Conserved organ-specific microbial assemblages in different populations of a terrestrial crab

Authors: Giovanni Bacci[1*], Sara Fratini[1*], Niccolò Meriggi[1*], Christine L. Y. Cheng[2], Ka Hei Ng[2], Massimo Pindo[3], Alessio Iannucci[1], Alessio Mengoni[1], Duccio Cavalieri[1]†, Stefano Cannicci[1,2]†

Affiliations:
[1] Department of Biology, University of Florence, Sesto Fiorentino, IT50019, Italy

[2] The Swire Institute of Marine Science and the Division of Ecology and Biodiversity, The University of Hong Kong, Hong Kong, Hong Kong SAR

[3] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, IT38098, Italy

* These authors contributed equally to this work

† Corresponding author. Email: cannicci@hku.hk (S.C.); duccio.cavalieri@unifi.it (D.C.)

Abstract

Brachyuran crabs originated in the oceans and evolved specific morphological and physiological adaptations to live in freshwater, intertidal and even terrestrial habitats but the role of a selection mechanism involving symbiotic microorganisms long these colonization processes are not known. In this work we investigated the associated microbiota of three populations of a terrestrial brachyuran crab, *Chiromantes haematocheir*, to find evidence of a conserved crab-specific microbiome unrelated to the population of origin and dissimilar from environmental microbial assemblages. Bacterial 16S rRNA gene and fungal ITS sequences were obtained from selected crab organs and environmental matrices to profile microbial communities. In spite of the presence of truly marine larval stages and no gregarious behaviour, favouring microbiota exchanges, we found common, organ-specific microbiota, associated to the gut and the gills of the crabs (with more than 15% of the genera detected specifically enriched only in one organ). Our results suggest an early establishment of a new common, stable microbiota in the transition from water to land.

Introduction

The evolution of a terrestrial lifestyle independently occurred multiple times across and within most of the metazoan phyla (*1–4*). Since the marine and terrestrial realms are ecologically very different, these several transitions involved profound morphological, physiological and behavioural adaptations, often converging towards similar structures and metabolic pathways (*1*, *4*). Among land adapted phyla, Arthropoda contribute the most to species diversity, which is the evolutionary outcome of a series of independent land invasions started in the Ordovician and still in progress at present (*4*, *5*). Indeed, insects and arachnids represent the most successful arthropods on land, in terms of biomass, functional roles and diversity, but crustaceans are also well represented. This largely marine group shows many examples of semi-terrestrial and terrestrial forms, mostly represented by Amphipoda, Isopoda and Decapoda, the latter including the largest extant terrestrial arthropod of the world, the iconic coconut crab *Birgus latro* (*6*).

In chronological order, the true crabs (Decapoda; Brachyura) were among the last groups of arthropods to perform the sea-land transition (*5*), but they have been particularly successful with more than twenty phylogenetically unrelated families that include species with intertidal, semiterrestrial and truly terrestrial lifestyles (*5*, *7*). Adaptations to terrestrial environments involved several morphological and physiological changes in sensory, locomotory, respiratory, excretory, osmoregulatory, digestive and reproductive systems (*8–18*). Interestingly, brachyuran gills have not proved to be fully functional for respiration in terrestrial environment (*1*) and are involved in osmotic regulation, urine dilution, nitrogen excretion and ion-regulation in most terrestrially adapted families (*11*, *19*, *20*). Other essential adaptations include herbivory and immune system recognition of potentially fungal pathogens, which are rare in marine environment but abundant on land, where they are

commonly known pathogens of terrestrial arthropods, including semiterrestrial mangrove crabs (*21*, *22*).

Due to the relatively high diversity of terrestrial forms, when related to their recent history of attempts to conquer the land, brachyurans have been considered an excellent animal model for the study of the evolution of terrestrial adaptations (*2*, *5*, *8*, *23*). They are, however, rarely surveyed in studies investigating the involvement of commensal and symbiotic microbiota in such a dramatic transition. Indeed, metazoans groups have adapted to different habitats also by creating mutualistic and symbiotic relationships with the microorganisms that inhabit their tissues and organs (*24*). These relationships can be so tight to drive the evolution and Darwinian selection of both host and microbiome, as a single unit (*25–27*).

Within this theoretical frame, we investigated the role of associated microbial communities in the terrestrial adaptations of a semi-terrestrial brachyuran crab, *Chiromantes haematocheir*, a common inhabitant of lowland forests of East Asia (*28*, *29*). We applied targeted metagenomics approaches to profile both prokaryotic and fungal communities isolated from 1) organs (i.e., gills, gonads and gut) collected from specimens belonging to different populations and 2) relevant environmental matrixes (i.e., soil, leaf litter, water and water debris) sampled in their microhabitats. Our null hypothesis was that microbial composition of crab organs could show the same microbial signature displayed by the matrices from the surrounding environments. This would indicate that the crab organs would host bacterial and fungal populations acquired from the environments without any specific selection and could be explained by a dynamic "loss and acquire" balance that ultimately would prevent the development of stable host-microbe associations. Our alternative hypothesis was that crab organs may show exclusive bacterial and/or fungal communities, in terms of diversity and composition, when compared to those present in the environment. Thus, our goal was to investigate, for the first time for terrestrial brachyurans, the presence of a selection process in

76 favour of specific microorganisms and, ultimately, the evolution of adaptive symbioses able to
77 contribute to the terrestrial life style of a bimodal crab only recently adapted to the land
78 environment.

79 # Results

80 ## Microbial composition of crab's organs and environmental samples

81 Sequencing produced 25,875,448 sequences for prokaryotic 16S rRNA amplicon and
82 15,788,842 sequences for the fungal internal transcribed region 1 (ITS1) that were clustered
83 according to DADA2 pipeline (see Supplementary methods for additional details). After the
84 clustering pipeline two fungal samples (namely GI_TKP_5 and MG_SH_1) produced no counts
85 in any of the fungal amplicon sequence variants (ASVs) and were therefore removed from
86 subsequent analyses. Since several fungal samples reported a low number of reads in crab's
87 organs, two samples were replicated and tested for correlation and accuracy to evaluate the
88 presence of fungal cells in the crabs' organs (Figure S1 and Table S1). Technical replicates
89 showed an average correlation coefficient (Spearman's rho) between replicates ranging
90 between 0.74 and 0.99, with an accuracy higher than 75% in all contrasts. These results mean
91 that the low recovery in biotic samples was not due to technical or analytical issues but
92 corresponded to the scarce presence of fungal DNA (Figure S2, Table S2 and Supplementary
93 method section).

94 Single nucleotide clustering detected a total of 56,233 ASVs including both 16S rRNA gene and
95 ITS1 amplicons. The 16S rRNA gene amplicons produced 55,036 ASVs (97.87%), whereas the
96 ITS1 region was clustered into 1,197 ASVs (2.13%). Most variants were assigned to the
97 Bacteria domain (54,800 ASVs, 97.45% of the total ASVs detected) and Fungi kingdom, even if
98 a sporadic presence of Archaea was found (236 ASVs, 0.42% of the total ASVS detected). In
99 general, microbial diversity was higher in the environmental samples than in crab's organs,

100 except for the gills where we found a high diversity of 16S rRNA amplicons (Figure 2a and b,
101 Table S3). The fungal diversity was low in all crab's organs, whereas environmental samples
102 were characterised by a highly diverse fungal assemblage, with soil, water, and water debris
103 reporting the highest diversities (Figure 2c & d). For each sample category and at each site,
104 the difference between extrapolated diversity—namely the inverse Simpson index calculated
105 by simulating a higher sequencing depth than the observed depth for each sample—and
106 observed diversity was low (Table 1), meaning that our sampling effort was effective in
107 assessing the natural microbial diversity for each category. Even the Good's coverage
108 estimator was higher than 99.9% in all sample types highlighting that only 1‰ of clustered
109 sequences came from ASVs detected only once in the whole dataset (Table 1 and Table S3).
110 Microbial diversity was rather uniform across sampling sites, except for To Kwa Peng (TKP),
111 which showed a significantly lower diversity with respect to the other sites in terms of ITS1
112 data (Figure S3).

113 **Microbial distribution across sample types**

114 The multidimensional ordination showed that both environmental samples and crab's organs
115 contributed to shape microbial community distribution, but their effect varied according to
116 both the amplicon type and the category of sample considered (Figure 3 and S4). In terms of
117 composition, the microbial communities found in the environmental samples overlapped with
118 each other more than the ones characteristic of the crab's organs (Figure 3a). The PCoA built
119 on 16S rRNA gene amplicons (Bacteria and Archaea) showed that soil and litter, and water
120 and water debris, respectively, formed two distinct clusters. The observed intra-cluster
121 variability is due to differences in microbial communities found across the three Hong Kong
122 sites (Figure S4). The same 16S rRNA amplicon dataset showed a clear separation across
123 crab's organs, with a very limited influence of the collection sites (Figure 3 and S4). The gut's
124 microbiome composition was similar across the two sampled sections (mid- and hindgut),

125 while the gills and gonads sharply separated from each other and the gut itself. These
126 significant differences were highlighted also by the permutational analysis of variance (Figure
127 3b, Table S4). With respect to 16S rRNA data, fungal distribution produced more overlap,
128 especially across crab's organs, with the gonads being the only organ to show significant
129 differences when compared to the other organs (Figure 3b, Table S4).

130 Both environmental samples and crab's organs possessed a large set of microbial species
131 which were unique to the category considered (Figure S5). As already mentioned, sampling
132 sites also contributed to shape microbial communities, but they predominantly affected
133 environmental samples distribution and showed a low effect on crab's organs (Figure S4a and
134 b, Figure S5a and c). In addition, the community composition of both prokaryotic and fungal
135 communities was more variable and dispersed at site level than at organ level (Figures S4 and
136 S6, Table S5), showing a higher specificity of such assemblages within the different organs
137 and across sites. Gills exchanged microbes with environmental matrices in direct contact with
138 them (soil and water) and even with the gut, which is in turn connected with the same
139 environmental matrices through defecation (Figure S5b and d).

140 **Defining characteristic patterns across sample types**

141 To inspect microbial distribution in different sample categories, we performed log-likelihood
142 ratio test on both prokaryotic and fungal communities using DESeq2. We found 250 ASVs
143 (218 bacteria and 32 fungi) reporting a different distribution across crab's organs and/or
144 environmental samples (Figure 4 and Table S6). Even if significant ASVs corresponded to
145 0.45% of the ASVs profiled in the whole community (250 on 55819), they accounted for more
146 than 50% of the total microbial abundance (with a mean in each sample of 56.5% and a
147 standard error of 2.10%) reflecting the presence of many rare and sporadic species
148 throughout sample types and sites. Divisive analysis of hierarchical clustering obtained using
149 variance-stabilized counts produced four distinct clusters, which show a peculiar pattern of

150 abundance of ASVs clearly linked to the ecological and biological settings of the study. Indeed,
151 clusters 1 and 2 (Figures 4 and S7) were composed by microbial assemblages highly
152 represented in the crab's organs, while the other two clusters, (namely, cluster 3 and 4)
153 represented the microorganisms mostly found in the environmental samples (Figures 4 and
154 S7). The clusters representing the microorganisms more abundant in the crab's organs were
155 represented by bacterial ASVs only and were split into two groups, formed by the bacteria of
156 the gut (cluster 1) and of the gills (cluster 2), respectively, with the latter being the smaller
157 one.

158 Taxonomic and functional enrichments in crab's organs and environmental 159 samples

160 An exploration of the distribution of the scaled variance-stabilized counts within each sample
161 group is shown in Figure 5a. Pairwise Wilkoxon test revealed that the ASVs present in the
162 cluster 1 were significant enriched in the gut, while ASVs included in cluster 2 were
163 significantly enriched in the gills (Figure 5a). Clusters 3 and 4 were significantly related to
164 environmental matrices, cluster 3 to the water and water debris samples, while cluster 4 with
165 litter and soil samples (Figure 5a).

166 Enrichment analysis estimated using $\log_2$ fold-changes allowed to assign a fair number of
167 ASVs to a specific taxon present in the previously described clusters (Figure 5).
168 Proteobacteria was the most represented phylum of the whole 16S dataset and comprised
169 ASVs belonging to clusters 1, 2 and 3 (Figure 5). The results of this analysis are displayed in
170 the sunburst plot of Figure 5b, which shows that some taxa were enriched in one or more
171 clusters. Significantly enriched taxa were reported from the highest taxonomic rank
172 considered (namely the domain level, the centre of the plot) to the lowest available taxonomic
173 level (namely genus, for bacteria/archaea and species for fungi) and were hierarchically
174 ordered. The intestinal cluster (i.e., cluster 1) was enriched in members of the Firmicutes

175 phylum, more specifically members of Bacilli class including *Candidatus Hepatoplasma*,
176 *Candidatus Bacilloplasma*, *Lactovum* and *Lactococcus* (Figure 5b). Members of the phylum
177 Bacteroidota and some taxa related to Protebacteria and Actinobacteriota were enriched in
178 both clusters 1 and 2. Within the phylum Bacteroidota, genera *Roseimarinus* and *Niabella*
179 were associated with cluster 1, *Fluviicola* and *Moheibacter* were related to cluster 2, and
180 *Leadbetterella* was shared between both clusters (Figure 5b). Other clear associations were
181 highlighted, with cluster 3 mainly associated with Proteobacteria and Ascomycota, while
182 cluster 2 mainly related to Actinobacteriota. The genera of these latter phyla are not
183 exclusively associated to a single cluster and are present in multiple clusters. Cluster 4
184 showed a taxonomic assignment totally related to the Fungi kingdom, in particular to the
185 phylum Ascomycota.

186 Molecular functions inferred from 16S rRNA gene amplicons showed that bacterial structures
187 associated with crab's organs are not only taxonomically defined but also functionally defined
188 (Figure 6a). Clusters of bacterial variants were mainly enriched/depleted by a unique set of
189 molecular functions that were absent in other clusters. The inferred genomic content of ASVs
190 detected in crab's gut (Cluster 1) was enriched by 359 GO terms and depleted by 21, with
191 more than a half (218 terms, 18 depleted and 200 enriched, Table S8) peculiar only to this
192 cluster (Figure 6b). Enzymes involved in the hydrolysis of complex carbohydrates—such as
193 cellulases and xylanases—and commonly associated to the gut of model herbivorous
194 arthropods were not enriched in the gut microbiome of the crab (Table S9). The gills (Cluster
195 2) had a population of bacteria enriched by 231 GO terms and depleted by 14 terms, with
196 roughly one third (88 terms, 12 depleted and 76 enriched) significantly found only in these
197 organs (Figure 6b). Functions associated to biofilm formation such as cellulose synthase (*30*,
198 *31*) were enriched only in this cluster (GO:0016760), whereas functions associated to the
199 nitrogen cycle—such as nitric- and nitrous-oxide reductase activity—were shared between

200 the gills and the environmental cluster 3. The latter, mainly composed of ASVs mostly found in
201 water and water debris samples, had the largest set of GO terms (401 terms, 22 depleted and
202 379 enriched) with more than a half (223 terms, 19 depleted and 204 enriched) exclusively
203 present in the inferred genome of ASVs highly abundant in that cluster (Figure 6b). Only
204 4.87% of the total GO terms was significantly enriched in all clusters (37 terms out of 759,
205 with no terms significantly depleted), indicating a functional role of microbial communities
206 both in crab's organs and in environmental samples (Figure 6b and c, and Table S8). ASVs
207 enriched in crab's gut and gills shared more functions (GO terms) with the environmental
208 cluster 3 (water and water debris) than between themselves (Figure 6b). In particular, 73 and
209 68 functions (with only 2 and 1 depleted terms) were shared between the gut and the gills,
210 respectively, and the ASVs detected in the environmental cluster 3 (Figure 6b, Table S8).
211 Considering the total number of terms significantly enriched/depleted, the gills were the
212 organs more impacted by the environment with roughly one third (27.8%) of molecular
213 functions shared with cluster 3. In contrast, the more populated organ, the gut, shared only
214 19.2% of the total number of functions with the environmental cluster.

215 Discussion

216 Our results show the presence of a specific microbiome associate to the semi-terrestrial
217 brachyuran crab *C. haematocheir*, strikingly different from that of the surrounding
218 environment This difference is more pronounced for Bacteria than Fungi. Further, the crab
219 microbiome profiling reveals the presence of consistent organ-specific microbial
220 communities. *C. haematocheir* represents one of the most recent attempts of land colonisation
221 by an arthropod. Consequently, the presence of a specific microbiota suggests its possible role
222 in the host's adaptation to terrestrial environment. The degree of specificity for both the
223 taxonomic and functional attributes in different organs can be explained by the adaptive

224 evolution of symbiotic host-microbe associations, which now represent a single biological unit
225 under selection from the recently colonised terrestrial habitat, i.e., a holobiont (*sensu 32*).
226 Temporary host-microbiota associations can be created, retained, or lost over time due to the
227 large number of organisms-microbe interactions. Thus, a crucial test needed to verify the
228 holobiont as a single selection unit is to assess the degree of specificity between the microbial
229 community and its host organism. Our sampling design and analytical approach were able to
230 define stable, organ-specific microbiota in both the gut and gills of individuals belonging to
231 different populations, suggesting a tight and persistent association between those microbial
232 communities and *C. haematocheir*. The organ specificity is supported by the observation that
233 such microbiota were consistently different among the different organs but importantly their
234 composition remained similar across sampling sites, where local microbial assemblages
235 proved to be site specific and different from those associated with the crabs.

236 The most enriched microorganisms in the gut microbiota of our model species are shared
237 with terrestrial isopods, which moved to the land roughly 300 million years ago, and are
238 supposed to play a central role in their adaptation to a diet based on vascular plant tissues
239 (*33*). One of the general assumptions of the hologenome concept, however, is that both the
240 host genome and the genetic information of the associated microbiome are transferred from
241 one generation to the next. This postulation underlies the critical role of mothers in
242 transferring microorganisms to their offspring (*34*), a challenging process for our model
243 organism that has highly dispersive planktonic larval stages followed by a recruitment phase.
244 Another crucial result is that fungal communities are exclusive to the environmental matrices
245 while almost absent in *C. haematocheir* organs. This suggests a specific selection process that
246 favours bacteria versus fungi, possibly supported by selection against yeasts and fungi since
247 several fungi are known pathogens for arthropods and crabs (*21*, *35*). To the best of our
248 knowledge, the present integrated set of results support the idea that semi-terrestrial crabs,

249 only recently migrated onto the land, are associated with microbial communities specifically
250 selected in response to the dramatic environmental pressures posed by the sea-to-land
251 transition. The next questions are where these microbes were acquired and where are their
252 reservoirs?

253 In the last decades there has been a proliferation of experimental studies supporting the view
254 of holobionts (hosts and their associated microbiota) as units of selection (*36*). These studies
255 showed how associated microbes have a central role in the host biology, ecology, and
256 evolution, since microbiota are genetically more dynamic and can change more rapidly than
257 the host genome in response to environmental pressures (*32*, *37*). Undeniably, stable
258 symbiotic insect-microbe interactions have been extensively investigated and proved to be
259 crucial for the evolution of many of the several lifestyles of insects (*38*, *39*). Studies on insects,
260 however, cannot shed light on the importance of such associations for the first steps towards
261 land colonisation, which was carried out by insects as far as the Ordovician (*4*). To bridge this
262 knowledge gap, in this study we focused on a semi-terrestrial brachyuran crab that, from a
263 paleontological point of view, just recently had to face the dramatic challenges represented by
264 excretion, breathing and digestion in terrestrial habitats (for a review see *8*).

265 The digestive system of brachyuran crabs comprises different tracts according to their
266 embryological origin and functional role. We intentionally focused on the midgut -where
267 central digestion and absorption occur- and the hindgut -which has a role in water and ion
268 transport (*8*, *17*). In contrast to the intestinal bacterial communities found in the freshwater
269 Chinese mitten crab, *Eriocheir sinensis* (*40*), and notwithstanding the different functions, the
270 selected intestinal tracts of *C. haematocheir* were homogeneous in terms of microbial
271 communities, suggesting, at least for this terrestrial crab species, a similar role of such
272 communities.

273 Host-associated microbiota are known to play a crucial role in the digestive process of many
274 terrestrial arthropods (*37*). Thus, it is not surprising that the highly diverse microbial
275 community found in *C. haematocheir* gut is clearly distinguished from all the other internal
276 and environmental assemblages. This semi-terrestrial species mainly relies on leaf litter and
277 supplements its diet with small arthropod preys (*28*, *29*). The ability to cope with difficult-to-
278 digest vascular plant compounds (i.e., cellulose, lignin, and polyphenols) has been brought
279 forward as a fundamental trait in the adaptive processes related to terrestrialisation (*17*, *41*,
280 *42*). The termites' ability to degrade lignocellulose is classical example of high gut microbiome
281 specialization (*37*). Conversely to the termites' gut microbiome, however, the bacterial
282 assemblages inhabiting the gut of *C. haematocheir* were not enriched in known functions that
283 would help the host to selectively digest cellulose products. Indeed, no gene with cellulase
284 activity were enriched in the gut, possibly reflecting a more variable diet of these crabs, not
285 strictly specialized on plant material digestion (*28*, *29*). The gut microbiome shared limited
286 similarities with the gut core microbiome of a marine predatory brachyuran, the
287 commercially important mud crab *Scylla paramamosain* (*43*). Core Proteobacteria shared by
288 *S. paramamosain* and *C. haematocheir* include only *Shewanella*, while Firmicutes include
289 *Lactococcus, Candidatus Hepatoplasma* and *Candidatus Bacilloplasma* (hereafter
290 *Hepatoplasma* and *Bacilloplasma*). *Hepatoplasma,* and *Bacilloplasma* are rare in marine
291 environments, while they were previously described as colonizers of the hepatopancreas of
292 the terrestrial isopod *Porcellio scaber* (*33*, *44*). *Hepatoplasma* was detected in *P. scaber*
293 specimens collected from very different geographical areas (Germany and western Canada),
294 proving the strong host-microbe association (*44*, *45*)., Other bacterial genera found in our
295 intestinal cluster (such as *Niabella*, *Paracoccus* and *Shewanella*) were also isolated from
296 terrestrial isopods (*41*). Of particular interest is also the presence of *Lactococcus*, which is a
297 genus of lactic acid bacteria known as homofermenters, producing a single product, lactic

298 acid, as the main, or only product of glucose fermentation. Lactic acid has a selective role since
299 it lowers the pH and selects the environmental microbes that can potentially thrive in the gut.
300 These microorganisms, common in the dairy industry, are also known to colonize the gut of
301 termites, *Lactococcus nasutitermitis* (*46*) and do not include marine species, although they are
302 occasionally described as pathogens of aquacultured species (*47*, *48*). Another microorganism
303 we found in the intestinal cluster that has never been described in the marine environment,
304 thus suggesting acquisition from terrestrial microorganisms, is *Lactovum*, a genus of bacteria
305 within the family Streptococcaceae. The genus includes a single species, *Lactovum miscens*, an
306 aerotolerant, anaerobic species originally isolated from soil of the Stiegerwald forest in
307 Germany (*49*). In addition, the intestinal cluster was enriched in *Erwiniaceae*, a family that
308 includes insect's symbionts such as *Buchnera aphidicola* (*50*). The gut microbiota we found in
309 the mid- and hindgut of *C. haematocheir* then shared similarities with the gut microbiota of
310 other terrestrial crustaceans and insects, suggesting the presence of functions/preferences of
311 those taxa that could help arthropods digestive systems, a hypothesis that deserves further
312 functional investigations.

313 The possibilities of a vertical transmission vs. acquisition from the environment of this
314 specialized gut microbiota deserves a deeper discussion and further experiments. First, land
315 and terrestrial crabs are neither social or gregarious and only few of them show some basic
316 degrees of parental care (*18*, *51*). Second, *C. haematocheir*, as well as the most land-adapted
317 family of brachyuran crabs, the Gecarcinidae, still retains an indirect development strategy
318 and releases planktonic larvae in coastal waters. Consequently, *C. haematoicher*, as well as the
319 land crabs with planktonic larvae in general, may have developed intimate gut-specific
320 symbiotic associations only through pseudo-vertical transmission (sensu *53*), i.e., by acquiring
321 and selecting their gut microbiota directly from food or, most probably, through consumption
322 of adults' faeces (coprophagy) found in or around their burrows. The megalopae of various

323 terrestrial and land crabs are known to specifically recruit at their spawning grounds, as
324 showed by the iconic mass recruitment of the Christmas Island red crab *Gecarcoidea natalis*
325 (*53*), or in areas where they can chemically detect adult populations, such as the case of the
326 swamp ghost crab *Ucides cordatus* (*54*). In some species, this selective recruitment towards
327 populations of conspecific adults ultimately results in the occurrence of juveniles and sub-
328 adults within secondary branches excavated along the burrows of adult crabs (*55–57*). Indeed,
329 these behaviours maximise the chances of finding adult populations and, consequently, a
330 successful pseudo-vertical transmission of microbiota through coprophagy. Further
331 metagenomic analyses on faeces are needed to clarify this point.

332

333 The gonads host a low microbial diversity with respect to the other organs and environmental
334 matrices we analysed, as expected for an internal organ in no direct contact with the
335 environment and not morphologically connected to the gut and the gills. Their microbial
336 community, moreover, is shared with both the other organs and the environment,
337 determining the absence of a specific taxonomic and functional cluster associated with them.
338 Microorganisms are known to be associated with the reproductive system of arthropods, in
339 both insects (*58*) and crustaceans (*59*). Few studies, however, have specifically characterised
340 these gonad-associated microbiota and most of them were focused on specific pathogens of
341 sexually transmitted infections or reproductive parasites, with the most notable example
342 being *Wolbachia* (*60*). When present, the vertical transmission of specific microbiota during
343 oogenesis or at birth is known to determine a colonization of the gonads that ultimately
344 affects the microbial composition of the offspring (*59*). In this view, the absence of an
345 exclusive microbiota in *C. haematoicher* gonads can be explained, as for the gut microbiota, by
346 the lack of vertical transmission through the parental-offspring pathways, because of the
347 presence of planktonic stages.

348 Although protected in the gill chambers, *C. haematocheir* gills are in direct communication
349 with the external environment, but they do show a unique resident microbiota, which is
350 different from the microbial communities detected in the environmental matrices. These gill-
351 associated microbial assemblages are only enriched in prokaryotes, unlike the soil/litter and
352 the water communities, which are composed almost entirely by fungi and a mix of
353 Proteobacteria and fungi, respectively. The prokaryotic cluster associated with the gills is
354 characterised by a strong uniformity in terms of both taxonomy and functions. Actinobacteria
355 associated to the gills include Microbacteriacea and *Illumatobacter*, which are known to play
356 an important role in marine organisms by producing bioactive compounds crucial in the
357 defence against pathogens (*61*). Some of the bacteria detected on the gills of *C. haematocheir*,
358 such as *Ilumatobacter* and *Albimonas*, were also found in the gills of the Chinese mitten crab
359 *Eriocheir sinensis* (*40*), but at this stage their metabolic functions are not clear. In contrast to
360 the gut, genes associated with cellulose synthase activity were found in the microbiota
361 associated with these organs (GO:0016760). This activity is typical of biofilm-forming bacteria
362 that use cellulose both as a physical barrier against armful molecules such as antibiotics but
363 also biocides and metallic cations, and as a molecular glue to help their interaction with the
364 host (*31*). Since the gills of *C. haematocheir* are exposed to external perturbations, the
365 presence of biofilm-forming functions may help to boost the resilience of bacteria stabilizing
366 host-microbiome interactions. Genes related to the reduction of nitric compounds (namely:
367 nitric oxide reductase activity, GO:0016966, and nitrous-oxide reductase activity,
368 GO:0050304) were enriched in both the gills and environmental water and water debris
369 samples. Besides their role in the anaerobic metabolism of nitrogen, these functions are
370 involved into pathogenesis and antibiotic resistance in bacteria (*62*) and may help tissue
371 colonization. In addition, the covalent incorporation of a nitric oxide molecule (nitrosylation)
372 of cytochrome c and quinol oxidases inhibits cellular respiration acting as an antimicrobial

373 molecule. We speculate that this mechanism may be used by bacteria inhabiting the gills of *C.*
374 *haematocheir* as a possible molecular defence against external pathogens helping the crab,
375 and themselves, to thrive in different terrestrial environments.

376 Our results show that fungi are strongly depleted in all crab's organs and that gills should be
377 considered an efficient selective filter between the environment and the host. In our opinion,
378 the substantial absence of fungal associations can be explained by different hypotheses that
379 all converge on the existence of defences from potential pathogens. We may hypothesize that
380 some physiological and anatomical characteristics of specific organs, such as gut and gills, may
381 inhibit adhesion and development of fungi. In insects, for instance, physiochemical
382 characteristics of the cuticle, or feeding habits, can counteract the entomopathogenic fungi
383 cuticular adhesion, germination of spores and hyphal growth (*63*). Moreover, bacterial
384 production of antifungal molecules may limit the growth of fungi. Events of fungal exclusion
385 by direct competition mediated by symbiotic bacteria have been extensively described in
386 beetles (*64*). *Streptomyces* sp. SPB74, symbiont of the beetle *Dendroctonus frontalis*, produces
387 an active compound, mycangimycin, which specifically inhibits antagonist fungi without
388 acting on the mutual ones (*64*). This protective role may be extended to the offspring.
389 Antifungal molecules produced by the symbiotic bacterium *Burkholderia gladioli* are known to
390 protect the eggs of the beetle *Lagria villosa* against pathogenic fungi (*35*, *65*). Pathogenic fungi
391 have been reported in intertidal and terrestrial crabs, such as the case of the swamp ghost
392 crab, *U. cordatus*, infected by species of black yeast, *Exophiala cancerae*, and *Fonsecaea*
393 *brasiliensis*, which cause a condition called "lethargic crab disease" (*21*, *22*). The role of
394 prokaryotic associations in the defence against pathogenic fungi can be of critical importance
395 for terrestrial crabs, since the immune molecules of crustaceans are less efficient than those of
396 insects against fungal infections (*66*).

397 In conclusion, this is the first attempt to ascertain the role of host-microbiome associations in
398 the transition to land of brachyuran crabs, the most recent arthropod taxa to perform such an
399 evolutionary leap. We found clear differentiations among the microbiota associated with
400 different crab's organs and the ones found in the environmental matrices, suggesting a
401 selective acquisition of the microbiota through "gain" and "loss" mechanisms. The differences
402 found among the organ-related and environmental clusters were not merely taxonomical
403 since the different clusters harbour different metabolic profiles. These results corroborate the
404 hypothesis that the recorded differences can be due to the presence of metabolic
405 complementation mechanisms that took place in those organs mostly impacted by the
406 challenges posed by the terrestrial environment. Our data also show possible evolutionary
407 convergences towards a uniform 'terrestrial intestinal microbiota' across different lineages of
408 arthropods, suggesting the presence of bacterial associations linked to terrestrial life. The
409 present study demonstrates mechanisms of specific microbial selection in the sesarmid crab
410 *C. heamatocheir* and strongly supports the hypothesis that semi- and terrestrial crabs are an
411 appropriate model system to study the evolution of arthropod-microbe interaction under the
412 selective pressures posed by the sea-land transition, which is happing right now in Brachyura
413 (*23*).

414

415 # Material and methods

416 ## Study species

417 *Chiromantes haematocheir* (Decapoda; Brachyura; Sesarmidae) is a semi-terrestrial crab
418 colonising the coastal vegetated areas from Taiwan to South East Asia (*28*, *67*). In Hong Kong,
419 it forms large populations in areas of lowland secondary forest adjacent to mangroves and in
420 pockets of riverine forests, where it was also observed climbing trees. Very little is known
421 about its ecology, apart from the fact that it digs deep burrows and, as in many sesarmids, it
422 releases pelagic larvae into the ocean. This species was selected because it is the most land
423 adapted among the Hong Kong brachyuran crabs.

424 ## Sample collection and total DNA extraction

425 We selected three large populations of *C. haematocheir* that colonised distant catchments
426 across the Hong Kong territory. The selected populations were sampled at Shui Hau (Southern
427 Lantau Island), To Kwa Peng (Eastern coast of Sai Kung Country Park, New Territories), and
428 Sai Keng (Three Fathoms Cove, Tolo Harbour New Territories). From each site, eleven sexual
429 mature adult crabs (carapace width range between 13.4 mm and 32.2 mm) were collected in
430 October 2018. To explore differences in microbial community composition across the various
431 crab's organs and the surrounding environment, we collected and analysed four organ
432 samples (gills, hindgut, midgut and gonads) and four environmental matrices (sediment, leaf
433 litter, freshwater and freshwater debris). Due to the intensive sampling, all environmental
434 samples and alive crabs were immediately frozen and subsequently transported to the
435 laboratories of the Division of Ecology and Biodiversity (The University of Hong Kong). The
436 dissections were then performed under sterile conditions. All dissection instruments were
437 sterilized over an open flame to eliminate residual DNA and washed with 75% EtOH to
438 prevent cross-contamination. After removing the carapace, gills, gonads, hindgut and midgut

439 from each crab were excised under a stereomicroscope and stored at -20°C in RNAlater
440 (Thermo Fisher Scientific) stabilization solution until DNA extraction.

441 Total DNA extraction from crab organs, sediment, leaf litter and freshwater debris was
442 performed using the DNeasy PowerLyzer PowerSoil Kit (QIAGEN) following manufacturer's
443 protocol. Total DNA extraction from water samples was performed using the DNeasy
444 PowerWater Kit (QIAGEN) following manufacturer's protocol, after having filtered 100ml of
445 freshwater through 0.2µm Thermo Scientific™ Nalgene™ Sterile Analytical Filter Units
446 (Thermo Fisher Scientific). Extracted DNA samples were stored at -20°C. Before the DNA
447 library's preparation, DNAs were quantified fluorometrically by using Qubit dsDNA HS Assay
448 Kit (Thermo Fisher Scientific).

449 **16S (V3-V4) rRNA gene amplification and sequencing**

450 The preparation and sequencing of the 16S library were performed at Laboratory of
451 Advanced Genomics, Department of Biology, University of Florence (Firenze, Italy). PCR
452 amplifications of the bacterial V3-V4 16S rRNA gene fragments were performed using KAPA
453 HiFi HotStart ReadyMix (Roche) and the primer pair 341F (5'-CCTACGGGNGGCWGCAG-3')
454 and 805R (5'-GACTACNVGGGTWTCTAATCC-3') (*68*) with overhang Illumina adapters. 16S
455 amplicon PCR protocol was set on 25 µl of final volume. In detail for each reaction, 12.5 µl of
456 2× KAPA HiFi HotStart ReadyMix (Roche), 10 µl of 1µM forward and reverse primers and 2.5
457 µl of template DNA (5–20 ng/µl) were combined. Amplicon PCR reaction was performed using
458 the GeneAmp PCR System 2700 (Thermo Fisher Scientific) and the following cycling
459 conditions: denaturation step at 95 °C for 3 minutes; 35 (gonads, water and water debris) and
460 25 (hindgut, midgut, gills, soil and litter) cycles: at 95 °C for 30 seconds, 55 °C for 30 seconds,
461 72 °C for 30 seconds; final extension step at 72 °C for 5 minutes. All PCR products were
462 checked through electrophoresis on 1.5% agarose gel and then purified using KAPA Pure
463 Beads (Roche) following the manufacturer's instructions. To apply the Illumina adapters

464 sequencing indexing using Nextera XT Index Kit V2 (Illumina), a second PCR amplification
465 was then performed by preparing a reaction mix in accordance with the 16S metagenomic
466 library preparation protocol (*69*). An indexing step was made for all samples by seven PCR
467 cycles. Amplicon products from indexing PCR were purified using KAPA Pure Beads (Roche)
468 and their quality check was performed using Agilent 2100 Bioanalyzer (Agilent Technologies)
469 with Agilent DNA 1000 Kit (Agilent Technologies). Subsequently, concentration check was
470 performed by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Finally, the barcoded
471 libraries were balanced and pooled at equimolar concentration, before being sequenced on an
472 Illumina MiSeq (PE300) platform (MiSeq Control Software 2.6.2.1).

473 **ITS1 rDNA region sequencing**

474 ITS1 library preparation and sequencing were performed at the Research and Innovation
475 Centre, Fondazione Edmund Mach (FEM) (S. Michele all'Adige, Trento, Italy). Fungal ITS1
476 fragments were amplified by PCR using the FastStart High Fidelity PCR System (Roche) for
477 environment matrixes and the Hot Start High-Fidelity DNA Polymerase (NEB) for animal
478 matrixes following the manufacturer instructions using the primers ITS1F (5'-
479 CTTGGTCATTTAGAGGAAGTAA-3') (*70*) and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3') (*71*) with
480 overhang Illumina adapters. ITS1 PCRs were performed in 25 µl of final volume. In detail for
481 each reaction, 2ul of 10uM forward and reverse primers were used in combination with 1 µl
482 of template DNA (5-20 ng/ul). Amplicon PCR reaction was performed using the GeneAmp PCR
483 System 9700 (Thermo Fisher Scientific) and the following cycling conditions: denaturation
484 step at 95 °C for 3 minutes; 25 (litter) and 33 (all other samples) cycles: at 95 °C for 20
485 seconds, 50 °C for 45 seconds, 72 °C for 90 seconds; final extension step at 72 °C for 10
486 minutes. All PCR products were checked on 1.5% agarose gel and purified using the CleanNGS
487 kit (CleanNA, the Netherlands) following the manufacturer's instructions. Subsequently a
488 second PCR was performed to apply the Illumina sequencing adapters Nextera XT Index

489 Primer (Illumina) (*69*). The indexing step was made for all samples by seven PCR cycles. After
490 Indexing PCR amplicon libraries were purified using the CleanNGS kit (CleanNA, the
491 Netherlands), and the quality control was performed on a Typestation 2200 platform (Agilent
492 Technologies, Santa Clara, CA, USA). Afterwards all barcoded libraries were mixed at
493 equimolar concentration, quantified by qPCR Kapa Library quantification kit (Roche) and
494 sequenced on an Illumina MiSeq (PE300) platform (MiSeq Control Software 2.5.0.5 and Real-
495 Time Analysis software 1.18.54.0).

496 **Amplicon sequence variant inference**

497 The DADA2 pipeline version 1.14.1 (*72*) was used to infer amplicon sequence variants (ASVs)
498 from raw sequences. Primers used for PCR amplification were removed using cutadapt
499 version 1.15 (*73*) in paired-end mode. If a primer was not found, the sequence was discarded
500 together with its mate to reduce possible contamination. For ITS amplicon sequences reads
501 containing both the forward and reverse primers were considered valid only if concordant
502 (one of the two primers must be present but reverse and complemented) with a cut-off length
503 of 70bp. Low quality reads were discarded using the "filterAndTrim" function with an
504 expected error threshold of 2 for both forward and reverse read pairs (namely only reads
505 with more than 2 expected errors were removed). Denoising was performed using the "dada"
506 function after error rate modelling ("learnErrors" function). Denoised reads were merged
507 discarding those with any mismatches and/or an overlap length shorter than 20bp
508 ("mergePairs" function). Chimeric sequences were removed using the
509 "removeBimeraDenovo" function whereas taxonomical classification was performed using
510 DECIPHER package version 2.14.0 against the latest version of the pre-formatted Silva small-
511 subunit reference database (*74*) (SSU version 138 available at:
512 http://www2.decipher.codes/Downloads.html) and the Warcup database for fungal ITS1
513 (*75*). All variants not classified as Bacteria, Archaea or Fungi were removed together with

514 sequences classified as chloroplasts or mitochondria (16S rRNA sequences only). Additional
515 information on the sequence variant inference pipeline used were reported in Supplementary
516 Materials.

517 **Inferring functional content of amplicon variants**

518 The genome content of bacterial ASVs was inferred using PICRUSt2 pipeline (*76*). Enzyme
519 Commission Numbers (EC numbers) were converted into Gene Ontology terms (GO terms)
520 using the mapping file available at: http://www.geneontology.org/external2go/ec2go. Gene
521 abundance was retrieved using the "--strtified" option to report gene abundances at species
522 level (ASVs. Additional information about functional content inference was reported in
523 Supplementary Material.

524 **Statistical analyses**

525 All statistical analyses were performed in the R environment (version 3.6). Briefly, alpha- and
526 beta-diversity analyses were conducted using the vegan package version 2.5 (*77*) in
527 combination with the iNEXT package version 2.0 (*78*). Normalization and differential
528 abundance analyses were performed with DESeq2 version 1.28 (*79*) whereas enrichment
529 analysis of taxa and functions was performed using hypergeometric test ("phyper" function of
530 R stats package). For additional details about tests data manipulation see Supplementary
531 Materials.

532 **References**

533 1. C. Little, *The Terrestrial Invasion. An Ecophysiological Approach to the Origins of Land*
534 *Animals* (Cambridge University Press, Cambridge, 1990).

535 2. C. Little, *The Colonisation of Land. Origins and Adaptations of Terrestrial Animals*
536 (Cambridge University Press, Cambridge, 2009).

537 3. D. J. Randall, W. W. Burggren, A. P. Farrell, M. S. Haswell, *The Evolution of Air Breathing*
538 *in Vertebrates* (Cambridge University Press, Cambridge, 2009).

539 4. J. Lozano-Fernandez, R. Carton, A. R. Tanner, M. N. Puttick, M. Blaxter, J. Vinther, J.
540 Olesen, G. Giribet, G. D. Edgecombe, D. Pisani, A molecular palaeobiological exploration

541 of arthropod terrestrialization. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150133 (2016).

542 5. F. Giomi, M. Fusi, A. Barausse, B. Mostert, H.-O. Pörtner, S. Cannicci, Improved heat
543 tolerance in air drives the recurrent evolution of air-breathing. *Proc. R. Soc. B Biol. Sci.*
544 **281**, 20132927 (2014).

545 6. M. M. Drew, S. Harzsch, M. Stensmyr, S. Erland, B. S. Hansson, A review of the biology
546 and ecology of the Robber Crab, Birgus latro (Linnaeus, 1767) (Anomura:
547 Coenobitidae). *Zool. Anz.* **249**, 45–67 (2010).

548 7. J. Krieger, P. Braun, N. T. Rivera, C. D. Schubart, C. H. G. Müller, S. Harzsch, Comparative
549 analyses of olfactory systems in terrestrial crabs (Brachyura): Evidence for aerial
550 olfaction? *PeerJ*. **3**, e1433 (2015).

551 8. W. W. Burggren, B. R. McMahon, *Biology of the land crabs* (Cambridge University Press,
552 Cambridge, 1988).

553 9. C. A. Farrelly, P. Greenaway, Land crabs with smooth lungs: Grapsidae, Gecarcinidae,
554 and Sundathelphusidae ultrastructure and vasculature. *J. Morphol.* **215**, 245–260
555 (1993).

556 10. S. C. Faria, R. O. Faleiros, F. A. Brayner, L. C. Alves, A. Bianchini, C. Romero, R. C.
557 Buranelli, F. L. Mantelatto, J. C. McNamara, Macroevolution of thermal tolerance in
558 intertidal crabs from Neotropical provinces: A phylogenetic comparative evaluation of
559 critical limits. *Ecol. Evol.* **7**, 3167–3176 (2017).

560 11. C. A. Farrelly, P. Greenaway, Gas Exchange Through the Lungs and Gills in Air-Breathing
561 Crabs. *J. Exp. Biol.* **187**, 113–130 (1994).

562 12. F. Paoli, C. S. Wirkner, S. Cannicci, The branchiostegal lung of Uca vocans (Decapoda:
563 Ocypodidae): Unreported complexity revealed by corrosion casting and MicroCT
564 techniques. *Arthropod Struct. Dev.* **44**, 622–629 (2015).

565 13. C. M. Wood, R. G. Boutilier, Osmoregulation, ionic exchange, blood chemistry, and
566 nitrogenous waste excration in the land crab Cardisoma carnifex: a field and laboratory
567 study. *Biol. Bull.* **169**, 267–290 (1985).

568 14. P. Greenaway, in *Biology of land crabs*, W. W. Burggren, B. R. McMahon, Eds. (Cambridge
569 University Press, New York, ed. 1, 1988), pp. 211–248.

570 15. S. Cannicci, R. Simoni, F. Giomi, Role of the embryo in crab terrestrialisation: an
571 ontogenetic approach. *Mar. Ecol. Prog. Ser.* **430**, 121–131 (2011).

572 16. R. Simoni, F. Giomi, D. Spigoli, H.-O. Pörtner, S. Cannicci, Adaptations to semi-terrestrial
573 life in embryos of East African mangrove crabs: a comparative approach. *Mar. Biol.* **160**,
574 2483–2492 (2013).

575 17. E. S. Lindquist, K. W. Krauss, P. T. Green, D. J. O'Dowd, P. M. Sherman, T. J. Smith III, Land
576 crabs as key drivers in tropical coastal forest recruitment. *Biol. Rev.* **84**, 203–223
577 (2009).

578 18. K. Anger, The conquest of freshwater and land by marine crabs: Adaptations in life-
579 history patterns and larval bioenergetics. *J. Exp. Mar. Bio. Ecol.* **193**, 119–145 (1995).

580 19. S. Morris, Neuroendocrine regulation of osmoregulation and the evolution of air-
581 breathing in decapod crustaceans. *J. Exp. Biol.* **204**, 979–989 (2001).

582 20. D. Weihrauch, S. Morris, D. W. Towle, Ammonia excretion in aquatic and terrestrial
583 crabs. *J. Exp. Biol.* **207**, 4491–504 (2004).

584 21. V. A. Vicente, R. Orélis-Ribeiro, M. J. Najafzadeh, J. Sun, R. S. Guerra, S. Miesch, A.
585 Ostrensky, J. F. Meis, C. H. Klaassen, G. S. de Hoog, W. A. Boeger, Black yeast-like fungi
586 associated with Lethargic Crab Disease (LCD) in the mangrove-land crab, Ucides
587 cordatus (Ocypodidae). *Vet. Microbiol.* **158**, 109–122 (2012).

588 22. W. A. Boeger, M. R. Pie, A. Ostrensky, L. Patella, Lethargic crab disease: Multidisciplinary
589 evidence supports a mycotic etiology. *Mem. Inst. Oswaldo Cruz.* **100**, 161–167 (2005).

590 23. S. Cannicci, S. Fratini, N. Meriggi, G. Bacci, A. Iannucci, A. Mengoni, D. Cavalieri, To the
591 Land and Beyond: Crab Microbiomes as a Paradigm for the Evolution of
592 Terrestrialization. *Front. Microbiol.* **11**, 575372 (2020).

593 24. L. Margulis, R. Fester, Eds., *Symbiosis as a Source of Evolutionary Innovation. Speciation*
594 *and Morphogenesis* (MIT Press, Cambridge, MA, 1991).

595 25. I. Zilber-Rosenberg, E. Rosenberg, Role of microorganisms in the evolution of animals
596 and plants: The hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723–735
597 (2008).

598 26. E. Rosenberg, I. Zilber-Rosenberg, The hologenome concept of evolution after 10 years.
599 *Microbiome*. **6**, 78 (2018).

600 27. E. Rosenberg, O. Koren, L. Reshef, R. Efrony, I. Zilber-Rosenberg, The role of
601 microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* **5**, 355–362
602 (2007).

603 28. K. F. Abd El-Wakeil, Trophic structure of macro- and meso-invertebrates in Japanese
604 coniferous forest: Carbon and nitrogen stable isotopes analyses. *Biochem. Syst. Ecol.* **37**,
605 317–324 (2009).

606 29. K. Miyake, K. Ura, S. Chida, Y. Ueda, Y. Baba, T. Kusube, S. Yanai, Guaiacol oxidation
607 activity of herbivorous land crabs, Chiromantes haematocheir and Chiromantes
608 dehaani. *J. Biosci. Bioeng.* **128**, 316–322 (2019).

609 30. R. V Augimeri, A. J. Varley, J. L. Strap, Establishing a Role for Bacterial Cellulose in
610 Environmental Interactions: Lessons Learned from Diverse Biofilm-Producing
611 Proteobacteria. *Front. Microbiol.* **6**, 1282 (2015).

612 31. H.-C. Flemming, J. Wingender, The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–633
613 (2010).

614 32. J.-C. Simon, J. R. Marchesi, C. Mougel, M.-A. Selosse, Host-microbiota interactions: from
615 holobiont theory to analysis. *Microbiome*. **7**, 5 (2019).

616 33. S. Fraune, M. Zimmer, Host-specificity of environmentally transmitted Mycoplasma-like
617 isopod symbionts. *Environ. Microbiol.* **10**, 2497–2504 (2008).

618 34. E. Rosenberg, I. Zilber-Rosenberg, The hologenome concept of evolution: do mothers

619 matter most? *BJOG An Int. J. Obstet. Gynaecol.* **127**, 129–137 (2020).

620 35. L. V Flórez, K. Scherlach, I. J. Miller, A. Rodrigues, J. C. Kwan, C. Hertweck, M. Kaltenpoth,
621 An antifungal polyketide associated with horizontally acquired genes supports
622 symbiont-mediated defense in Lagria villosa beetles. *Nat. Commun.* **9**, 2478 (2018).

623 36. E. Rosenberg, G. Sharon, I. Zilber-Rosenberg, The hologenome theory of evolution
624 contains Lamarckian aspects within a Darwinian framework. *Environ. Microbiol.* **11**,
625 2959–2962 (2009).

626 37. A. Brune, C. Dietrich, The Gut Microbiota of Termites: Digesting the Diversity in the
627 Light of Ecology and Evolution. *Annu. Rev. Microbiol.* **69**, 145–166 (2015).

628 38. T. Kuriwada, T. Hosokawa, N. Kumano, K. Shiromoto, D. Haraguchi, T. Fukatsu,
629 Biological Role of Nardonella Endosymbiont in Its Weevil Host. *PLoS One.* **5**, e13101
630 (2010).

631 39. N. Nikoh, T. Tsuchida, T. Maeda, K. Yamaguchi, S. Shigenobu, R. Koga, T. Fukatsu,
632 Genomic Insight into Symbiosis-Induced Insect Color Change by a Facultative Bacterial
633 Endosymbiont, "Candidatus Rickettsiella viridis." *MBio.* **9**,
634 e00890-18 (2018).

635 40. J. Dong, X. Li, R. Zhang, Y. Zhao, G. Wu, J. Liu, X. Zhu, L. Li, Comparative analysis of the
636 intestinal bacterial community and expression of gut immunity genes in the Chinese
637 Mitten Crab (Eriocheir sinensis). *AMB Express.* **8** (2018), doi:10.1186/s13568-018-
638 0722-0.

639 41. D. Bouchon, M. Zimmer, J. Dittmer, The Terrestrial Isopod Microbiome: An All-in-One
640 Toolbox for Animal–Microbe Interactions of Ecological Relevance. *Front. Microbiol.* **7**,
641 1472 (2016).

642 42. M. Zimmer, Nutrition in terrestrial isopods (Isopoda: Oniscidea): An evolutionary-
643 ecological approach. *Biol. Rev.* **77**, 455–493 (2002).

644 43. H. Wei, H. Wang, L. Tang, C. Mu, C. Ye, L. Chen, C. Wang, High-throughput sequencing
645 reveals the core gut microbiota of the mud crab (Scylla paramamosain) in different
646 coastal regions of southern China. *BMC Genomics.* **20**, 1–12 (2019).

647 44. Y. Wang, U. Stingl, F. Anton-erxleben, S. Geisler, A. Brune, M. Zimmer, "Candidatus
648 Hepatoplasma crinochetorum," a New, Stalk-Forming Lineage of. *Appl. Environ.*
649 *Microbiol.* **70**, 6166–6172 (2004).

650 45. Y. Wang, A. Brune, M. Zimmer, Bacterial symbionts in the hepatopancreas of isopods:
651 Diversity and environmental transmission. *FEMS Microbiol. Ecol.* **61**, 141–152 (2007).

652 46. S. Yan Yang, Y. Zheng, Z. Huang, X. Min Wang, H. Yang, Lactococcus nasutitermitis sp.
653 nov. isolated from a termite gut. *Int. J. Syst. Evol. Microbiol.* **66**, 518–522 (2016).

654 47. A. Williams, Lactococcus piscium sp. nov. a new Lactococcus species from salmonid fish.
655 *FEMS Microbiol. Lett.* **68**, 109–113 (1990).

656 48. R. M. Sakala, H. Hayashidani, Y. Kato, C. Kaneuchi, M. Ogawa, Isolation and
657 characterization of Lactococcus piscium strains from vacuum-packaged refrigerated
658 beef. *J. Appl. Microbiol.* **92**, 173–179 (2002).

49. C. Matthies, A. Gößner, G. Acker, A. Schramm, H. L. Drake, Lactovum miscens gen. nov., sp. nov., an aerotolerant, psychrotolerant, mixed-fermentative anaerobe from acidic forest soil. *Res. Microbiol.* **155**, 847–854 (2004).

50. X. Chen, M. D. Hitchings, J. E. Mendoza, V. Balanza, P. D. Facey, P. J. Dyson, P. Bielza, R. Del Sol, Comparative Genomics of Facultative Bacterial Symbionts Isolated from European Orius Species Reveals an Ancestral Symbiotic Association . *Front. Microbiol.* . **8** (2017), p. 1969.

51. E. Hornung, Evolutionary adaptation of oniscidean isopods to terrestrial life: Structure, physiology and behavior. *Terr. Arthropod Rev.* **4**, 95–130 (2011).

52. D. M. Wilkinson, The Role of Seed Dispersal in the Evolution of Mycorrhizae. *Oikos.* **78**, 394 (1997).

53. A. M. Adamczewska, S. Morris, Ecology and Behavior of Gecarcoidea natalis , the Christmas Island Red Crab, During the Annual Breeding Migration. *Biol. Bull.* **200**, 305–320 (2001).

54. K. Diele, D. J. B. Simith, Effects of substrata and conspecific odour on the metamorphosis of mangrove crab megalopae, Ucides cordatus (Ocypodidae). *J. Exp. Mar. Bio. Ecol.* **348**, 174–182 (2007).

55. L. E. Agusto, S. Fratini, P. J. Jimenez, A. Quadros, S. Cannicci, Structural characteristics of crab burrows in Hong Kong mangrove forests and their role in ecosystem engineering. *Estuar. Coast. Shelf Sci.* **248**, 106973 (2021).

56. M. Vannini, S. Cannicci, R. Berti, G. Innocenti, Cardisoma carnifex (Brachyura): Where have all the babies gone? *J. Crustac. Biol.* **23**, 55–59 (2003).

57. T. Stieglitz, P. V. Ridd, P. Müller, Passive irrigation and functional morphology of crustacean burrows in a tropical mangrove swamp. *Hydrobiologia*. **421**, 69–76 (2000).

58. C. M. Gibson, M. S. Hunter, Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecol. Lett.* **13**, 223–234 (2010).

59. J. I. Perlmutter, S. R. Bordenstein, Microorganisms in the reproductive tissues of arthropods. *Nat. Rev. Microbiol.* **18**, 97–111 (2020).

60. F. Chevalier, J. Herbinière-Gaboreau, D. Charif, G. Mitta, F. Gavory, P. Wincker, P. Grève, C. Braquart-Varnier, D. Bouchon, Feminizing Wolbachia: a transcriptomics approach with insights on the immune response genes in Armadillidium vulgare. *BMC Microbiol.* **12 Suppl 1**, S1–S1 (2012).

61. K. Valliappan, W. Sun, Z. Li, Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl. Microbiol. Biotechnol.* **98**, 7365–7377 (2014).

62. A. Vázquez-Torres, A. J. Bäumler, Nitrate, nitrite and nitric oxide reductases: from the last universal common ancestor to modern bacterial pathogens. *Curr. Opin. Microbiol.* **29**, 1–8 (2016).

63. A. Ortiz-Urquiza, N. Keyhani, Action on the Surface: Entomopathogenic Fungi versus the

Insect Cuticle. *Insects.* **4**, 357–374 (2013).

64. D.-C. Oh, J. J. Scott, C. R. Currie, J. Clardy, Mycangimycin, a Polyene Peroxide from a Mutualist Streptomyces sp. *Org. Lett.* **11**, 633–636 (2009).

65. L. V Flórez, K. Scherlach, P. Gaube, C. Ross, E. Sitte, C. Hermes, A. Rodrigues, C. Hertweck, M. Kaltenpoth, Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat. Commun.* **8**, 15172 (2017).

66. T. H. Ng, J. Kurtz, Dscam in immunity: A question of diversity in insects and crustaceans. *Dev. Comp. Immunol.* **105**, 103539 (2020).

67. C. D. Schubart, P. K. L. Ng, Revision of the intertidal and semiterrestrial crab genera Chiromantes Gistel , 1848 , and Pseudosesarma Serène & Soh , 1970 (Crustacea : Brachyura : Sesarmidae), using morphology and molecular phylogenetics , with the establishment of nine new genera a. *Raffles Bull. Zool.* **68**, 891–994 (2020).

68. F. Rosso, V. Tagliapietra, D. Albanese, M. Pindo, F. Baldacchino, D. Arnoldi, C. Donati, A. Rizzoli, Reduced diversity of gut microbiota in two Aedes mosquitoes species in areas of recent invasion. *Sci. Rep.* **8**, 16091 (2018).

69. Illumina, *16S metagenomic sequencing library preparation protocol: preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. Part no. 15044223 Rev B* (Illumina, San Diego, CA., 2013; https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf).

70. M. Gardes, T. D. Bruns, ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).

71. T. J. White, T. Bruns, S. Lee, J. Taylor, in *PCR Protocols* (Elsevier, 1990; https://linkinghub.elsevier.com/retrieve/pii/B9780123721808500421), pp. 315–322.

72. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods.* **13**, 581–583 (2016).

73. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* **17**, 10 (2011).

74. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

75. V. Deshpande, Q. Wang, P. Greenfield, M. Charleston, A. Porras-Alfaro, C. R. Kuske, J. R. Cole, D. J. Midgley, N. Tran-Dinh, Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia*. **108**, 1–5 (2016).

76. G. M. Douglas, V. J. Maffei, J. R. Zaneveld, S. N. Yurgel, J. R. Brown, C. M. Taylor, C. Huttenhower, M. G. I. Langille, PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **38**, 685–688 (2020).

77. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R.

B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, *Vegan: community ecology package* (2019; http://cran.r-project.org/package=vegan).

78. T. C. Hsieh, K. H. Ma, A. Chao, iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).

79. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgments

We thank the friends of the Integrate Mangrove Ecology laboratory of the Swire Institute of Marine Science, HKU for their help in sampling and lab work. We also thank Prof Colin Little and Gray A. Williams for inspiring and discussing the original idea behind the study.

Author contributions: G.B. performed bioinformatics analyses and wrote the manuscript; S.F. supervised lab work and wrote the manuscript; N.M. performed 16S laboratory analyses and wrote the manuscript; C.L.Y.C. and K.H.N carried out the samplings and performed DNA extractions; D.C. and A.M. helped conceiving the original idea; and S.C. conceived the original idea, collected the samples and wrote the manuscript. S.C. and D.C provided funding for the experiments. All the authors commented and validated the manuscript, participated to plan the sampling design and critically discussed the results.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: data needed to evaluate the conclusions are present in the paper and in Supplementary Materials. All codes used in the work have been uploaded to a public Github repository available at: https://github.com/GiBacci/Chiromantes_haematocheir_microbiome. Sequencing data were uploaded to the European Nucleotides Archive (ENA) under the project ID: PRJEB43930.

Funding: This work was supported by the project "The sky's the limit: the irresistible ascent to land and trees by crabs", sponsored by TUYF Charitable Trust funds, Hong Kong, (HKU no 260008686.088562.26000.400.01) and by the Eighth Government Matching Grant, Hong Kong Government (HKU no 207080320.088562.26020.430.01) to S.C. and by the RAE Improvement Fund from the Faculty of Science, HKU (HKU no 000250449.088562.26000.100.01) to S.C and D.C.

Figures

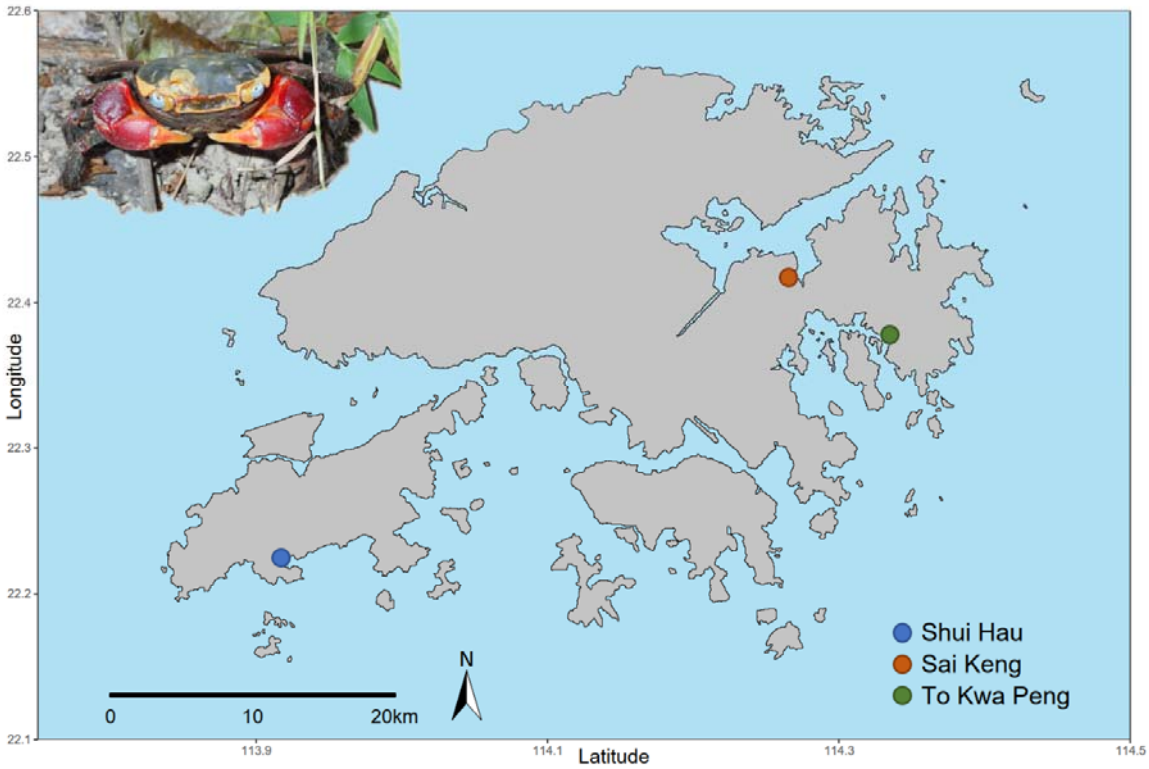


Figure 1. Geographical distribution of the sampled populations of *C. haematocheir*. The map shows the names and location of the three sampling sites visited for the study in the New Territories and on Lantau Island and an adult male *C. haematocheir* in its natural environment (top left).

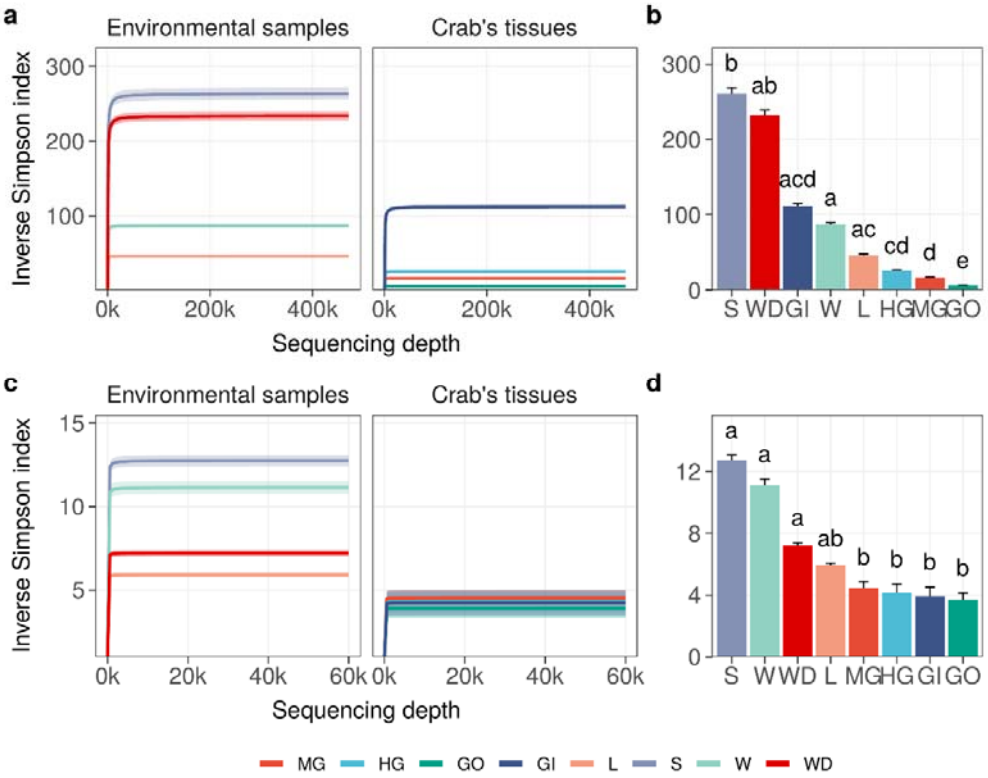


Figure 2: Microbial diversity in crab's organs and environmental samples. The average of the inverse Simpson index was reported with increasing sampling effort for all types of samples. Interpolated and extrapolated diversity was reported in panel a and c (16S rRNA gene and ITS-1 region, respectively), whereas observed diversity was reported in panel b and d (16S rRNA gene and ITS-1 region respectively). Significant differences in microbial diversity (Wilcoxon non-parametric test) were reported using lowercase letters (panel b and d) whereas colours and acronyms on the x-axis correspond to different sample types (MG, midgut; HG, hindgut; GO, gonads; GI, gills; L, litter; S, soil; W, water; WD, water debris). If two means were significantly different, all letters on top of the two boxes must be different; if two means were equal, at least one letter must be the same.

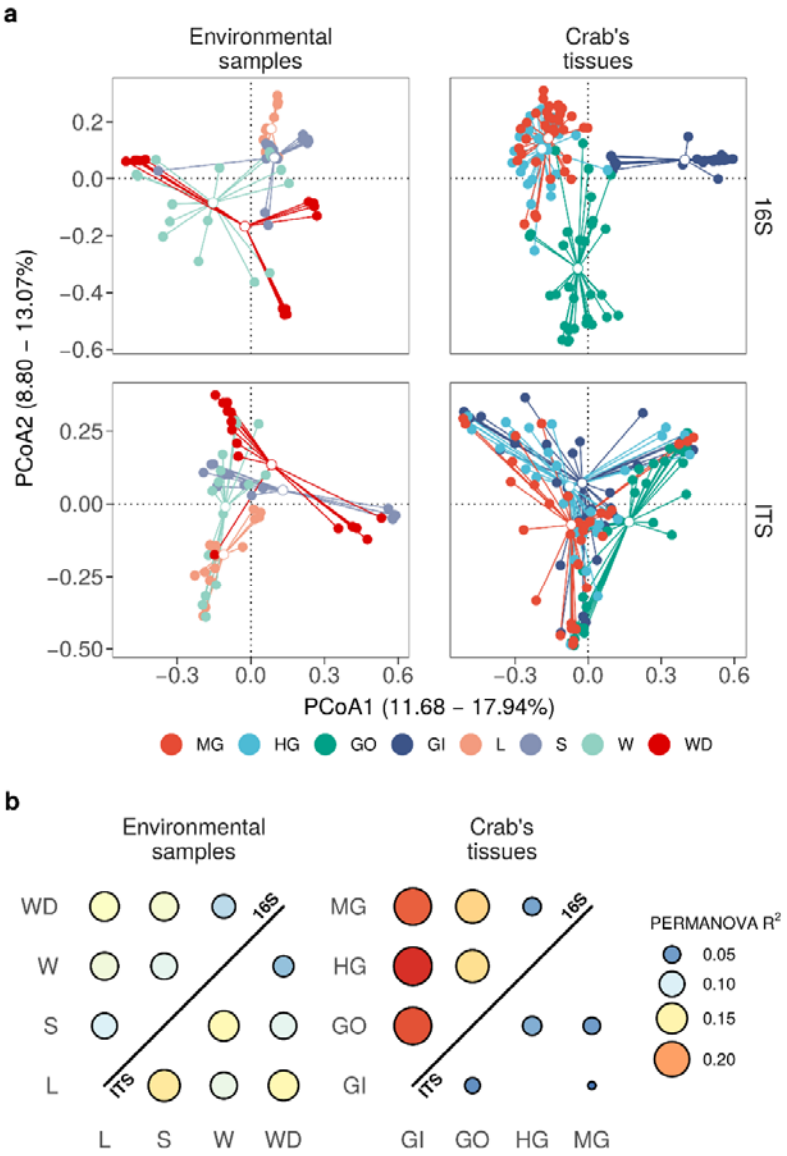


Figure 3: Microbial distribution according to Bray-Curtis distance. a) Principal coordinates analyses on Bray-Curtis distances inferred form 16S rRNA and ITS-1 metabarcoding. Environmental samples and crab's organs (top side of the panels) were reported separately as distribution obtained with the two markers reported (right side of each panel). Solid-coloured points represent different samples whereas white-filled points represent centroids. The variance of the objects along each axis has been reported between squared brackets (from the lowest to the highest) and different sample types were reported using different colours (MG, midgut; HG, hindgut; GO, gonads; GI, gills; L, litter; S, soil; W, water; WD, water debris). b) Permutational analysis of variance on ordinations reported in panel a. For each pair of organs and environmental samples a permutational analysis of variance was performed. R-squared values of significant contrasts were reported for both 16S (upper triangle) and ITS (lower triangle) counts.

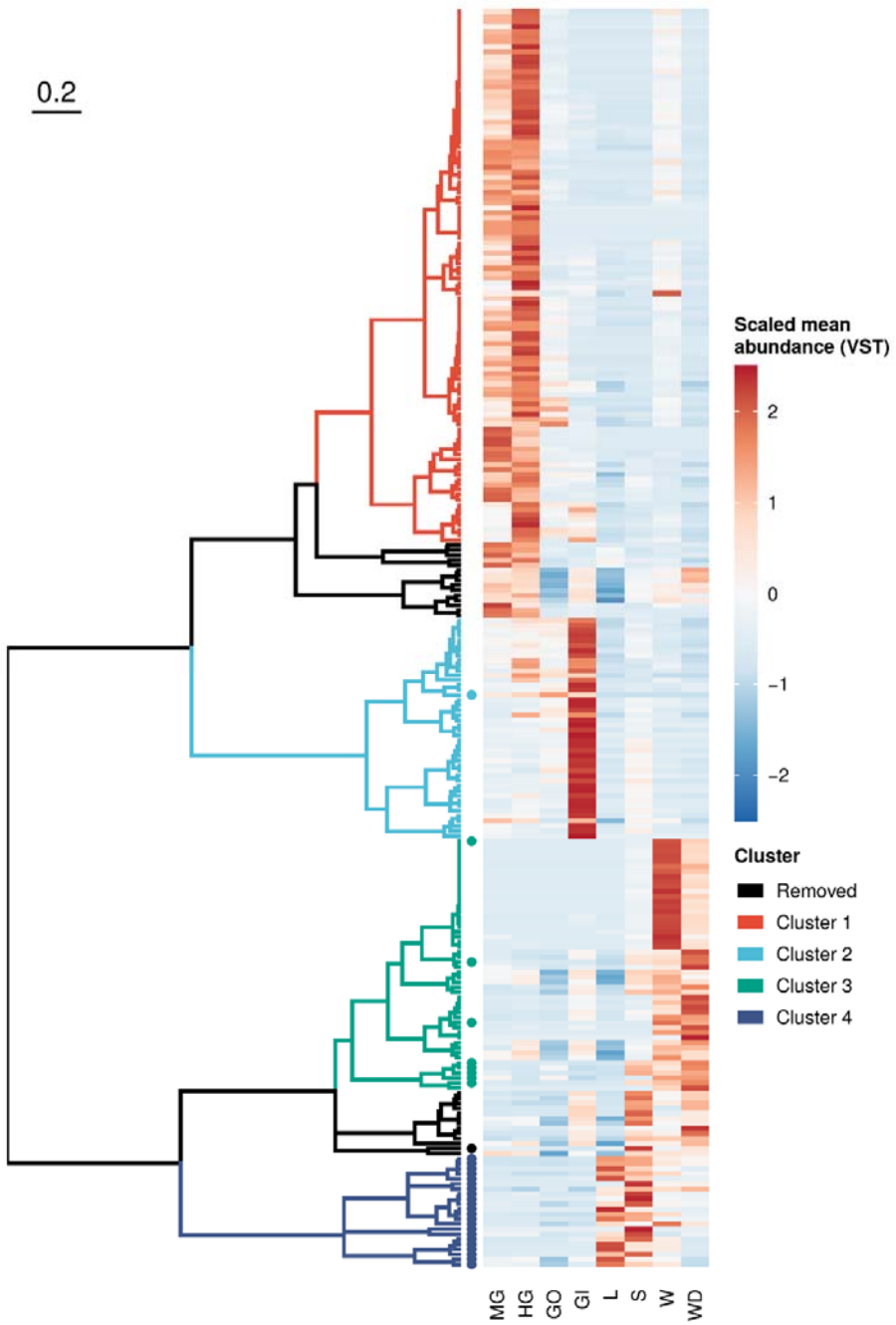


Figure 4: Sequence variant clustering according to their abundance along sample types. Amplicon sequence variants reporting a different abundance pattern in one or more sample types (loglikelihood ratio test of DESeq2) were clustered according to their mean variance-stabilized abundance. Abundance values were reported using different colours after clustering based on Kendall correlation (right side of the plot). Clusters were coloured according to the scheme reported in the legend whereas removed clusters (namely those composed with less than 10 variants) where reported in black. Sequence variants inferred from ITS-1 amplicon sequencing were reported using a solid dot.

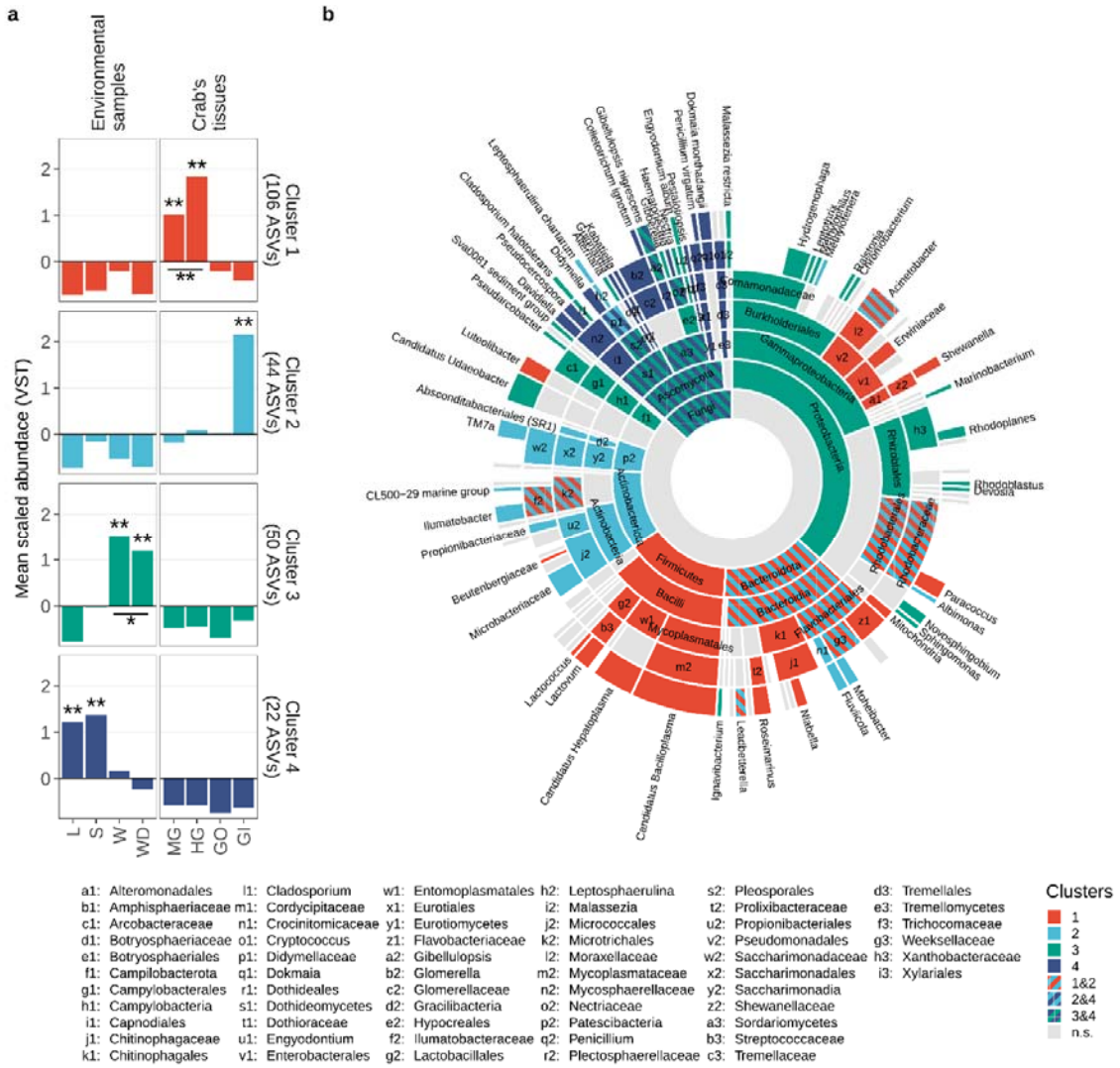


Figure 5: Taxonomic enrichment in clusters. a) Scaled variance-stabilized counts were tested for differences along sample types within all clusters detected. A pairwise Wilcoxon test was performed, and results were reported highlighting significant differences with one asterisk (p-value < 0.05) or two (p-value < 0.01). A complete overview of significant differences was reported in Figure S7. b) Sunburst plot of taxa showing a significant enrichment in a given cluster were coloured according to the colour scale showed in panel a. If a given taxonomy was significantly enriched in more than one cluster, the corresponding sector of the plot was coloured using a striped pattern (as reported in the legend). Leaves represent the most specific level at which a given variant has been classified, namely genus level for 16S rRNA amplicons (Bacteria and Archaea) and species level for ITS-1 region amplicons (Fungi).

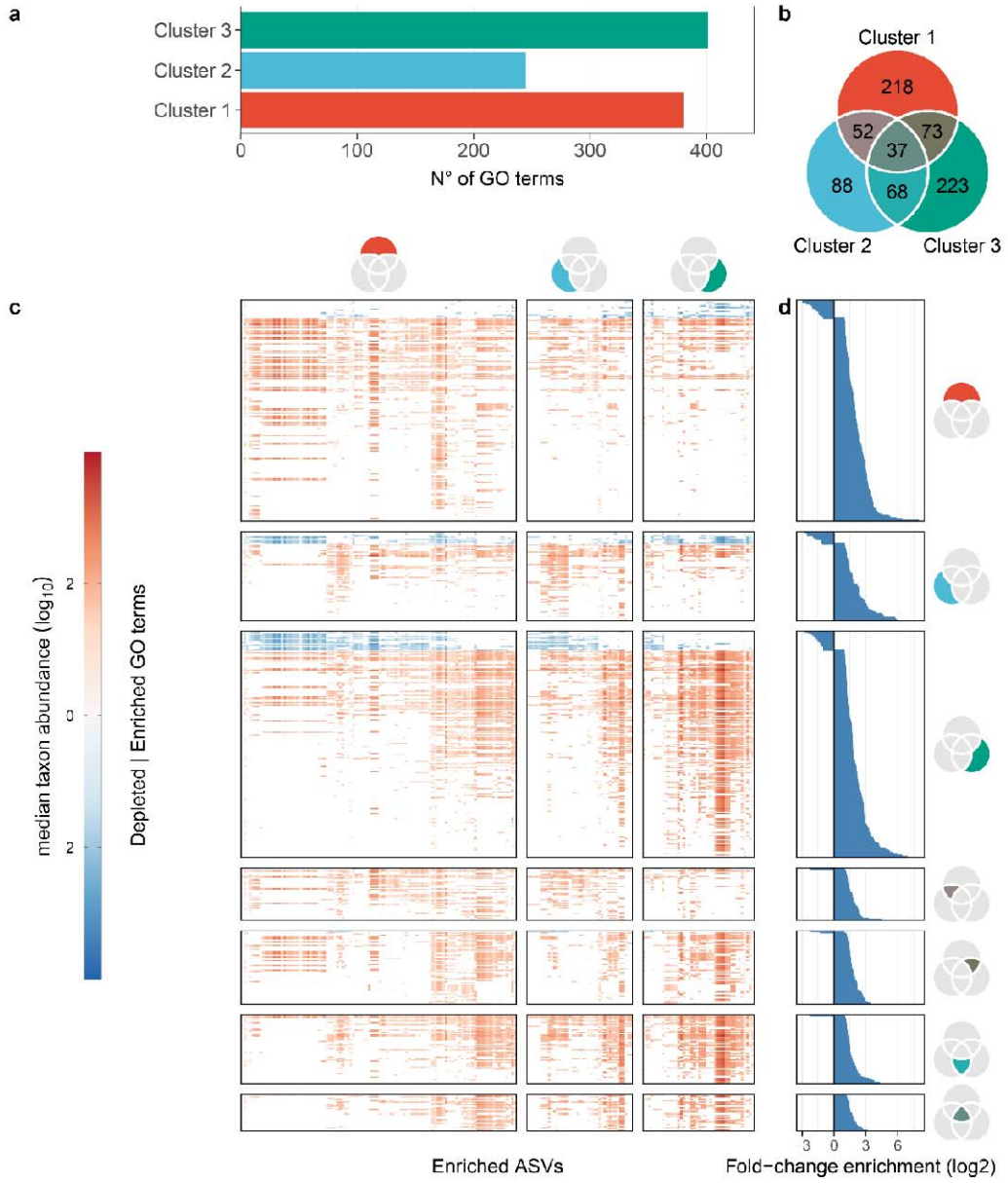


Figure 6: Enrichment analysis of "Gene Ontology" terms associated with detected functions. a) Number of GO terms associated to bacterial function detected in each cluster (cluster 4 was not reported since it was entirely composed of Fungi). b) Venn diagram of GO terms detected in the ASVs of each cluster. Diagram sections in common between two or more clusters were coloured by interpolating colours reported in panel a. c) Median taxon abundance of GO terms (y-axis) in all ASVs detected in each cluster (x-axis). The median taxon abundance was reported using different shades of red—for enriched terms, namely those detected with a higher frequency in respect to the whole population—and blue—for depleted terms, namely those detected with a lower frequency than the rest of the population. The plot was vertically divided according to ASV clusters whereas it was horizontally divided according to venn diagram sections reported in panel b. d) Mean enrichment fold changes associated to each term. Fold-changes were transformed using the logarithmic function (with base equal two) to report enriched and depleted terms symmetrically around the zero.

842

843 # Tables
844

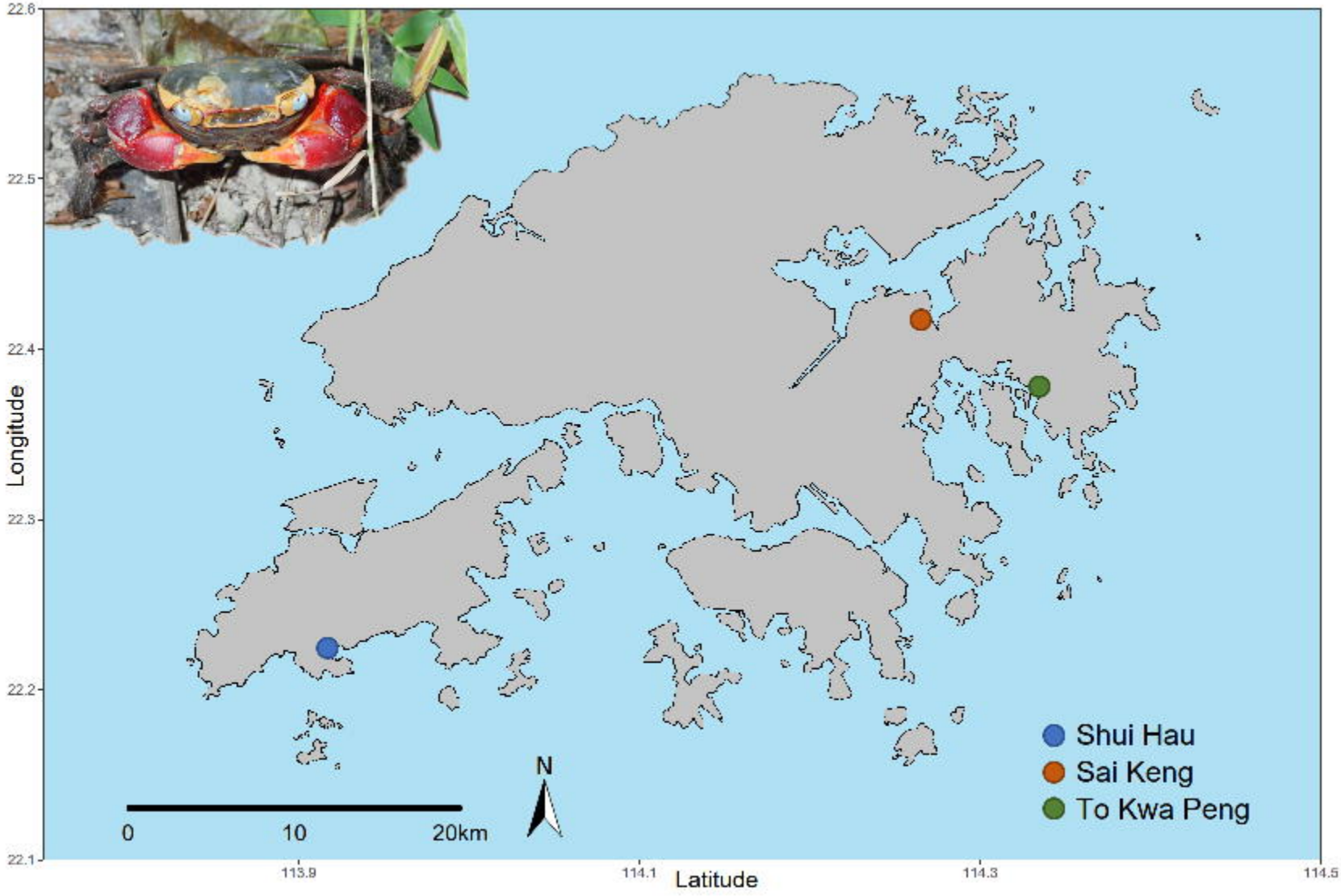
845 **Table 1: Mean differences between extrapolated diversity and observed diversity in**
846 **each sample type and site.** The differences between extrapolated Simpson diversity (Inverse
847 Simpson index computed for a sequencing depth higher than the observed one) and observed
848 diversity (Inverse Simpson index computed for a sequencing depth equal to the real
849 sequencing depth of the sample) was reported using the average value ± the standard error
850 on the mean for each sample type and site. Good's coverage estimator was also reported using
851 the same notation used for Simpson diversity. Abbreviations as in Figures 1 and 2.
852

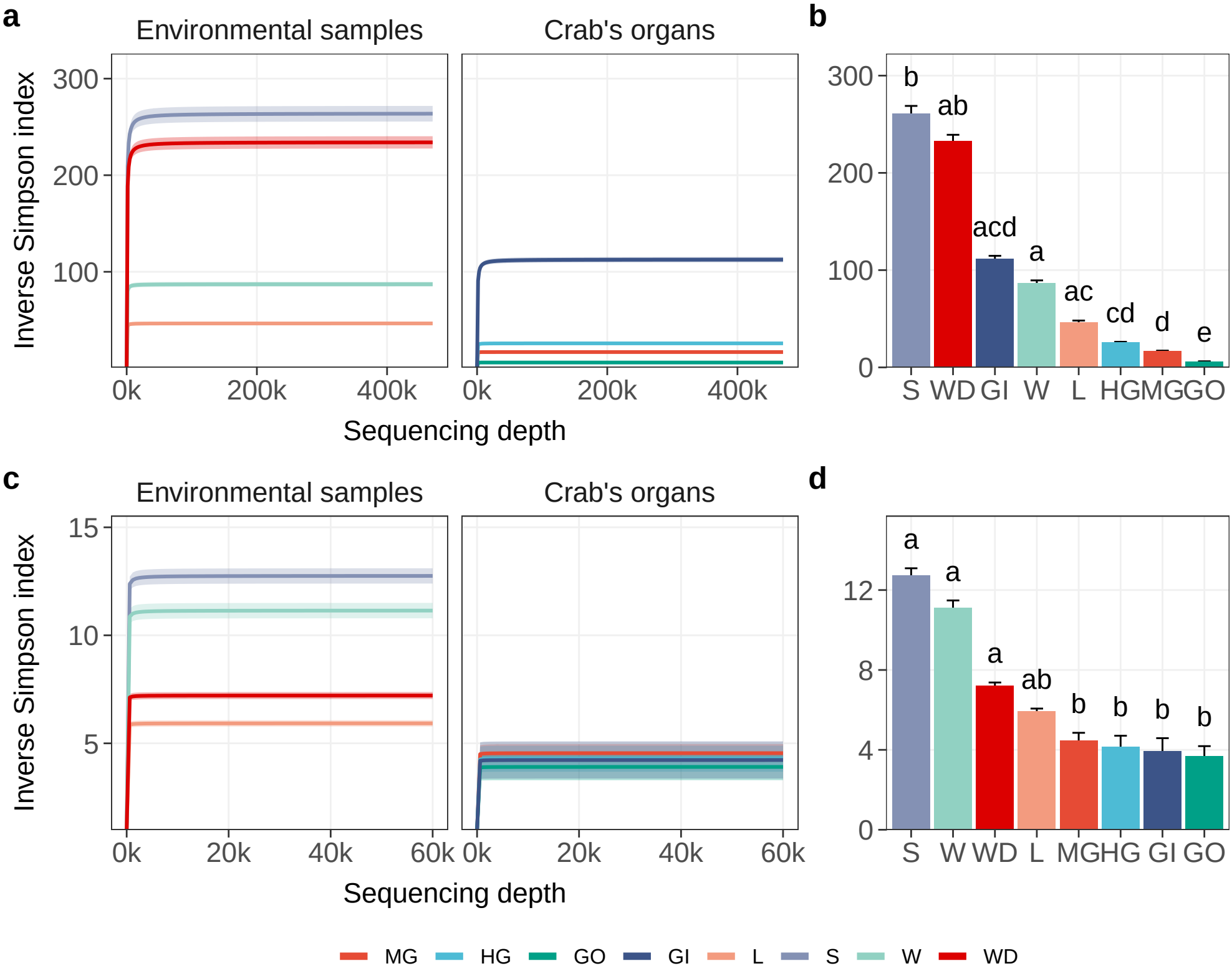
		Inverse Simpson (ext - obs)	Good's coverage estimator
Sample type			
	MG	0.051 ± 0.040	99.995 ± 0.001
	HG	0.132 ± 0.126	99.997 ± 0.001
	GO	0.085 ± 0.029	99.987 ± 0.005
	GI	0.575 ± 0.230	99.994 ± 0.002
	L	0.185 ± 0.114	99.963 ± 0.007
	S	2.741 ± 0.809	99.993 ± 0.001
	W	0.246 ± 0.066	99.997 ± 0.001
	WD	1.149 ± 0.293	99.992 ± 0.002
Site			
	SH	0.705 ± 0.236	99.987 ± 0.003
	SK	0.502 ± 0.146	99.990 ± 0.002
	TKP	0.240 ± 0.067	99.996 ± 0.001

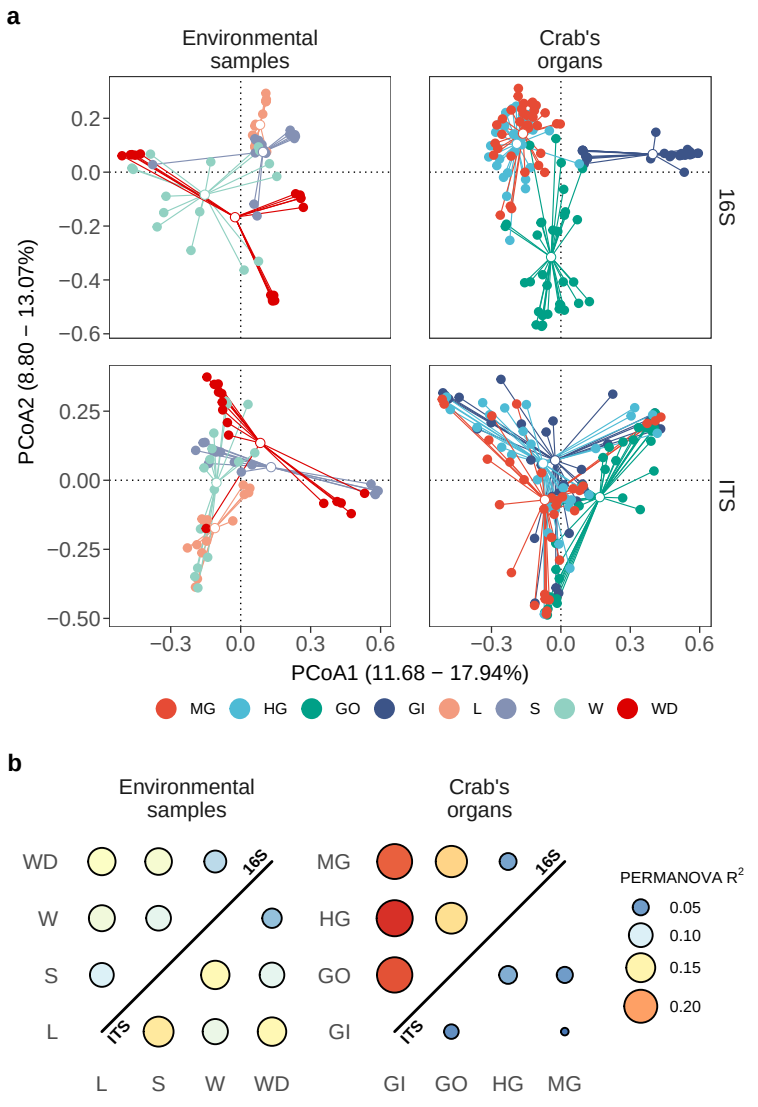
853

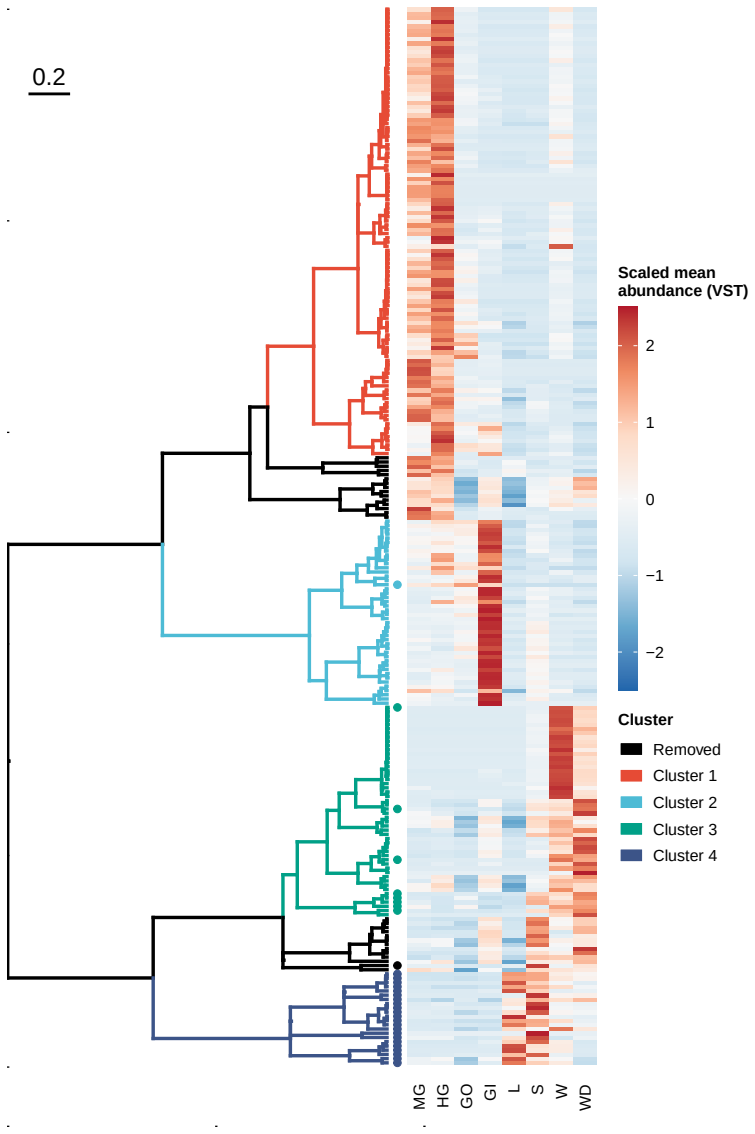
854

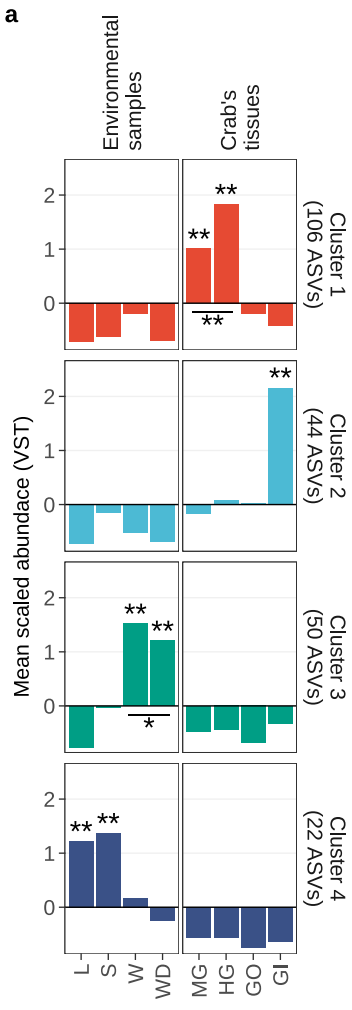
855

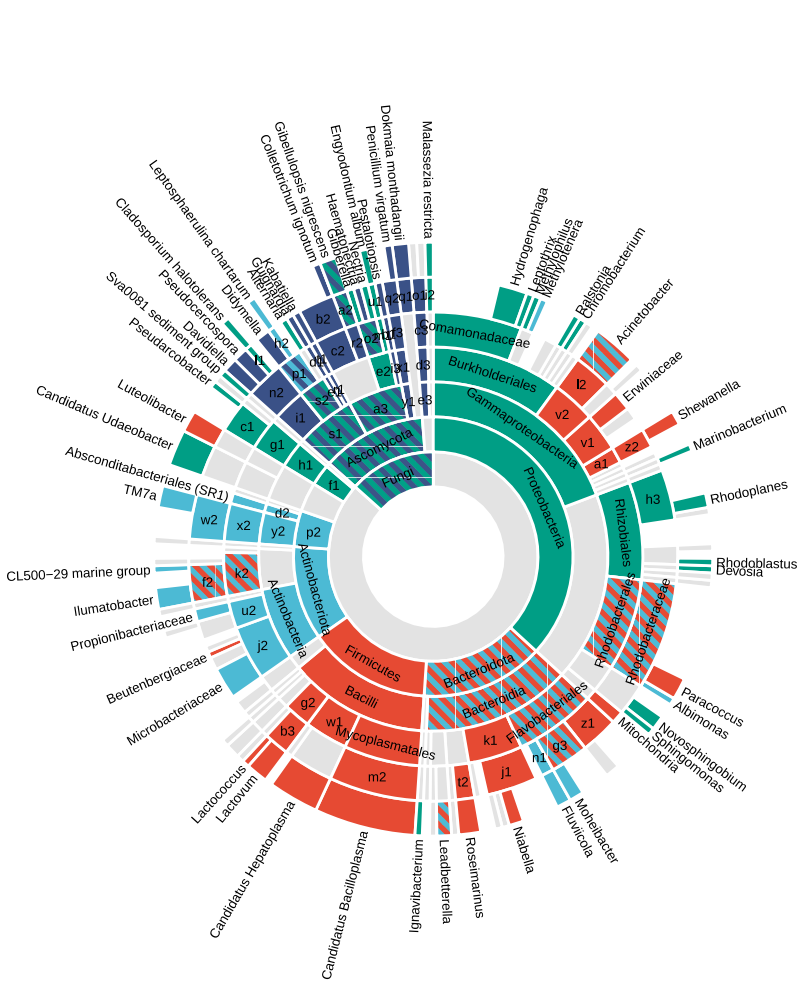

22.6
22.5
22.4
22.3
22.2
22.1
Longitude
113.9
114.1
114.3
114.5
Latitude
0
10
20km
N
Shui Hau
Sai Keng
To Kwa Peng


a
Environmental samples
Crab's organs
Inverse Simpson index
300
200
100
0k
200k
400k
Sequencing depth
b
b
ab
acd
a
ac
cd
d
e
S
WD
GI
W
L
HG
MG
GO
c
Environmental samples
Crab's organs
Inverse Simpson index
15
10
5
0k
20k
40k
60k
Sequencing depth
d
a
a
a
ab
b
b
b
b
S
W
WD
L
MG
HG
GI
GO
MG
HG
GO
GI
L
S
W
WD


a
Environmental samples
Crab's organs
16S
ITS
PCoA2 (8.80 − 13.07%)
PCoA1 (11.68 − 17.94%)
0.2
0.0
−0.2
−0.4
−0.6
0.25
0.00
−0.25
−0.50
−0.3
0.0
0.3
0.6
MG
HG
GO
GI
L
S
W
WD
b
Environmental samples
Crab's organs
WD
W
S
L
16S
ITS
MG
HG
GO
GI
L
S
W
WD
GI
GO
HG
MG
PERMANOVA R2
0.05
0.10
0.15
0.20

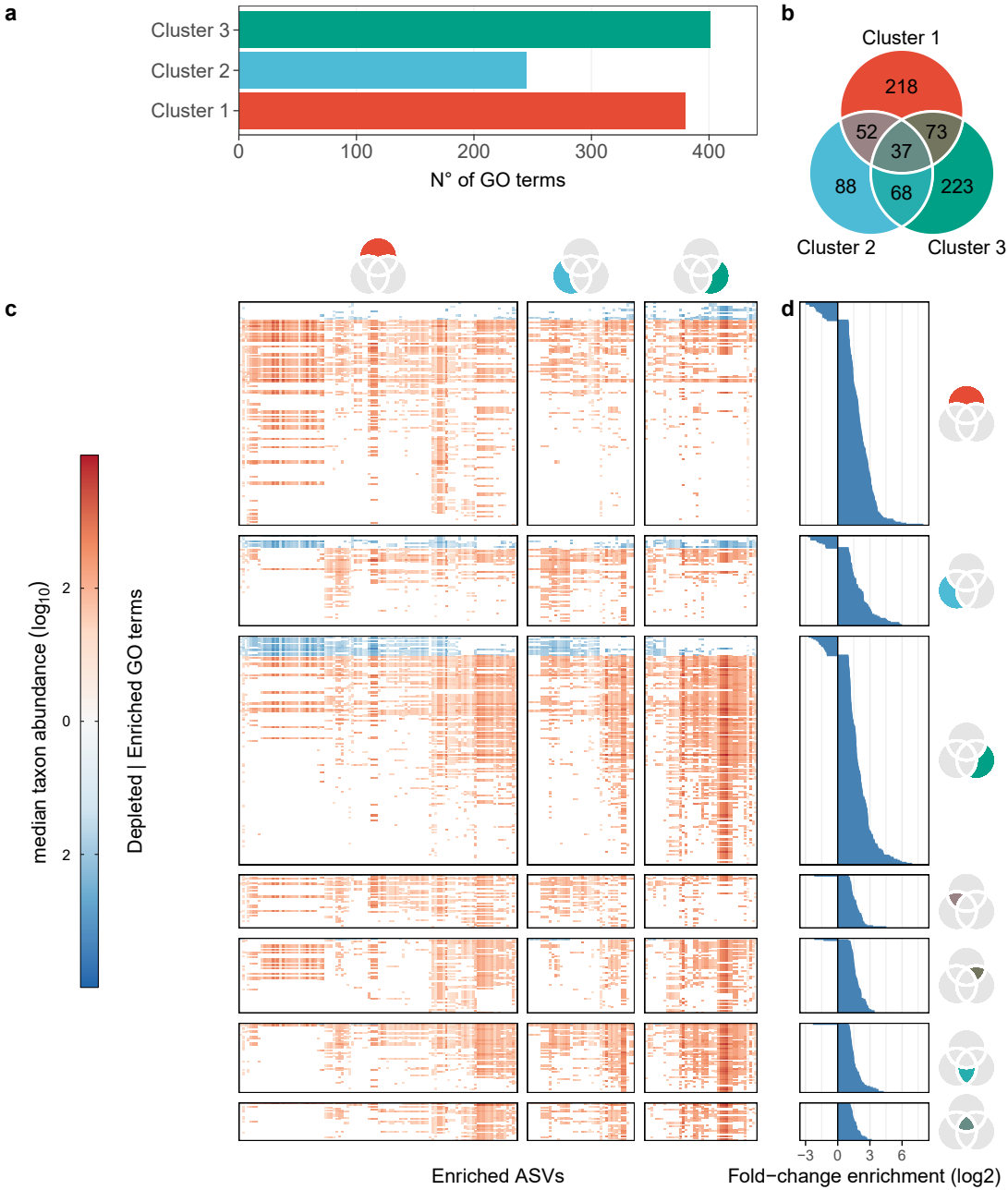

0.2
Scaled mean abundance (VST)
2
1
0
−1
−2
Cluster
Removed
Cluster 1
Cluster 2
Cluster 3
Cluster 4
MG
HG
GO
GI
L
S
W
WD





a1: Alteromonadales
b1: Amphisphaeriaceae
c1: Arcobacteraceae
d1: Botryosphaeriaceae
e1: Botryosphaeriales
f1: Campilobacterota
g1: Campylobacterales
h1: Campylobacteria
i1: Capnodiales
j1: Chitinophagaceae
k1: Chitinophagales
l1: Cladosporium
m1: Cordycipitaceae
n1: Crocinitomicaceae
o1: Cryptococcus
p1: Didymellaceae
q1: Dokmaia
r1: Dothideales
s1: Dothideomycetes
t1: Dothioraceae
u1: Engyodontium
v1: Enterobacterales
w1: Entomoplasmatales
x1: Eurotiales
y1: Eurotiomycetes
z1: Flavobacteriaceae
a2: Gibellulopsis
b2: Glomerella
c2: Glomerellaceae
d2: Gracilibacteria
e2: Hypocreales
f2: Ilumatobacteraceae
g2: Lactobacillales
h2: Leptosphaerulina
i2: Malassezia
j2: Micrococcales
k2: Microtrichales
l2: Moraxellaceae
m2: Mycoplasmataceae
n2: Mycosphaerellaceae
o2: Nectriaceae
p2: Patescibacteria
q2: Penicillium
r2: Plectosphaerellaceae
s2: Pleosporales
t2: Prolixibacteraceae
u2: Propionibacteriales
v2: Pseudomonadales
w2: Saccharimonadaceae
x2: Saccharimonadales
y2: Saccharimonadia
z2: Shewanellaceae
a3: Sordariomycetes
b3: Streptococcaceae
c3: Tremellaceae
d3: Tremellales
e3: Tremellomycetes
f3: Trichocomaceae
g3: Weeksellaceae
h3: Xanthobacteraceae
i3: Xylariales

Clusters
1
2
3
4
1&2
2&4
3&4
n.s.


a
Cluster 3
Cluster 2
Cluster 1
0
100
200
300
400
N° of GO terms
b
Cluster 1
218
52
73
37
88
68
223
Cluster 2
Cluster 3
c
median taxon abundance (log10)
2
0
2
Depleted | Enriched GO terms
Enriched ASVs
d
−3
0
3
6
Fold−change enrichment (log2)