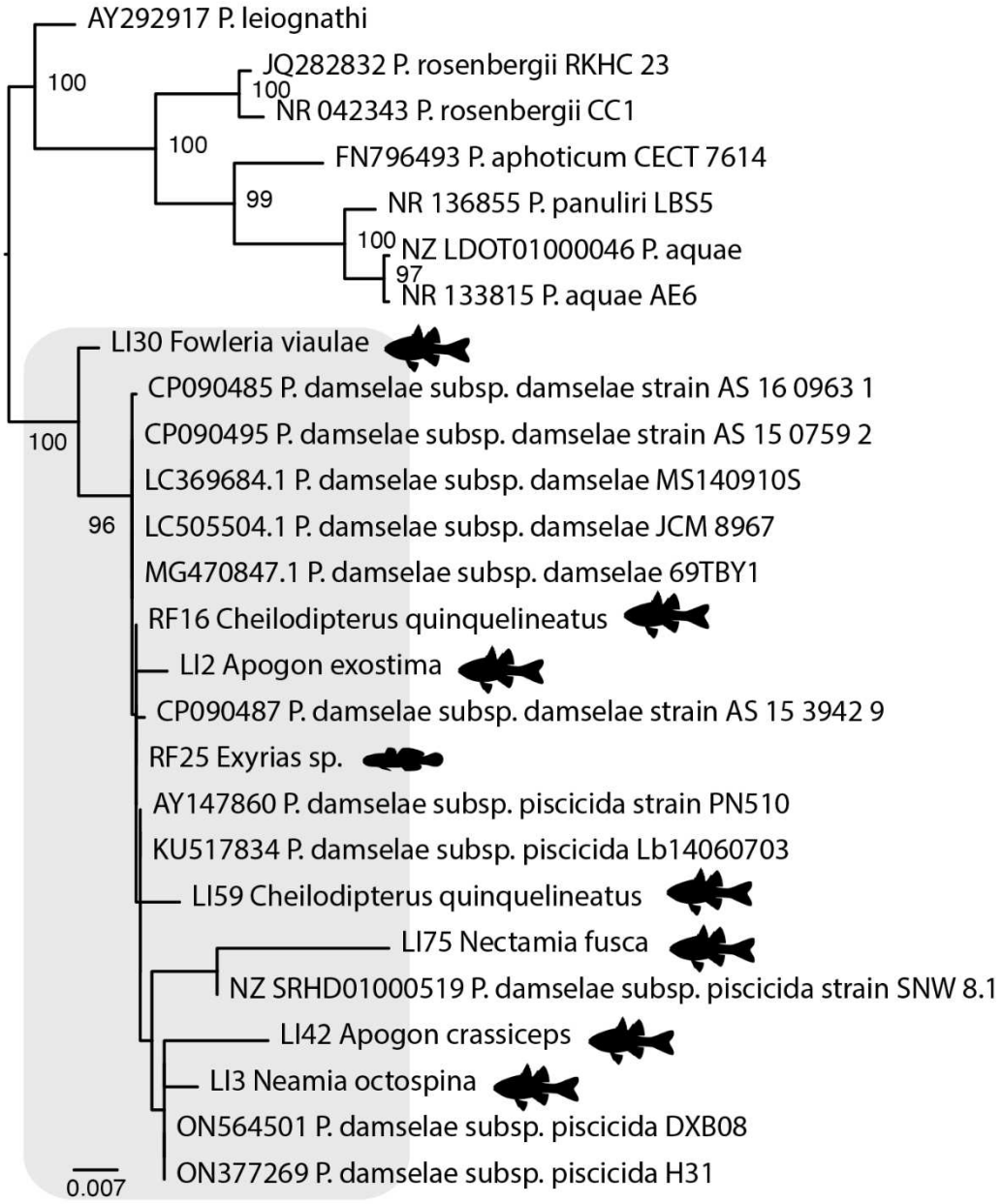


Figure 6. Maximum likelihood phylogeny of the genus *Photobacterium*, estimated using nucleotide sequences of the 16S gene. Fish silhouettes represent individuals identified in this study. The scale bar represents the number of nucleotide substitutions per site. Tree was midpoint rooted for clarity only.

331 **Composition of single-celled eukaryotes**

332 Finally, we identified transcripts representing single-celled eukaryotes from 15 phyla.
333 Dinoflagellates were the most abundant group (46% of the total eukaryotic reads) followed
334 by Bacillariophyta (21.6%), Foraminifera (12.6%), Cercozoa (5.8%), Euglenozoa (2.9%),
335 Apicomplexa (2.7%), Ciliophora (2.5%), Endomyxa (1.1%), Parabasalia, Haptista, Fornicata,
336 and Heterolobosea (all less than one percent). Fungi—Ascomycota, Basidiomycota,
337 Microsporidia—were identified at much lower frequencies, representing only 3.3% of the
338 total reads in this category.

339 Using these data, we performed a PCoA based on fish location, taxonomy and ecology. As
340 with the analysis of virome composition, this revealed overlapping microbial communities
341 between both islands ($F = 1.208$, $R^2 = 0.015$, $p = 0.181$). However, there were significant
342 differences in microbial communities between fish taxonomic groups
343 ($F = 1.384$, $R^2 = 0.119$, $p = 0.002$), which may be driven by the separation of pomacentrids
344 from other groups, in turn reflecting the larger diversity of microorganisms identified in this
345 group (Figure 7a-b). When assessing the impact of host ecology, we similarly identified
346 substantial overlap with no clear partitioning (Figure 7). Apart from *Symbiodinium* spp.,
347 which form symbiotic relationships with coral rather than fish and hence are dietary-
348 associated, we found no evidence for the same microbial species present in multiple fishes.
349 Among the Apicomplexa, which are common parasites of vertebrates, we detected reads
350 associated with *Goussia* spp. in *Chaetodon aureofasciatus*, *Halichores melanurus*, and *Eviota*
351 *melasma*. Overall, the parasitic families, such as the Eimeriidae (Apicomplexa),
352 Trypanosomatidae, and Ichthyobodonidae (both Euglenzoa) were found in eight libraries at
353 low transcript abundances, representing an average of 0.19% of the reads in each library.

354

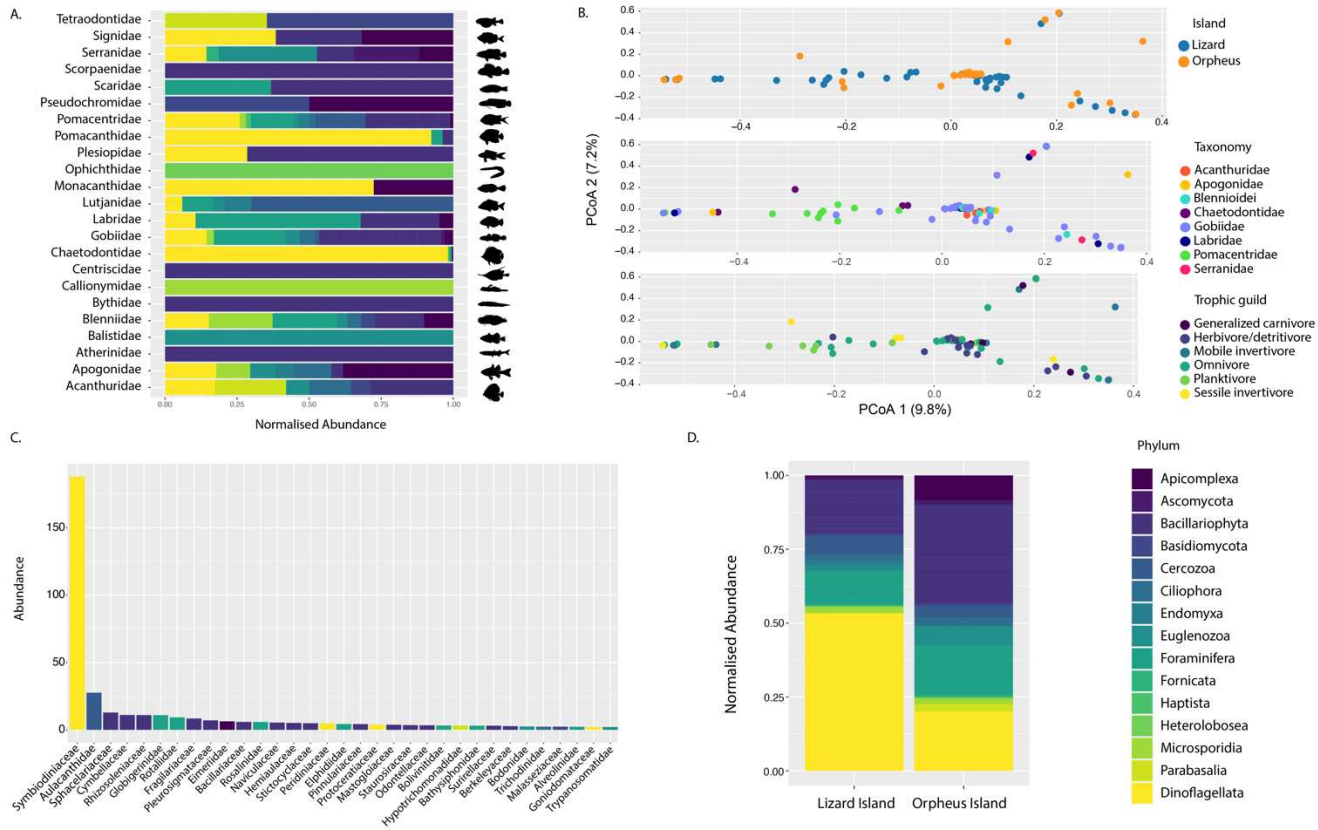


Figure 7. (A) Normalised abundance of single-celled eukaryotic phyla for each reef fish family. (B) Principal coordinate analysis (PCoA) plots of single-celled eukaryotic communities for island, host taxonomy and trophic guild using Bray-Curtis dissimilarity matrix. (C) Abundance of single-celled eukaryotic families coloured by phylum. (D) Normalised abundance of single-celled eukaryotic phyla for each island.

Discussion

Our metatranscriptomic analysis of 128 reef fish species revealed no significant differences in the composition of viral and microbial families/subfamilies between two islands located ~450 kilometres apart in the Australian Great Barrier Reef. Fish sampled from islands were associated with the presence of the *Picornaviridae*, *Astroviridae*, *Parvoviridae*, *Hantaviridae*, *Orthomyxoviridae*, *Coronaviridae*, *Hepeviridae* and *Iridoviridae*, as well as *Proteobacteria*

368 of the *Vibrionaceae, Enterobacteriaceae*, and *Endozoicomonadaceae*. Also of note was that
369 within each group of viruses we observed high levels of genetic diversity, with minimal
370 evidence for the same virus being shared among fish species (both within and between
371 islands), despite strong ecological interactions in the reef ecosystem and our spatially
372 restricted sampling.

373 These findings offer further support for recent studies of virus ecology in fish that show that
374 host phylogeny has a strong influence on virome composition (16, 33). For example, closely
375 related African cichlid species, that have evolved *in situ* within Lake Tanganyika over the last
376 10 million years, exhibit highly similar viromes with high levels of cross-species
377 transmission and viral generalism (33). African cichlids are members of a single family, the
378 Cichlidae, and exhibit some of the lowest pairwise genetic distances observed between
379 vertebrates (e.g., differences of 0.03% between some species). In contrast, reef fish
380 communities are composed of several divergent families, including 28 examined in this study
381 (7, 37), many of which were established around 66 Ma, occupying the biogeographic region –
382 Tethys – now covered by Europe and the Mediterranean Sea (27). During the Oligocene there
383 was a shift in species richness from Tethys to the IAA, eventually forming the IAA
384 biodiversity hotspot (27, 38). Most reef fish genera formed within the IAA around 30-15 Ma,
385 with accelerated speciation during the Miocene (27).

386 A notable difference between cichlids and reef fishes, which may explain their strikingly
387 contrasting levels of cross-species transmission, is that cichlids rapidly evolved within Lake
388 Tanganyika, while reef fishes diversified within the IAA, over many millions of years, prior
389 to their settlement on the GBR (27, 39). As the GBR that exists today formed after
390 Pleistocene sea level rises, reef fish communities only became established during the last
391 ~10,000 years, such that the genetic boundaries inhibiting cross-species transmission were
392 already established in the IAA before their settlement at reef locations. In marked contrast,

393 the adaptive radiation of the African cichlids would have provided a more favourable
394 environment for cross-species transmission as their diversification occurred rapidly within
395 the confinements of Lake Tanganyika, generating a large pool of closely related host species
396 for infection. Indeed, a time-calibrated phylogeny of cichlid hepaciviruses showed that
397 elevated rates of virus diversification coincided with a period of rapid cichlid speciation (33).

398 These ecological and evolutionary patterns were similarly observed across the bacteriome,
399 with *P. damselae* the only bacterium that was shared among multiple fish species. Indeed, *P.*
400 *damselae* was primarily identified in cardinalfishes (Figure 6), further illustrating a
401 phylogenetic effect. However, it is likely that this association represents a long-term
402 symbiotic relationship between cardinalfishes and *P. damselae*, particularly as it was detected
403 in *Cheilodipterus quinquelineatus* from both islands, as well as its presence in 88% of the
404 cardinalfish libraries examined. This strongly suggests that *P. damselae* forms part of the
405 natural microbiome in these species. It is noteworthy that most cardinalfish species are
406 nocturnal on coral reefs, and the genus *Photobacterium* is renowned for its
407 bioluminescence—via the expression of *lux* genes—including some strains of *P. damselae*
408 (40). Indeed, *Photobacterium mandapamensis* is a bioluminescent symbiont of the urchin
409 cardinalfish, *Siphamia tubifer*, where it provides light to attract prey (40, 41). Several
410 cardinalfish species have evolved specialized "light organs" that harbour bioluminescent
411 *Photobacterium*; however, these specialized organs are not present in all species (42). It is
412 important to note that we did not observe the expression of *lux* genes in any of our libraries,
413 and these species are not recognized for possessing a bioluminescent system. The presence of
414 *P. damselae* in almost all Apogonidae libraries implies a longstanding relationship between
415 *Photobacterium* and cardinalfishes on coral reefs, suggesting that this genus might have
416 originated and diversified in the reef environment. Overall, these data show that
417 cardinalfishes serve as natural hosts for *P. damselae* and should be monitored closely,

418 particularly if interacting with farmed populations that are often severely affected by *P.*
419 *damselae* infection (43).

420 The identification of two iridoviruses that were related to Erythrocytic necrosis virus (ENV;
421 83-89% MCP similarity) was also noteworthy. ENV is an important pathogen in the North
422 Atlantic and North Pacific oceans (44), and the presence of an ENV-like virus at Lizard
423 Island is consistent with a previous study that identified viral erythrocytic necrosis (VEN) in
424 a juvenile triggerfish (*Rhinecanthus aculeatus*) at this location (45). Moreover, this study
425 identified VEN-like bodies in *R. aculeatus* erythrocytes that were also found in the digestive
426 tract of associated gnathiid isopods that are common blood feeding parasites. Similarly, we
427 detected reads from gnathiids in *C. rollandi* from Lizard Island, implying that gnathids could
428 act as vectors in the marine environment (45), although we did not detect any gnathiid reads
429 in *Enneapterygius tutuilae* from Orpheus Island. This association warrants further
430 investigation and has the potential to improve control measures against pathogenic ENV in
431 susceptible species such as pink (*Oncorhynchus gorbuscha*) and chum (*Oncorhynchus keta*)
432 salmon (44).

433 It was notable that we identified a low number of protozoan parasites (n = 8), such as
434 apicomplexans and trypanosomes. In a similar manner, we detected low levels of fungal
435 reads (e.g., Ascomycota, Basidiomycota and Microsporidia), supporting the idea that the
436 marine environment, particularly coral reefs, contains low fungal biomass perhaps because of
437 the oligotrophic conditions compared to nutrient-rich terrestrial environments (46). The vast
438 majority of reads in this category came from dinoflagellates, diatoms, and foraminifera that
439 likely came from the reef environment, particularly dinoflagellates that form symbioses with
440 corals (1). Indeed, we discovered six viruses that fell within the "Weivirus" group—a group
441 of dinoflagellate, poriferan and mollusc viruses—grouping with a virus recently identified in
442 *Symbiodinium spp.* (47, 48) (Figure 4).

While our microbial profiling was able to detect diverse microbial communities, it is important to note that our sampling was initially performed for virological analysis (16), such that our methodology was optimized for the detection of viruses (e.g. ribosomal depletion). This may have limited our ability to detect bacteria and microbial eukaryotes. Moreover, there were necessary limitations in our sampling that impacted the power of our statistical analyses. For example, there was a considerably higher number of gobies sampled compared to other reef fish families (Figure 1). Moreover, these comparisons are based on combined tissues, such as liver and gills or whole fish (i.e. cryptobenthic reef fishes). Overall, given that metatranscriptomics is solely based on RNA-sequencing, we were only able to detect microbes and DNA viruses that were expressing genes during the time of sampling.

In summary, while our metatranscriptomic of reef fish communities from islands separated by ~450 kilometres revealed the same viral and bacterial families across both islands, there was strikingly little evidence for cross-species transmission within reef fish communities. As such, these data support the concept that fish are rich in microbial diversity, but that there are strong barriers to infection in host communities that display high levels of genetic diversity.

Materials and methods

Ethics

Fish were collected under a Great Barrier Reef Marine Park Authority permit (G16/37684.1) and James Cook University Animal Ethics permit A2752.

Fish sample collection

Building on transcriptome data from our previous sampling of reef fish from Orpheus Island (n = 192 fishes; 16 reef fish families) (16), an additional 163 individuals (24 families) were

collected at Lizard Island (14°40'08″S 145°27'34″E) during January 2022 (Supplementary Table 3). All fishes were intact with no visible signs of disease, and the vast majority were adults. These animals were captured using an enclosed clove oil method (16) from both Mermaid Cove and the Lagoon entrance, located on the north and south side of Lizard Island (Supplementary Table 1). All fish caught were placed either dissected (liver and gills) or whole in RNAlater and then transported to the lab on ice (Supplementary Table 1). Specimens were stored at −80°C until RNA extraction. Overall, we sampled 1-12 individuals per species (with a mean of 3 per species).

RNA extraction, metagenomic library preparation and next-generation sequencing

As described previously (16), tissue specimens (e.g. liver and gills or whole fish) were collectively processed as a single extraction for each individual fish sample. The combined tissues were submerged in 600 μl of lysis buffer containing 30 μl of foaming reagent (Reagent DX, Qiagen) and 60 μl of β-mercaptoethanol (Sigma-Aldrich). Tissue samples were homogenized with a TissueRuptor (Qiagen) for up to 1 minute at 5,000 rpm. The homogenate was centrifuged at maximum speed for three minutes to remove tissue residues. The RNA was then extracted from the resulting clear supernatant using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's guidelines.

RNA quantification was conducted utilizing a UV–Vis cuvette spectrophotometer (DeNovix, Delaware, USA) and a parallel capillary electrophoresis instrument (Fragment Analyzer; Agilent, CA, USA). RNA from individual fishes were pooled according to species, resulting in 79 RNA sequencing libraries newly generated from Lizard Island. All libraries were prepared using the TruSeq Total RNA Library Preparation Protocol (Illumina). Ribo-Zero Plus Kit (Illumina) was employed for host ribosomal RNA depletion, and paired-end sequencing (150 bp) was performed on the NovaSeq 6000 platform (Illumina). To mitigate

index hopping and minimize false virus–host assignments, each library was sequenced on two different lanes. Library construction and metatranscriptomic sequencing were performed by the Australian Genome Research Facility.

Assembly of reef fish viromes

We replicated the same methodology we used previously (16). Accordingly, raw RNA sequencing reads were quality trimmed using Trimmomatic v.0.38, employing the parameters SLIDINGWINDOW:4:5, LEADING:5, TRAILING:5, and MINLEN:25, and assembled into contigs using MEGAHIT v.1.2.9, with default parameter settings (49, 50). Assembled contigs were compared against the NCBI non-redundant protein (nr) and nucleotide (nt) databases (August 2022) using DIAMOND (BLASTX) (v.2.0.9) and BLASTn (51). To enable the identification of divergent viral sequences, we used an e-value search threshold of 1×10^{-5}. Contigs with top matches to the kingdom "Viruses" (NCBI taxid: 10239) were predicted as open reading frames (ORFs) using Geneious Prime (v.2022.0) (www.geneious.com) (52). To remove false positives, all putative viral ORFs were translated into amino acid sequences and used as a query to perform a second search (BLASTP) against the NCBI nr database using Geneious Prime. ORFs with top matches to fish genes were deemed as false positives and removed from further analysis. To determine whether our putative viral contigs were expressed endogenous viral elements (EVEs) we screened for disrupted ORFs and flanking host regions using CheckV and BLASTn (16, 33). Viral contig contamination and completion was determined using CheckV (53). Transcript abundances of both host (RPS13 gene) and virus (16, 17, 54) were calculated using RNA-Seq by Expectation Maximization (RSEM) (v.1.3.0) and coverage was assessed by mapping using Bowtie2 (v.2.3.3.1) (55, 56).

Taxonomic assignment and genome annotation of reef fish viruses

We aligned the amino acid sequences of our putative viruses (partial or complete) with the complete sequences of related viruses available on NCBI/GenBank using the E-INS-i algorithm in MAFFT v.7.450 (57). To determine whether our viruses were novel species, we used levels of sequence similarity and phylogenetic relationships (see below) as specified by the International Committee of Viral Taxonomy (ICTV) (https://talk.ictvonline.org) for each viral genus/family. We used these criteria to determine whether a virus was likely infecting reef fishes (i.e. vertebrate-associated) or of "non-vertebrate" origin, such as those derived from fish diet, microbiome or environment (16, 17, 54). Viral genomes were annotated with the Live Annotate and Predict tool in Geneious using reference sequences from NCBI/GenBank, with a similarity threshold of 20%. We also used the NCBI conserved domain (CDD) search tool and InterProScan with the TIGRFAMs (v.15.0), SFLD (v.4.0), PANTHER (v.15.0), SuperFamily (v.1.75), PROSITE (v.2022_01), CDD (v.3.18), Pfam (v.34.0), SMART (v.7.1), PRINTS (v.42.0), and CATH-Gene3D databases (v.4.3.0) (58).

Viral phylogenetic analysis

To infer the evolutionary relationships of both the vertebrate and non-vertebrate associated viruses, we aligned the translated contigs with background protein sequences from each viral family/subfamily/genus selected from the ICTV classification and obtained from NCBI/GenBank. For RNA viruses, we used the conserved RNA-dependent RNA polymerase (RdRp), while for DNA viruses we used the DNA polymerase. Amino acid sequence alignments were trimmed using TrimAl (v.1.2) with a gap threshold of 0.9 and a variable conserve value (59). The best-fit model of amino acid substitution was estimated with the "ModelFinder Plus" (-m MFP) flag in IQ-TREE (v.1.6.12) (60, 61). We used a maximum likelihood approach to estimate phylogenetic trees using IQ-TREE, with 1000 bootstrap replicates. Trees were annotated using FigTree (v.1.4.4) (http://tree.bio.ed.ac.uk/software%20/figtree/).

Virus nomenclature

Viruses were provisionally named (i.e., awaiting ICTV confirmation) according to host species (e.g. *Halichoeres melanurus ranavirus*) as described previously (16).

Microbial profiling

To screen for transcripts associated with bacteria or single-celled eukaryotes, we aligned our contigs to a custom database comprising all nucleotide sequences available on NCBI (with the removal of environment or artificial sequences) using the KMA aligner and CCMetagen (62, 63). We also used this output to assess metazoan reads—e.g., arthropod, mollusc, platyhelminth, nematode—that may represent potential vectors for virus transmission. For instances in which CCMetagen identified a microbe at the species level, we validated these taxonomic assignments by: (i) performing an additional search (BLASTn) against a custom 16S (bacteria) or 18S (eukaryote) rRNA database, and (ii) analysing the BLASTX output (see above) for hits to bacterial or eukaryotic proteins. The contigs from these BLAST hits were predicted into ORFs, translated into amino acid sequences, and used as a query to perform a second search against the NCBI using BLASTP for further validation. The 16S or 18S rRNA gene was then utilised for phylogenetic analysis as a final validation.

Analysis of Beta diversity

To compare viral and microbial communities between reef fish assemblages, we calculated beta diversity using a Bray–Curtis distance matrix with the phyloseq package in R (64). The variables assessed were host taxonomy, location (i.e. island) and trophic guild. Accordingly, fish species were categorised into six trophic guilds: carnivores, mobile invertivores, omnivores, planktivores, sessile invertivores, herbivores/detritivores as described in (7). We based our analysis on groups with three or more species. These data were then tested using

permutational multivariate analysis of variance (PERMANOVA) with the vegan package (adonis) (65). All plots were constructed using ggplot2 in R (66).

Data Availability

Raw sequence reads have been deposited in the Sequence Read Archive (NCBI/SRA) under BioProject PRJNA1078998. All viral sequences discovered have been deposited in NCBI/GenBank under the accessions XXXX-XXXX. All phylogenetic trees, tables, and chrona plots are available on GitHub under the repository https://github.com/vcosta16/reeffishvirome.

Acknowledgements

This work was funded by an Australian Research Council Discovery Project (DP200102351) and a National Health and Medical Research Council Investigator grant (GNT2017197) to ECH, and an ARC Laureate Fellowship (FL 190100062) to DRB. JLG is funded by a New Zealand Royal Society Rutherford Discovery Fellowship (RDF-20-UOO-007) and a Marsden Fund Fast Start (20-UOO-105). We acknowledge the University of Sydney for providing the high-performance computer cluster “Artemis”, that was used for this study.

Author Contributions

VAC and ECH conceptualised the study. VAC, DRB and JCOM performed the sampling. VAC performed the analyses. VAC wrote and prepared the original draft. VAC, JCOM, ECH, EH, DRB, and JLG edited and revised the manuscript. ECH and JLG funded the project. EH and ECH supervised the project.

Supplemental Information

583 **Supplementary Figure 1.** Maximum likelihood phylogeny of the *Vibrio harveyi* clade
584 estimated using nucleotide sequences of the 16S gene. Fish silhouettes represent individuals
585 identified in this study. The scale bar represents the number of nucleotide substitutions per
586 site. Tree was midpoint rooted for clarity only.

587 **Supplementary Table 1.** Composition of host sequence reads.

588 **Supplementary Table 2.** Description of vertebrate-associated viruses.

589 **Supplementary Table 3.** Host library information.

590

591 ## References

592 1. Roth, MS. 2014. The engine of the reef: photobiology of the coral–algal
593 symbiosis. *Front Microbiol* 5:422.

594 2. Spalding, M, Grenfell, A. 1997. New estimates of global and regional coral reef
595 areas. *Coral Reefs* 16:225–230.

596 3. Eschmeyer WN, Fricke R, Fong JD, Polack DA. 2010. Marine fish diversity: history of
597 knowledge and discovery (Pisces). *Zootaxa* 2525:19-50.

598 4. Hughes TP, Bellwood DR, Connolly SR. 2002. Biodiversity hotspots, centres of
599 endemicity, and the conservation of coral reefs. *Ecol Lett* 5:775-84.

600 5. Bellwood DR, Wainwright PC. 2002. The history and biogeography of fishes on coral
601 reefs. In *Coral reef fishes: dynamics and diversity in a complex ecosystem*. 5:32.

602 6. Brandl SJ, Goatley CH, Bellwood DR, Tornabene L. 2018. The hidden half: ecology and
603 evolution of cryptobenthic fishes on coral reefs. *Biol Rev* 93:1846-73.

604 7. Siqueira AC, Morais RA, Bellwood DR, Cowman PF. 2020. Trophic innovations fuel
605 reef fish diversification. *Nat Commun* 11:2669.

606 8. Mihalitsis M, Morais RA, Bellwood DR. 2022. Small predators dominate fish predation
607 in coral reef communities. *PLoS Biol* 20:e3001898.

608 9. Bellwood DR, Hoey AS, Ackerman JL, Depczynski M. 2006. Coral bleaching, reef fish
609 community phase shifts and the resilience of coral reefs. *Glob Change Biol* 12:1587-94.

610 10. Grutter AS. 1999. Cleaner fish really do clean. *Nature* 398:672-3.

611 11. Robertson DR. 1982. Fish feces as fish food on a Pacific coral reef. *Mar Ecol Prog*
612 7:253-65.

613 12. Clements KD, Angert ER, Montgomery WL, Choat JH. 2014. Intestinal microbiota in
614 fishes: what's known and what's not. *Mol Ecol* 23:1891-8.

615 13. Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K,
616 Rawls JF. 2011. Evidence for a core gut microbiota in the zebrafish. *ISME J* 5:1595-608.

617 14. Ghanbari M, Kneifel W, Domig KJ. 2015. A new view of the fish gut microbiome:
618 advances from next-generation sequencing. *Aquaculture* 448:464-75.

619 15. Pratte ZA, Besson M, Hollman RD, Stewart FJ. 2018. The gills of reef fish support a
620 distinct microbiome influenced by host-specific factors. *Appl Environ Microbiol*
621 84:e00063-18.

622 16. Costa VA, Bellwood DR, Mifsud JC, Van Brussel K, Geoghegan JL, Holmes EC,
623 Harvey E. 2023. Limited cross-species virus transmission in a spatially restricted coral
624 reef fish community *Virus Evol* 9:vead011.

17. Geoghegan JL, Di Giallonardo F, Wille M, Ortiz-Baez AS, Costa VA, Ghaly T, Mifsud JC, Turnbull OM, Bellwood DR, Williamson JE, Holmes EC. 2021. Virome composition in marine fish revealed by meta-transcriptomics. *Virus Evol* 7:veab005.

18. Wang AR, Ran C, Ringø E, Zhou ZG. 2018. Progress in fish gastrointestinal microbiota research. *Rev Aquacult* 10:626-40.

19. Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight RO, Kilham SS, Russell JA. 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol Ecol* 21:3363-78.

20. Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Knight R, Caporaso JG, Svanbäck R. 2014. Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). *Ecol Lett* 17:979-87.

21. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol* 72:457-70.

22. French RK, Anderson SH, Cain KE, Greene TC, Minor M, Miskelly CM, Montoya JM, Wille M, Muller CG, Taylor MW, Digby A, Kakapo Recovery Team, Holmes EC. 2023. Host phylogeny shapes viral transmission networks in an island ecosystem. *Nat Ecol Evol* 7:1834-43.

23. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. 2011. Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog* 7:e1002260.

24. Gupta P, Vishnudas CK, Robin VV, Dharmarajan G. 2020. Host phylogeny matters: Examining sources of variation in infection risk by blood parasites across a tropical montane bird community in India. *Parasit Vectors* 13:1-3.

25. Shaw LP, Wang AD, Dylus D, Meier M, Pogacnik G, Dessimoz C, Balloux F. 2020. The phylogenetic range of bacterial and viral pathogens of vertebrates. *Mol Ecol* 29:3361-79.

26. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. 2014. The evolution and genetics of virus host shifts. *PLoS Pathog* 10:e1004395.

27. Bellwood DR, Goatley CH, Bellwood O. 2017. The evolution of fishes and corals on reefs: form, function and interdependence. *Biol Rev* 92:878-901.

28. Webster JM, Braga JC, Humblet M, Potts DC, Iryu Y, Yokoyama Y, Fujita K, Bourillot R, Esat TM, Fallon S, Thompson WG. 2018. Response of the Great Barrier Reef to sea-level and environmental changes over the past 30,000 years. *Nat Geosci* 11:426-32.

29. Lefèvre CD, Nash KL, González-Cabello A, Bellwood DR. 2016. Consequences of extreme life history traits on population persistence: do short-lived gobies face demographic bottlenecks? *Coral Reefs* 35:399-409.

30. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, Takahashi S, Tatem AJ, Wagner CE, Wang LF, Wesolowski A. 2022. Infectious disease in an era of global change. *Nat Rev Microbiol* 20:193-205.

31. Altan E, Kubiski SV, Boros Á, Reuter G, Sadeghi M, Deng X, Creighton EK, Crim MJ, Delwart E. 2019. A highly divergent picornavirus infecting the gut epithelia of zebrafish (*Danio rerio*) in research institutions worldwide. *Zebrafish* 16:291-9.

32. Miller AK, Mifsud JC, Costa VA, Grimwood RM, Kitson J, Baker C, Brosnahan CL, Pande A, Holmes EC, Gemmell NJ, Geoghegan JL. 2021. Slippery when wet: cross-species transmission of divergent coronaviruses in bony and jawless fish and the evolutionary history of the *Coronaviridae*. *Virus Evol* 7:veab050.

670 33. Costa VA, Ronco F, Mifsud JC, Harvey E, Salzburger W, Holmes EC. 2024. Host
671 adaptive radiation is associated with rapid virus diversification and cross-species
672 transmission in African cichlid fishes. *Curr Biol*, in press.

673 34. Liu W, Zhang Y, Ma J, Jiang N, Fan Y, Zhou Y, Cain K, Yi M, Jia K, Wen H, Liu W.
674 2020. Determination of a novel parvovirus pathogen associated with massive mortality in
675 adult tilapia. *PLoS Pathog* 16:e1008765.

676 35. Pénzes JJ, de Souza WM, Agbandje-McKenna M, Gifford RJ. 2019. An ancient lineage
677 of highly divergent parvoviruses infects both vertebrate and invertebrate hosts. *Viruses*
678 11:525.

679 36. Sun X, Li Y, Yang Q, Zhang H, Xu N, Tang Z, Wu S, Jiang Y, Mohamed HF, Ou D,
680 Zheng X. 2023. Identification of quorum sensing-regulated *Vibrio fortis* as potential
681 pathogenic bacteria for coral bleaching and the effects on the microbial shift. *Front*
682 *Microbiol* 14:1116737.

683 37. Svardal H, Salzburger W, Malinsky M. 2021. Genetic variation and hybridization in
684 evolutionary radiations of cichlid fishes. *Annu Rev Anim Biosci* 9:55-79.

685 38. Cowman PF, Bellwood DR 2013. The historical biogeography of coral reef fishes: global
686 patterns of origination and dispersal. *J Biogeogr* 40:209-24.

687 39. Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, Indermaur A,
688 Malinsky M, Ricci V, Kahmen A, Jentoft S., Salzburger, W. 2021. Drivers and dynamics
689 of a massive adaptive radiation in cichlid fishes. *Nature* 589:76-81.

690 40. Urbanczyk H, Ast JC, Dunlap PV. 2011. Phylogeny, genomics, and symbiosis of
691 *Photobacterium*. *FEMS Microbiol Rev* 35:324-42.

41. Gould AL, Henderson JB, Lam AW. 2022. Chromosome-Level Genome assembly of the bioluminescent cardinalfish *Siphamia tubifer*: an emerging model for symbiosis research. *Genome Biol Evol* 14:evac044.

42. Thacker CE, Roje DM. 2009. Phylogeny of cardinalfishes (Teleostei: Gobiiformes: Apogonidae) and the evolution of visceral bioluminescence. *Mol Phylogenet Evol* 52:735-45.

43. Gouife M, Chen S, Huang K, Nawaz M, Jin S, Ma R, Wang Y, Xue L, Xie J. 2022. *Photobacterium damselae subsp. damselae* in mariculture. *Aquac Int* 30:1453-80.

44. Pagowski VA, Mordecai GJ, Miller KM, Schulze AD, Kaukinen KH, Ming TJ, Li S, Teffer AK, Tabata A, Suttle CA. 2019. Distribution and phylogeny of erythrocytic necrosis virus (ENV) in salmon suggests marine origin. *Viruses* 11:358.

45. Davies AJ, Curtis L, Grutter AS, Smit NJ. 2009. Suspected viral erythrocytic necrosis (VEN) in a juvenile blackbar triggerfish, *Rhinecanthus aculeatus*, from Lizard Island, Great Barrier Reef, Australia. *Mar Biodivers Rec* 2:e149.

46. Roik A, Reverter M, Pogoreutz C. 2022. A roadmap to understanding diversity and function of coral reef-associated fungi. *FEMS Microbiol Rev* 46:fuac028.

47. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J. 2016. Redefining the invertebrate RNA virosphere. *Nature* 540:539-43.

48. Benites LF, Stephens TG, Bhattacharya D. 2022. Multiple waves of viral invasions in *Symbiodiniaceae* algal genomes. *Virus Evol* 8:veac101.

49. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-20.

714 50. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-
715 node solution for large and complex metagenomics assembly via succinct de Bruijn
716 graph. *Bioinformatics* 31:1674–1676.

717 51. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using
718 DIAMOND. *Nat Methods* 12:59-60.

719 52. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper
720 A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012.
721 Geneious Basic: an integrated and extendable desktop software platform for the
722 organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

723 53. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021.
724 CheckV assesses the quality and completeness of metagenome-assembled viral
725 genomes. *Nat Biotechnol* 39:578–585.

726 54. Costa VA, Mifsud JC, Gilligan D, Williamson JE, Holmes EC, Geoghegan JL. 2021.
727 Metagenomic sequencing reveals a lack of virus exchange between native and invasive
728 freshwater fish across the Murray–Darling Basin, Australia. *Virus Evol* 7:veab034.

729 55. Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data
730 with or without a reference genome. *BMC Bioinformatics* 12:323.

731 56. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*
732 *Methods* 9:357–359.

733 57. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
734 improvements in performance and usability. *Mol Biol Evol* 30:772-80.

735 58. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,
736 Mitchell A, Nuka G, Pesseat S. 2014. InterProScan 5: genome-scale protein function
737 classification. *Bioinformatics* 30:1236-40.

738 59. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
739 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

740 60. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and
741 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol*
742 *Evol* 32:268–274.

743 61. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017.
744 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat*
745 *Methods* 14:587–589.

746 62. Clausen PT, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads
747 against redundant databases with KMA. *BMC Bioinformatics* 19:1–8.

748 63. Marcelino VR, Clausen PTLC, Buchmann JP, Wille M, Iredell JR, Meyer W, Lund O,
749 Sorrell TC, Holmes EC. 2020. CCMetagen: comprehensive and accurate identification of
750 eukaryotes and prokaryotes in metagenomic data. *Genome Biol* 21:103.

751 64. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive
752 analysis and graphics of microbiome census data. *PloS one* 8:e61217.

753 65. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci*
754 14:927-30.

755 66. Valero-Mora PM. 2010. ggplot2: elegant graphics for data analysis. *J Stat Softw* 35:1-3.