Organ-specific microbiota enhances the terrestrial lifestyle of a brachyuran crab

Authors: Giovanni Bacci[1*], Sara Fratini[1*], Niccolò Meriggi[1*], Christine L. Y. Cheng[2], Ka Hei Ng[2], 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

* These authors contributed equally to this work

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

Abstract

The transition to terrestrial environments has occurred repeatedly and at different geological times in arthropods, but almost no information is available about the role of symbiotic microbiota in such process. Here we investigated the associated microbiota of a terrestrial brachyuran crab, *Chiromantes haematocheir*, using a targeted metagenomic approach. Bacterial 16S rRNA gene and fungal ITS sequences were obtained from selected crab organs and environmental matrices to profile microbial communities. We found stable and organ-specific communities of microorganisms associated to the gut and the gills of the crabs, the former involved in the digestion of vascular plant tissues. These communities were mainly composed by prokaryotic organisms and significantly differed from the fungi-dominated ones present in the environment. Our results suggest that the establishment of a specific, stable microbiota may be crucial to drive evolutionary transitions, as colonization of terrestrial environments.

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*, *9*, *18*, *10–17*). 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 (*10*, *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,

50 where they are commonly known pathogens of terrestrial arthropods, including
51 semiterrestrial mangrove crabs (*21*, *22*).

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

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

75 and possibly the evolution of adaptive symbioses able to contribute to the baseline life
76 processes of this semiterrestrial crab and to drive its adaptations to the terrestrial habitat it
77 inhabits. Thus, our final aim was to shed light on the possible role of animal-microbiome
78 symbioses involved in the evolution of adaptive traits coupled with the ongoing transition of
79 crabs from the sea to terrestrial environments.

80 # Results

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

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

94 Single nucleotide clustering detected a total of 56,233 amplicon sequence variants (ASVs)
95 including both 16S rRNA gene and ITS1 amplicons. The 16S rRNA gene amplicons produced
96 55,036 ASVs (97.87%), whereas the ITS1 region was clustered into 1,197 ASVs (2.13%). Most
97 variants were assigned to the Bacteria domain (54,800 ASVs, 97.45% of the total ASVs
98 detected) and Fungi kingdom, even if a sporadic presence of Archaea was found (236 ASVs,

99 0.42% of the total ASVS detected). In general, microbial diversity was higher in the
100 environmental samples than in crab's organs, except for the gills where we found a high
101 diversity of 16S rRNA amplicons (Figure 2a and b, Table S3). The fungal diversity was low in
102 all crab's organs, whereas environmental samples were characterised by a highly diverse
103 fungal assemblage, with soil, water, and water debris reporting the highest diversities (Figure
104 2c & d). For each sample category and at each site, the difference between extrapolated
105 diversity—namely the inverse Simpson index calculated by simulating a higher sequencing
106 depth than the observed depth for each sample—and observed diversity was really low
107 (Table 1), meaning that our sampling effort was effective in assessing the natural microbial
108 diversity for each category. Even the Good's coverage estimator was higher than 99.9% in all
109 sample types highlighting that only 1‰ of clustered sequences came from ASVs detected only
110 once in the whole dataset (Table 1 and Table S3). Microbial diversity was rather uniform
111 across sampling sites, except for To Kwa Peng (TKP), which showed a significantly lower
112 diversity with respect to the other sites in terms of ITS1 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 (S) and litter (L), and
120 water (W) and water debris (WD), respectively, formed two distinct clusters, although
121 differences in microbial communities found across the three Hong Kong sites increased the
122 intra-cluster variability (Figure S4). On the other hand, the same 16S rRNA amplicon dataset
123 showed a clear separation across crab's organs, with a very limited influence of the collection

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

131 Both environmental samples and crab's organs possessed a large set of microbial species
132 which were unique to the category considered (Figure S5). As already mentioned, sampling
133 sites also contributed to shape microbial communities, but they predominantly affected
134 environmental samples distribution and showed a low effect on crab's organs (Figure S4a and
135 b, Figure S5a and c). In addition, the community composition of both prokaryotic and fungal
136 communities resulted to be more variable and dispersed at site level than at organ level
137 (Figures S4 and S6, Table S5), showing a higher specificity of such assemblages within the
138 different organs and across sites. The microbiota found on the gills appeared to be less organ-
139 specific and shared a consistent portion of their microbiomes with soil and water, as well as
140 gut samples (Figure S5b and d).

141 Defining characteristic patterns across sample types

142 To inspect microbial distribution in different sample categories, we performed log-likelihood
143 ratio test on both prokaryotic and fungal communities using DESeq2. We found 250 ASVs
144 (218 bacteria and 32 fungi) reporting a different distribution across crab organs and/or
145 environmental samples (Figure 4 and Table S6). Even if significant ASVs corresponded to
146 0.45% of the ASVs profiled in the whole community (250 on 55819), they accounted for more
147 than 50% of the total microbial abundance (with a mean in each sample of 56.5% and a
148 standard error of 2.10%) reflecting the presence of many rare and sporadic species

149 throughout sample types and sites. Divisive analysis of hierarchical clustering obtained using
150 variance-stabilized counts produced four distinct clusters, which show a peculiar pattern of
151 abundance of ASVs clearly linked to the ecological and biological settings of the study. Indeed,
152 clusters 1 and 2 (Figures 4 and S7) were composed by microbial assemblages highly
153 represented in the crab's organs, while the other two clusters, (namely, cluster 3 and 4)
154 represented the microorganisms mostly found in the environmental samples (Figures 4 and
155 S7). The clusters representing the microorganisms more abundant in the crab's organs were
156 represented by bacterial ASVs only and were split into two groups, formed by the bacteria of
157 the gut (cluster 1) and of the gills (cluster 2), respectively, with the latter being the smaller
158 one.

159 Taxonomic and functional enrichments in crab's organs and environmental 160 samples

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

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

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

187 Functional profiling obtained from 16S rRNA gene amplicons confirmed that the conserved
188 bacterial structures associated with crab's organs are not just composed by a defined set of
189 taxa but also by a defined set of molecular functions (Figure 6a). Clusters of bacterial variants
190 were mainly enriched/depleted by a unique set of molecular functions that did not report the
191 same effect in ASVs from other clusters. The inferred genomic content of variants detected in
192 crab's gut (Cluster 1) was enriched by 359 GO terms and depleted by 21, with more than a half
193 (218 terms, 18 depleted and 200 enriched, Table S8) peculiar only to this cluster. Molecular
194 functions, as cellullases and xylanases, commonly associated to commensal microorganisms
195 present in the gut of model herbivorous arthropods, were not enriched in the gut microbiome
196 of the crab (Table S9). The gills (Cluster 2) had a population of bacteria enriched by 231 GO
197 terms and depleted by 14 terms, with roughly one third (88 terms, 12 depleted and 76
198 enriched) significantly found only in these organs. Functions associated to biofilm formation

199 such as cellulose synthase (*30*, *31*) were enriched only in this cluster (GO:0016760), whereas
200 functions associated to the nitrogen cycle - such as nitric- and nitrous-oxide reductase activity
201 - were shared between the gills and the environmental cluster 3. The latter - mainly composed
202 of ASVs mostly found in water and water debris samples - had the largest set of GO terms (401
203 terms, 22 depleted and 379 enriched) with more than a half (223 terms, 19 depleted and 204
204 enriched) exclusively present in the inferred genome of ASVs highly abundant in that cluster.
205 Only 4.87% of the total GO terms was significantly enriched in all clusters (37 terms out of
206 759, with no terms significantly depleted), indicating a functional role of microbial
207 communities both in crab's organs and in environmental samples (Figure 6b and c, and Table
208 S8). Amplicon sequence variants significantly enriched in crab's gut and gills shared more
209 functions (GO terms) with the environmental cluster 3 (water and water debris) than
210 between themselves. In particular, 73 and 68 functions (with only 2 and 1 depleted terms)
211 were shared between the gut and the gills, respectively, and the ASVs detected in the
212 environmental cluster 3 (Figure 6b, Table S8). Considering the total number of terms
213 significantly enriched/depleted, the gills were the organs more impacted by the environment
214 with roughly one third (27.8%) of molecular functions shared with cluster 3. In contrast, the
215 more populated organ, the gut, shared only 19.2% of the total number of functions.

216 Discussion

217 Our results show a dramatic difference between the microbiome of the semi-terrestrial
218 brachyuran crab *C. haematocheir*, which represents one of the most recent attempts of land
219 colonisation by an arthropod, and the surrounding environmental microbiome. This
220 difference is more pronounced for the bacterial component than for the fungal one. Further,
221 the crab microbiome analysis shows the presence of consistent organ-specific microbial
222 communities. The degree of specificity of the taxonomic and functional microbial signatures

223 we found at organ level can be explained by the adaptive evolution of symbiotic host-microbe
224 associations, which now represent a single biological unit under selection from the recently
225 colonised terrestrial habitat, i.e., a holobiont (*sensu 32*). Temporary host-microbiota
226 associations can be created, retained or lost over time, due to the large number of organisms-
227 microbe interactions. Thus, a crucial verification needed to define the holobiont as a single
228 selection unit is to define the degree of specificity of the microbial signature of an organism.
229 Our sampling design and analytical approach were able to define stable, organ-specific
230 microbiota in both the gut and gills of individuals belonging to different populations,
231 suggesting a tight and persistent association between those prokaryotic communities and *C.*
232 *haematocheir*. The organ specificity is supported by the observation that such microbiota
233 were consistently dissimilar among the different organs but their composition did not
234 significantly change across sampling sites, where local microbial assemblages proved to be
235 site specific and different from the ones associated with the crabs. The most enriched
236 microorganisms in the gut microbiota of our model species are shared with terrestrial
237 isopods, which moved to the land roughly 300 million years ago, and are supposed to play a
238 central role in their adaptation to a diet based on vascular plant tissues (*33*). One of the
239 general assumptions of the hologenome concept, however, is that both the host genome and
240 the genetic information of the associated microbiome are transferred from one generation to
241 the next. This postulation underlies the critical role of mothers in transferring
242 microorganisms to their offspring (*34*), a challenging process for our model organism that has
243 highly dispersive planktonic larval stages followed by a recruitment phase. Another crucial
244 result is that fungal communities are exclusive to the environmental matrices while almost
245 absent in *C. haematocheir* organs. This suggests a specific selection process that favours
246 bacteria versus fungi, possibly supported by selection against yeasts and fungi since several
247 fungi are known pathogens for arthropods and crabs (*21*, *35*). To the best of our knowledge,

248 the present integrated set of results support the novel idea that semi-terrestrial crabs, only
249 recently migrated onto the land, are associated with microbial communities that play specific
250 functional roles in response to the dramatic environmental pressures posed by the sea-to-
251 land transition. The next questions are where these microbes were acquired and where are
252 their 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*, *16*). In contrast with the intestinal bacterial communities found in the
269 freshwater Chinese mitten crab, *Eriocheir sinensis* (*40*), and notwithstanding the different
270 functions, the selected intestinal tracts of *C. haematocheir* were homogeneous in terms of
271 microbial communities, suggesting, at least for this terrestrial crab species, a similar role of
272 such traits.

273 Host-associated microbiota are known to play a crucial role in the digestive process of many
274 terrestrial arthropods (*41*). 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 integrates 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 (*16*, *42*,
280 *43*). The termites' ability to degrade lignocellulose is classical example of high gut microbiome
281 specialization (*41*). 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*). 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* (*44*). 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*, *45*). *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 (*45*, *46*)., Other bacterial genera found in our
295 intestinal cluster (such as *Niabella*, *Paracoccus* and *Shewanella*) were also isolated from
296 terrestrial isopods (*42*). 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* (*47*)and do not include marine species, although they are
302 occasionally described as pathogens of aquacultured species (*48*, *49*). 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 (*50*). In addition, the intestinal cluster was enriched in *Erwiniaceae*, a family that
308 includes insect's symbionts such as *Buchnera aphidicola* (*51*). 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 (*17*, *52*). Second, *C. haematocheir*, as well as the most land-adapted
317 family of brachyuran crabs, the Gecarcinidae, still retain an indirect development strategy and
318 release planktonic larvae in coastal waters. Consequently, *C. haematoicher*, as well as the land
319 crabs with planktonic larvae in general, may have developed intimate gut-specific symbiotic
320 associations only through pseudo-vertical transmission (sensu *53*), i.e., by acquiring and
321 selecting their gut microbiota directly from food or, most probably, through consumption of
322 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 (*54*), or in areas where they can chemically detect adult populations, such as the case of the
326 swamp ghost crab *Ucides cordatus* (*55*). 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 (*56–58*). 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 (*59*) and crustaceans (*60*). 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* (*61*). 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 (*60*). 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, as a result 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 (*62*). 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 (*63*) 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 (*64*). Moreover, prokaryotes-
384 fungi direct competition through bacterial production of antifungal molecules may limit the
385 growth of the latter. Events of fungal exclusion by direct competition mediated by symbiotic
386 bacteria have been extensively described in beetles (*65*). *Streptomyces* sp. SPB74, symbiont of
387 the beetle *Dendroctonus frontalis*, produces an active compound, mycangimycin, which
388 specifically inhibits antagonist fungi without acting on the mutual ones (*65*). This protective
389 role may be extended to the offspring. Antifungal molecules produced by the symbiotic
390 bacterium *Burkholderia gladioli* are known to protect the eggs of the beetle *Lagria villosa*
391 against pathogenic fungi (*35*, *66*). Pathogenic fungi have been reported in intertidal and
392 terrestrial crabs, such as the case of the swamp ghost crab, *U. cordatus*, infected by species of
393 black yeast, *Exophiala cancerae*, and *Fonsecaea brasiliensis*, which cause a condition called
394 "lethargic crab disease" (*21*, *22*). The role of prokaryotic associations in the defence against
395 pathogenic fungi can be of critical importance for terrestrial crabs, since the immune
396 molecules of crustaceans are less efficient than the ones of insect against fungal infections
397 (*67*).

398 In conclusion, this is the first attempt to ascertain the role of host-microbiome associations in
399 the transition to land of brachyuran crabs, the most recent arthropod taxa to perform such an
400 evolutionary leap. Using a targeted metagenomic approach, we demonstrated persistent
401 organ-specific crab-microbiome associations which help a semi-terrestrial crab coping with
402 novel environmental challenges posed by its recently occupied habitats, i.e., the lowland
403 forests or East Asia. We found clear differentiations among the microbiota associated with
404 different crab's organs and the ones found in the environmental matrices, suggesting a
405 selective acquisition of the microbiota through "gain" and "loss" mechanisms. The differences
406 found among the organ-related and environmental clusters were not merely taxonomical
407 since the different clusters harbour different metabolic profiles. These results corroborate the
408 hypothesis that the recorded differences can be due to the presence of metabolic
409 complementation mechanisms that took place in those organs mostly impacted by the
410 challenges posed by the terrestrial environment. Our data also show possible evolutionary
411 convergences towards a uniform 'terrestrial intestinal microbiota' across different lineages of
412 arthropods, suggesting the presence of bacterial associations linked to terrestrial life. The
413 present study strongly supports the hypothesis that semi- and terrestrial crabs are an
414 appropriate model system to study the evolution of arthropod-microbe interaction under the
415 selective pressures posed by the sea-land transition, which is happing right now in Brachyura
416 (*23*).

417

418 # Material and methods

419 ## Study species

420 *Chiromantes haematocheir* (Decapoda; Brachyura; Sesarmidae) is a semi-terrestrial crab
421 colonising the coastal vegetated areas from Taiwan to South East Asia (*28*, *68*). In Hong Kong,
422 it forms large populations in areas of lowland secondary forest adjacent to mangroves and in
423 pockets of riverine forests, where it was also observed climbing trees. Very little is known
424 about its ecology, apart from the fact that it digs deep burrows and, as in many sesarmids, it
425 releases pelagic larvae into the ocean. This species was selected because it is the most land
426 adapted among the Hong Kong brachyuran crabs.

427 ## Sample collection and total DNA extraction

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

443 stored at -20°C in RNAlater (Thermo Fisher Scientific) stabilization solution until DNA
444 extraction.

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

453 **16S (V3-V4) rRNA gene amplification and sequencing**

454 The preparation and sequencing of the 16S library were performed at Laboratory of
455 Advanced Genomics, Department of Biology, University of Florence (Firenze, Italy). PCR
456 amplifications of the bacterial V3-V4 16S rRNA gene fragments were performed using KAPA
457 HiFi HotStart ReadyMix (Roche) and the primer pair 341F (5'-CCTACGGGNGGCWGCAG-3) and
458 805R (5'-GACTACNVGGGTWTCTAATCC-3') (*69*) with overhang Illumina adapters. 16S
459 amplicon PCR protocol was set on 25 µl of final volume. In detail for each reaction, 12.5 µl of
460 2× KAPA HiFi HotStart ReadyMix (Roche), 10 µl of 1µM forward and reverse primers and
461 2.5 µl of template DNA (5–20 ng/µl) were combined (PCR cycling conditions are reported in
462 Supplementary materials). Amplicon PCR reaction was performed using the GeneAmp PCR
463 System 2700 (Thermo Fisher Scientific) and the following cycling conditions: denaturation
464 step at 95 °C for 3 minutes; 35 (gonads, water and water debris) and 25 (hindgut, midgut,
465 gills, soil and litter) cycles: at 95 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds;
466 final extension step at 72 °C for 5 minutes. All PCR products were checked through
467 electrophoresis on 1.5% agarose gel and then purified using KAPA Pure Beads (Roche)

468 following the manufacturer's instructions. To apply the Illumina adapters sequencing
469 indexing using Nextera XT Index Kit V2 (Illumina), a second PCR amplification was then
470 performed by preparing a reaction mix in accordance with the Illumina 16S metagenomic
471 library preparation protocol (*70*). An indexing step was made for all samples by seven PCR
472 cycles. Amplicon products from indexing PCR were purified using KAPA Pure Beads (Roche)
473 and their quality check was performed using Agilent 2100 Bioanalyzer (Agilent Technologies)
474 with Agilent DNA 1000 Kit (Agilent Technologies). Subsequently, concentration check was
475 performed by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Finally, the barcoded
476 libraries were balanced and pooled at equimolar concentration, before being sequenced on an
477 Illumina MiSeq (PE300) platform (MiSeq Control Software 2.6.2.1).

478 **ITS1 rDNA region sequencing**

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

493 second PCR was performed to apply the Illumina sequencing adapters Nextera XT Index
494 Primer (Illumina). An indexing step was made for all samples by seven PCR cycles. After
495 Indexing PCR amplicon libraries were purified using the CleanNGS kit (CleanNA, the
496 Netherlands), and the quality control was performed on a Typestation 2200 platform (Agilent
497 Technologies, Santa Clara, CA, USA). Afterwards all barcoded libraries were mixed at
498 equimolar concentration, quantified by qPCR Kapa Library quantification kit (Roche) and
499 sequenced on an Illumina MiSeq (PE300) platform (MiSeq Control Software 2.5.0.5 and Real-
500 Time Analysis software 1.18.54.0).

501 **Amplicon sequence variant inference**

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

518 (*76*). All variants not classified as Bacteria, Archaea or Fungi were removed together with
519 sequences classified as chloroplasts or mitochondria (16S rRNA sequences only). Additional
520 information on the sequence variant inference pipeline used were reported in Supplementary
521 Materials.

522 **Inferring functional content of amplicon variants**

523 The genome content of bacterial ASVs was inferred using PICRUSt2 pipeline (*77*). Enzyme
524 Commission Numbers (EC numbers) were converted into Gene Ontology terms (GO terms)
525 using the mapping file available at: http://www.geneontology.org/external2go/ec2go. Gene
526 abundance was retrieved using the "--strtified" option to report gene abundances at species
527 level (ASVs. Additional information about functional content inference was reported in
528 Supplementary Material.

529 **Statistical analyses**

530 All statistical analyses were performed in the R environment (version 3.6). Briefly, alpha- and
531 beta-diversity analyses were conducted using the vegan package version 2.5 (*78*) in
532 combination with the iNEXT package version 2.0 (*79*). Normalization and differential
533 abundance analyses were performed with DESeq2 version 1.28 (*80*) whereas enrichment
534 analysis of taxa and functions was performed using hypergeometric test ("phyper" function of
535 R stats package). For additional details about tests data manipulation see Supplementary
536 Materials.

537 **References**

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

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

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

544 4. J. Lozano-Fernandez, R. Carton, A. R. Tanner, M. N. Puttick, M. Blaxter, J. Vinther, J.
545 Olesen, G. Giribet, G. D. Edgecombe, D. Pisani, A molecular palaeobiological exploration
546 of arthropod terrestrialization. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150133 (2016).

547 5. F. Giomi, M. Fusi, A. Barausse, B. Mostert, H.-O. Pörtner, S. Cannicci, Improved heat
548 tolerance in air drives the recurrent evolution of air-breathing. *Proc. R. Soc. B Biol. Sci.*
549 **281**, 20132927 (2014).

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

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

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

558 9. C. A. Farrelly, P. Greenaway, Land crabs with smooth lungs: Grapsidae, Gecarcinidae,
559 and Sundathelphusidae ultrastructure and vasculature. *J. Morphol.* **215**, 245–260
560 (1993).

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

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

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

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

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

573 15. R. Simoni, F. Giomi, D. Spigoli, H.-O. Pörtner, S. Cannicci, Adaptations to semi-terrestrial
574 life in embryos of East African mangrove crabs: a comparative approach. *Mar. Biol.* **160**,
575 2483–2492 (2013).

576 16. E. S. Lindquist, K. W. Krauss, P. T. Green, D. J. O'Dowd, P. M. Sherman, T. J. Smith III, Land
577 crabs as key drivers in tropical coastal forest recruitment. *Biol. Rev.* **84**, 203–223
578 (2009).

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

581 18. S. C. Faria, R. O. Faleiros, F. A. Brayner, L. C. Alves, A. Bianchini, C. Romero, R. C.
582 Buranelli, F. L. Mantelatto, J. C. McNamara, Macroevolution of thermal tolerance in
583 intertidal crabs from Neotropical provinces: A phylogenetic comparative evaluation of

584 critical limits. *Ecol. Evol.* **7**, 3167–3176 (2017).

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

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

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

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

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

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

600 25. I. Zilber-Rosenberg, E. Rosenberg, Role of microorganisms in the evolution of animals
601 and plants: The hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723–735
602 (2008).

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

605 27. E. Rosenberg, O. Koren, L. Reshef, R. Efrony, I. Zilber-Rosenberg, The role of
606 microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* **5**, 355–362
607 (2007).

608 28. K. F. Abd El-Wakeil, Trophic structure of macro- and meso-invertebrates in Japanese
609 coniferous forest: Carbon and nitrogen stable isotopes analyses. *Biochem. Syst. Ecol.* **37**,
610 317–324 (2009).

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

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

617 31. H.-C. Flemming, J. Wingender, The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–633
618 (2010).

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

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

34. E. Rosenberg, I. Zilber-Rosenberg, The hologenome concept of evolution: do mothers matter most? *BJOG An Int. J. Obstet. Gynaecol.* **127**, 129–137 (2020).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

50. 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).

51. 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.

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

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

54. 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).

55. 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).

56. 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).

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

58. 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).

59. 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).

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

61. 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).

62. 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).

63. A. Vázquez-Torres, A. J. Bäumler, Nitrate, nitrite and nitric oxide reductases: from the

703 last universal common ancestor to modern bacterial pathogens. *Curr. Opin. Microbiol.*
704 **29**, 1–8 (2016).

705 64. A. Ortiz-Urquiza, N. Keyhani, Action on the Surface: Entomopathogenic Fungi versus the
706 Insect Cuticle. *Insects.* **4**, 357–374 (2013).

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

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

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

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

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

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

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

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

730 73. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes,
731 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods.*
732 **13**, 581–583 (2016).

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

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

738 76. V. Deshpande, Q. Wang, P. Greenfield, M. Charleston, A. Porras-Alfaro, C. R. Kuske, J. R.
739 Cole, D. J. Midgley, N. Tran-Dinh, Fungal identification using a Bayesian classifier and the
740 Warcup training set of internal transcribed spacer sequences. *Mycologia.* **108**, 1–5
741 (2016).

742 77. G. M. Douglas, V. J. Maffei, J. R. Zaneveld, S. N. Yurgel, J. R. Brown, C. M. Taylor, C.

743 Huttenhower, M. G. I. Langille, PICRUSt2 for prediction of metagenome functions. *Nat.*
744 *Biotechnol.* **38**, 685–688 (2020).

745 78. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R.
746 B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, *Vegan:*
747 *community ecology package* (2019; http://cran.r-project.org/package=vegan).

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

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

752

753 ## Acknowledgments

754 We thank the friends of the Integrate Mangrove Ecology laboratory of the Swire Institute of
755 Marine Science, HKU for their help in sampling and lab work. We thank Massimo Pindo
756 (Research and Innovation Centre, Fondazione Edmund Mach) for the support in the
757 metagenomic library preparation. We also thank Prof Colin Little and Gray A. Williams for
758 inspiring and discussing the original idea behind the study.

759

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

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

768 **Data and materials availability**: data needed to evaluate the conclusions are present in the
769 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

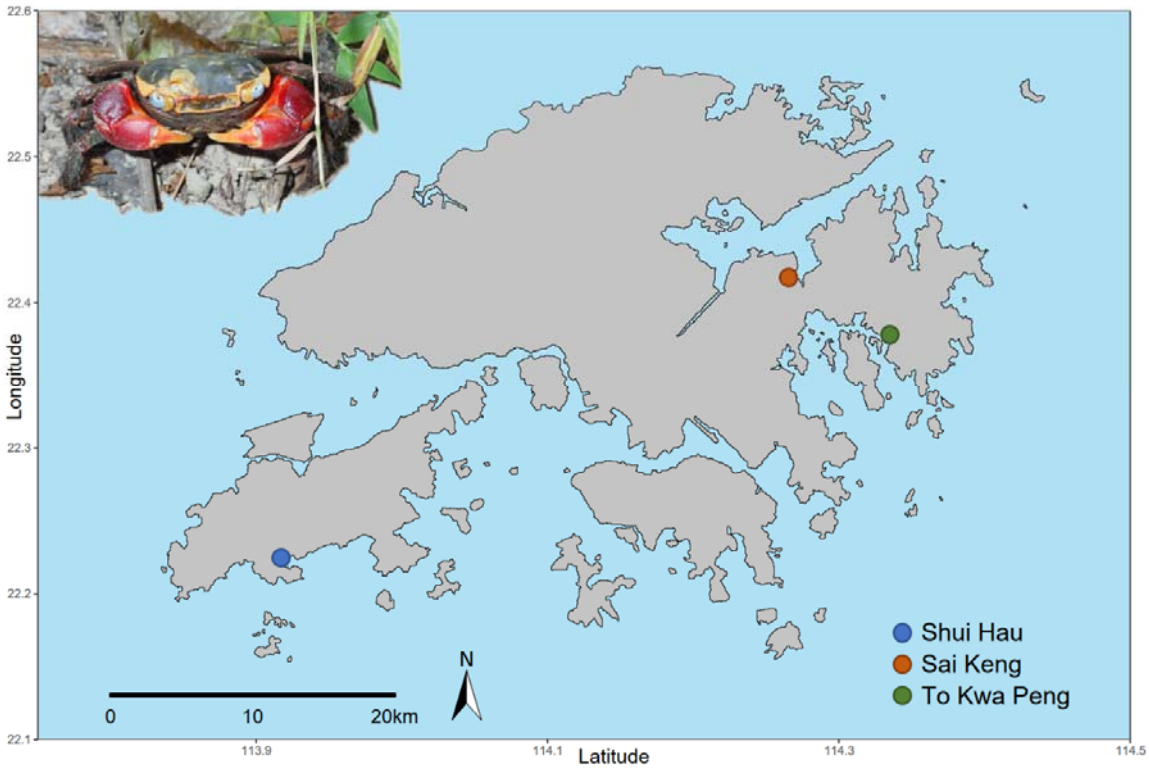


Figure 1. Geographical distribution of the sampled populations of *C. haematocheir*. The map shows the names and location of the three sampling sites visited for the study in the New Territories and on Lantau Island and an adult male *C. haematocheir* in its natural environment (top left).

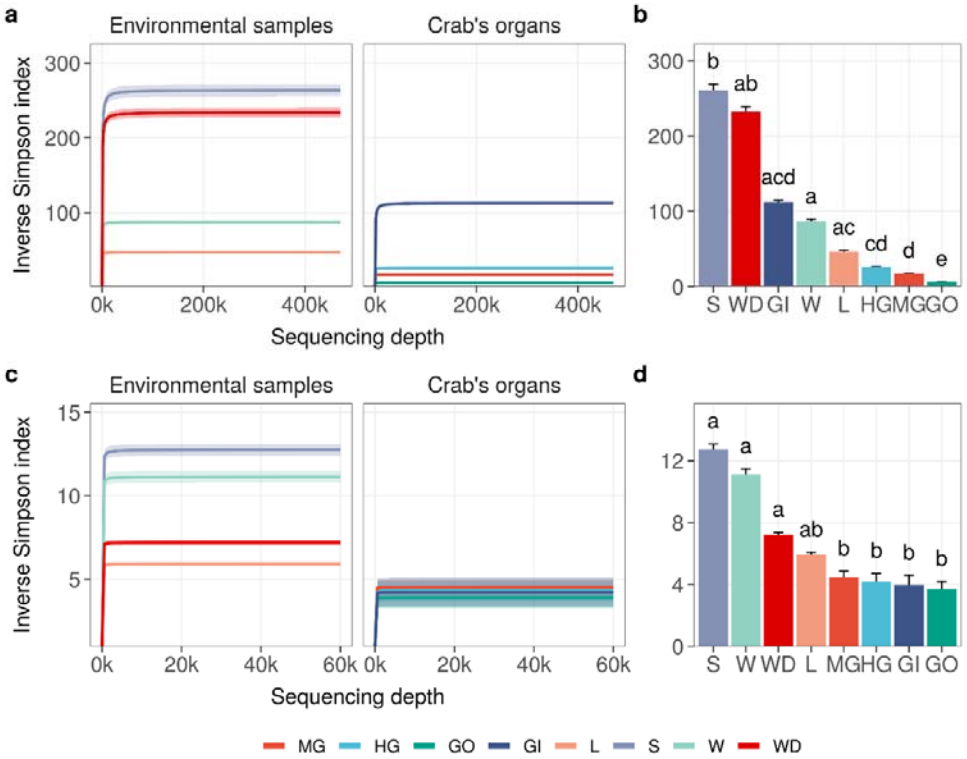


Figure 2: Microbial diversity in crab's organs and environmental samples. The average of the inverse Simpson index was reported with increasing sampling effort for all types of samples. Interpolated and extrapolated diversity was reported in panel a and c (16S rRNA gene and ITS-1 region, respectively), 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.

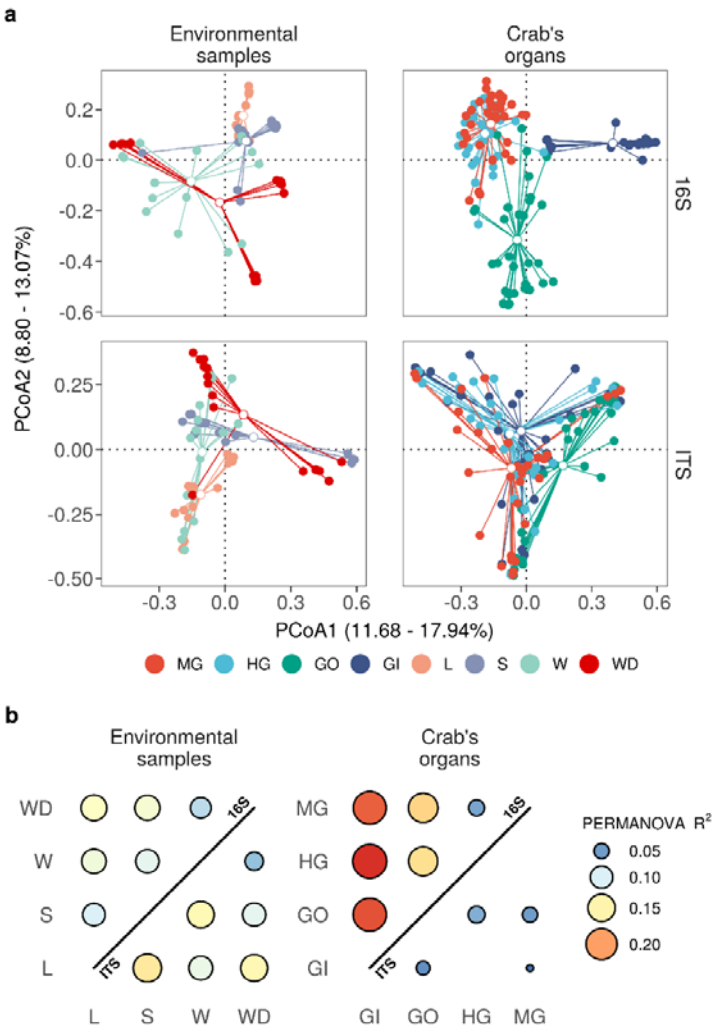


Figure 3: Microbial distribution according to Bray-Curtis distance. a) Principal coordinates analyses on Bray-Curtis distances inferred form 16S rRNA and ITS-1 metabarcoding. Environmental samples and crab's organs (top side of the panels) were reported separately as distribution obtained with the two 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.

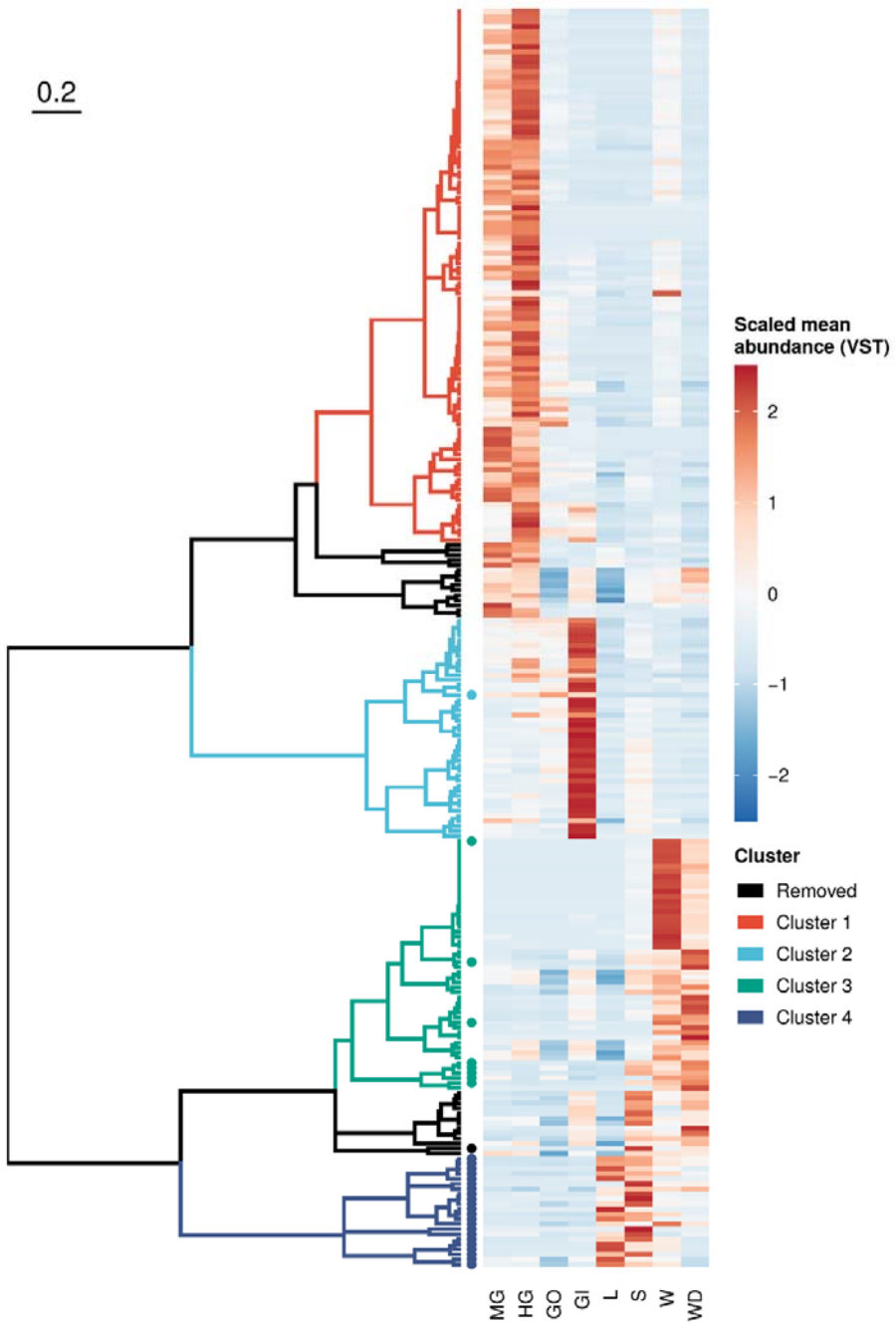


Figure 4: Sequence variant clustering according to their abundance along sample types. Amplicon sequence variants reporting a different abundance pattern in one or more sample types (loglikelihood ratio test of DESeq2) were clustered according to their mean variance-stabilized abundance. Abundance values 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.

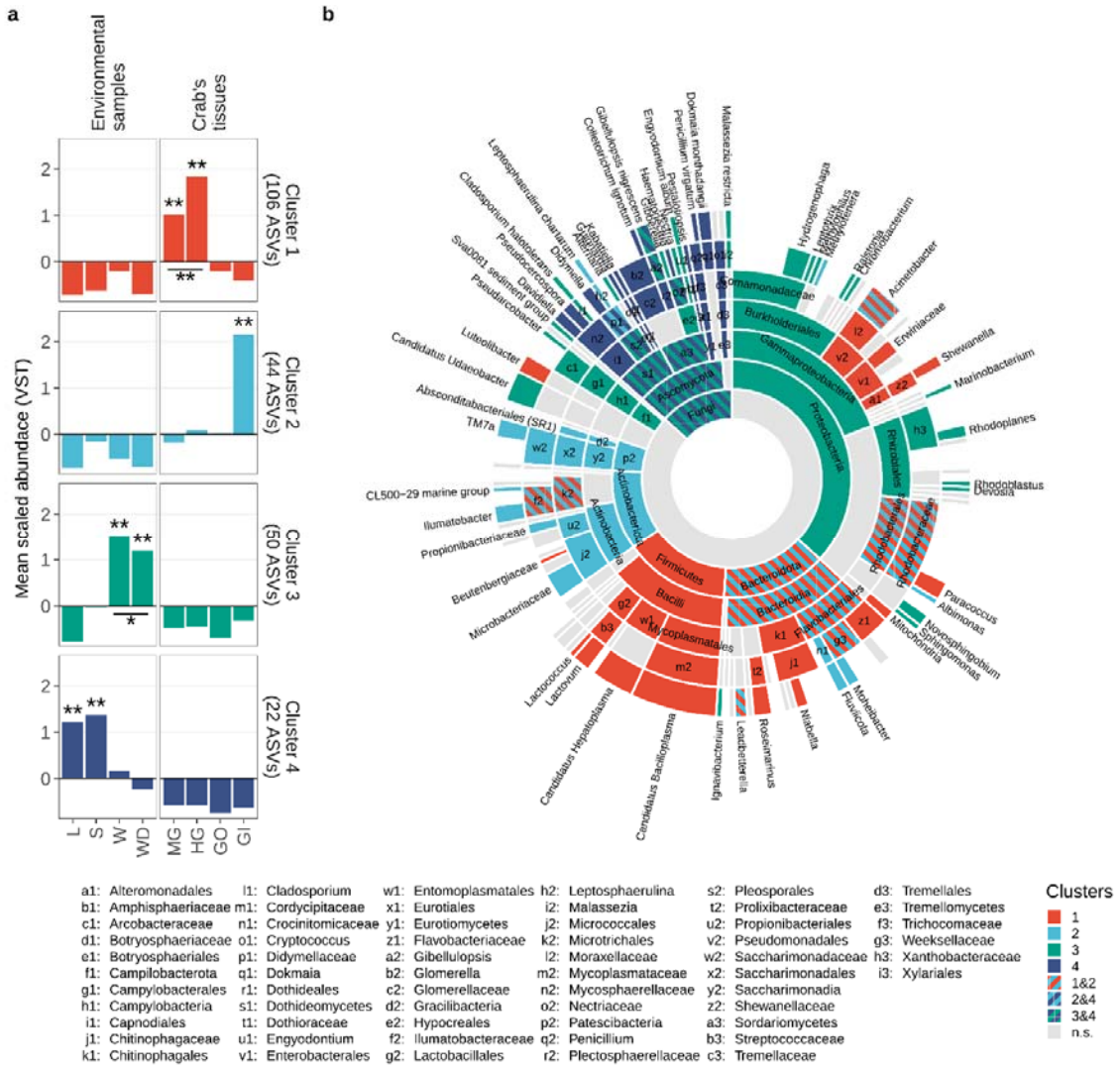


Figure 5: Taxonomic enrichment in clusters. a) Scaled variance-stabilized counts were tested for differences along sample types within all clusters detected. A pairwise Wilcoxon test was performed, and results were reported highlighting significant differences with one asterisk (p-value < 0.05) or two (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).

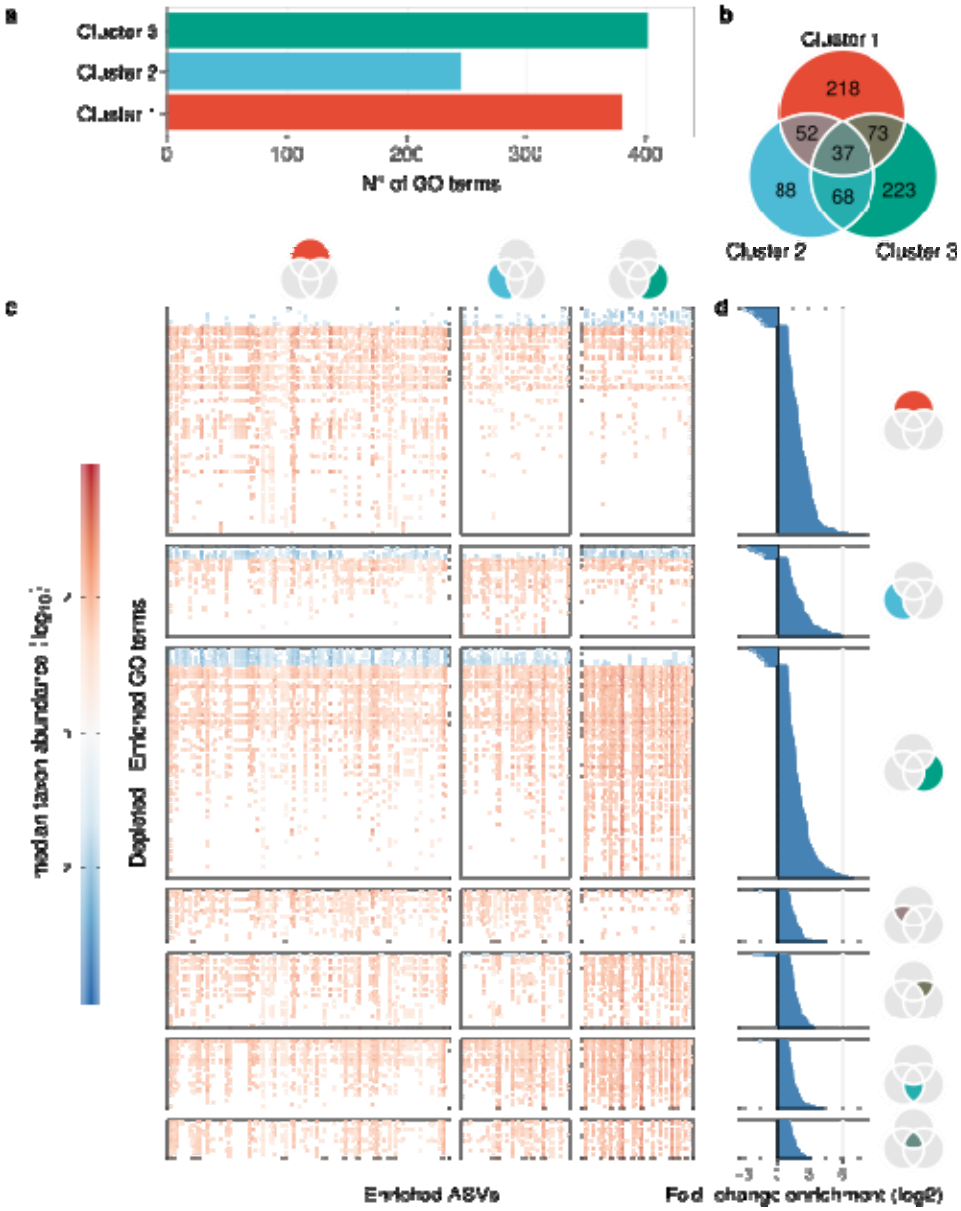


Figure 6: Enrichment analysis of "Gene Ontology" terms associated with detected functions. a) Number of GO terms associated to bacterial function detected in each cluster (cluster 4 was not reported since it was entirely composed of Fungi). b) Venn diagram of GO terms detected in the ASVs of each cluster. 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.

852 # Tables
853

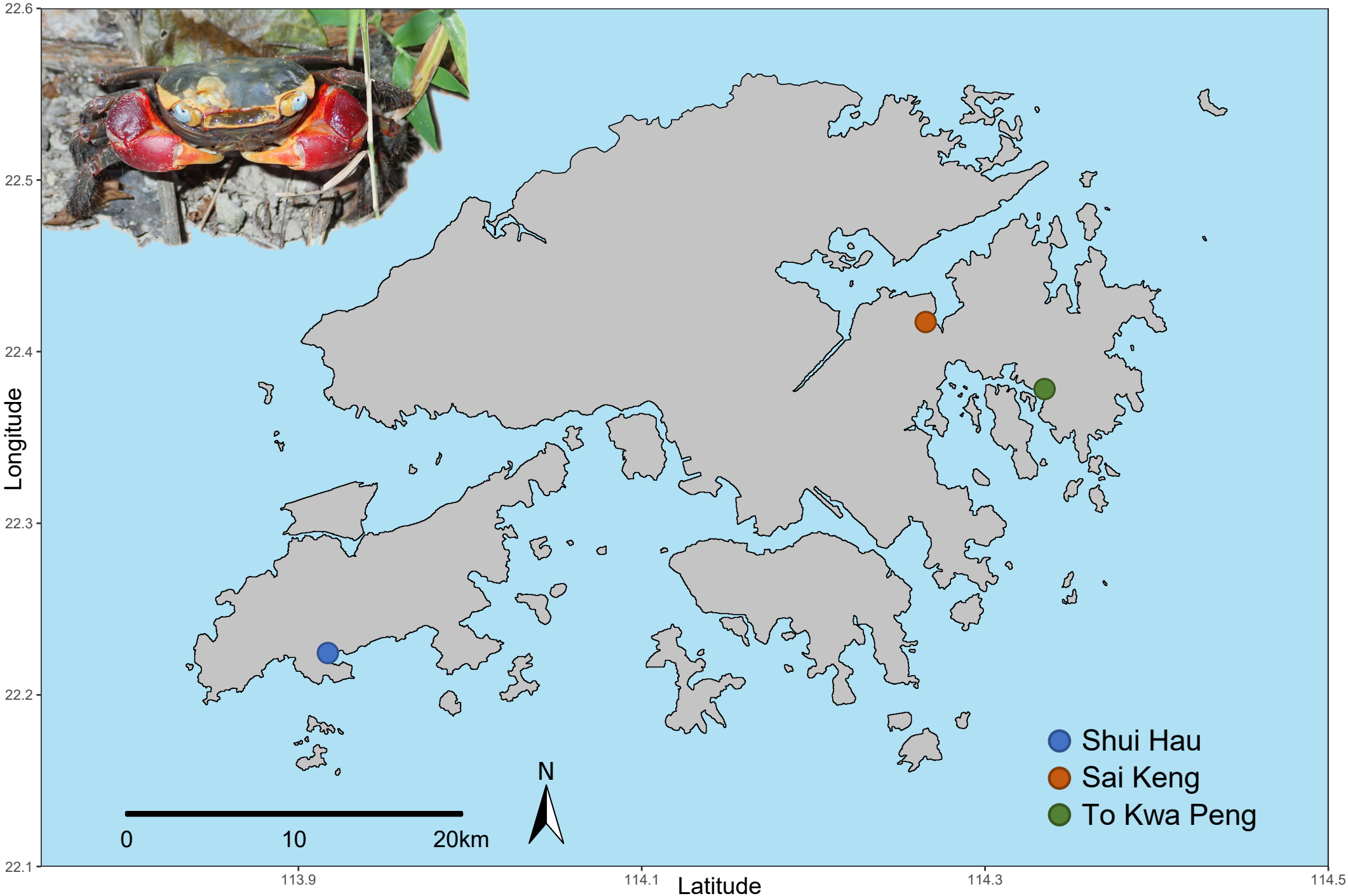
854 **Table 1: Mean differences between extrapolated diversity and observed diversity in**
855 **each sample type and site.** The differences between extrapolated Simpson diversity (Inverse
856 Simpson index computed for a sequencing depth higher than the observed one) and observed
857 diversity (Inverse Simpson index computed for a sequencing depth equal to the real
858 sequencing depth of the sample) was reported using the average value ± the standard error
859 on the mean for each sample type and site. Good's coverage estimator was also reported using
860 the same notation used for Simpson diversity. Abbreviations as in Figures 1 and 2.
861

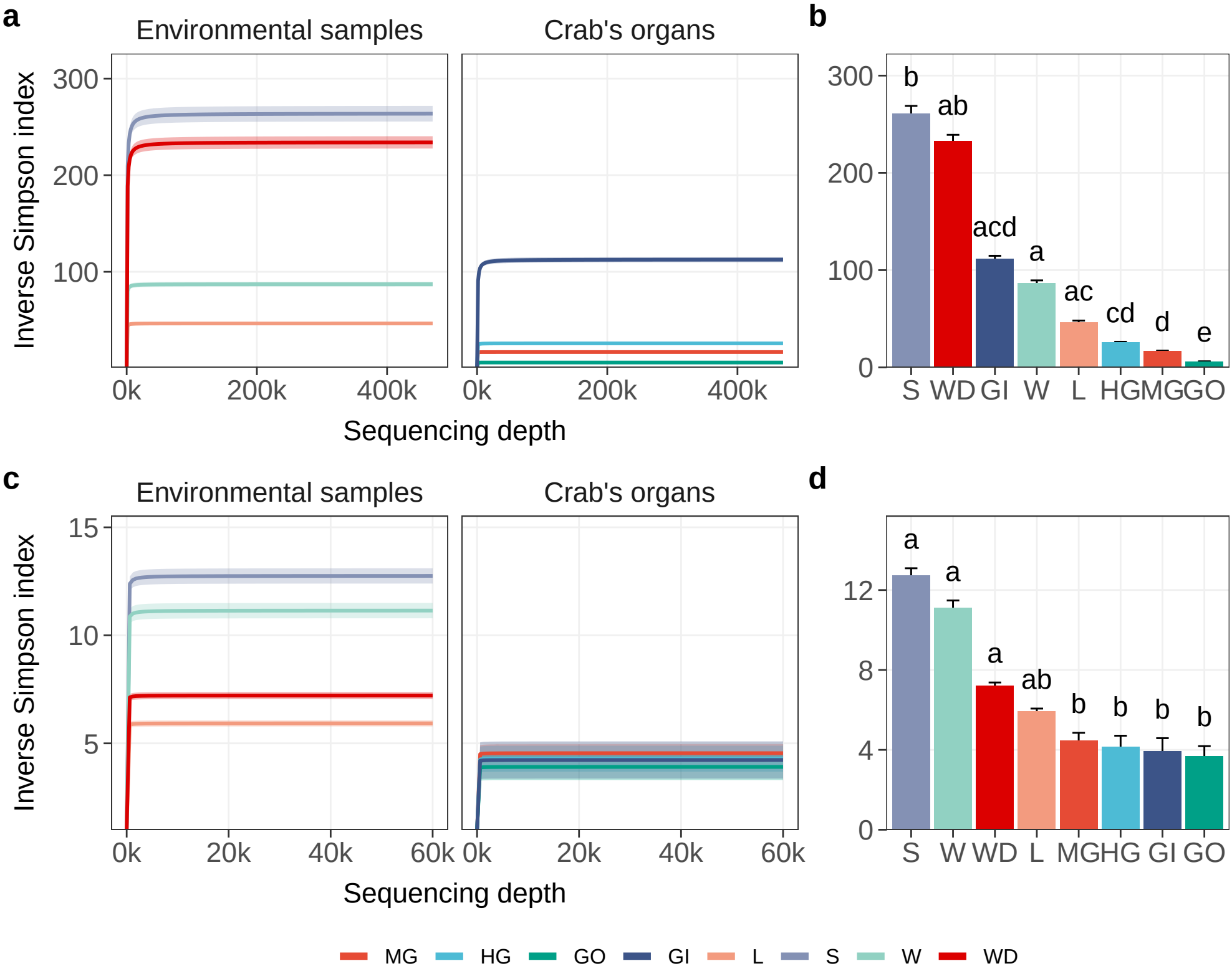
		Inverse Simpson (ext - obs)	Good's coverage estimator
Sample type			
	MG	0.051 ± 0.040	99.995 ± 0.001
	HG	0.132 ± 0.126	99.997 ± 0.001
	GO	0.085 ± 0.029	99.987 ± 0.005
	GI	0.575 ± 0.230	99.994 ± 0.002
	L	0.185 ± 0.114	99.963 ± 0.007
	S	2.741 ± 0.809	99.993 ± 0.001
	W	0.246 ± 0.066	99.997 ± 0.001
	WD	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

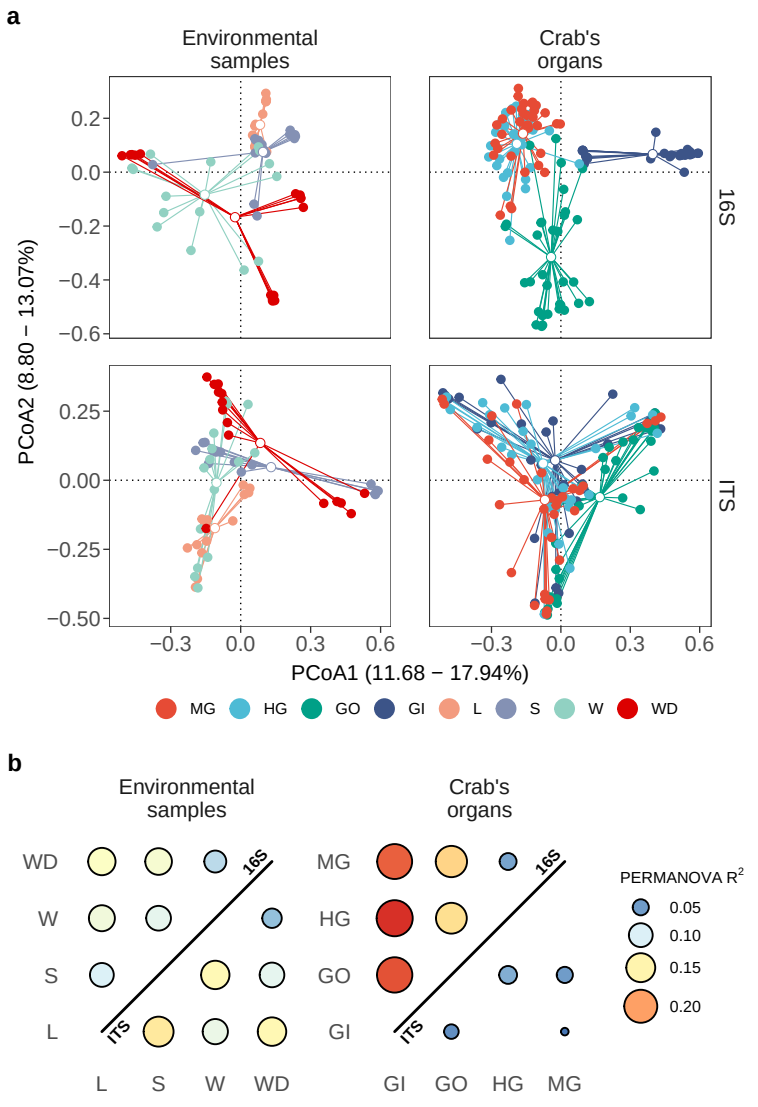
862

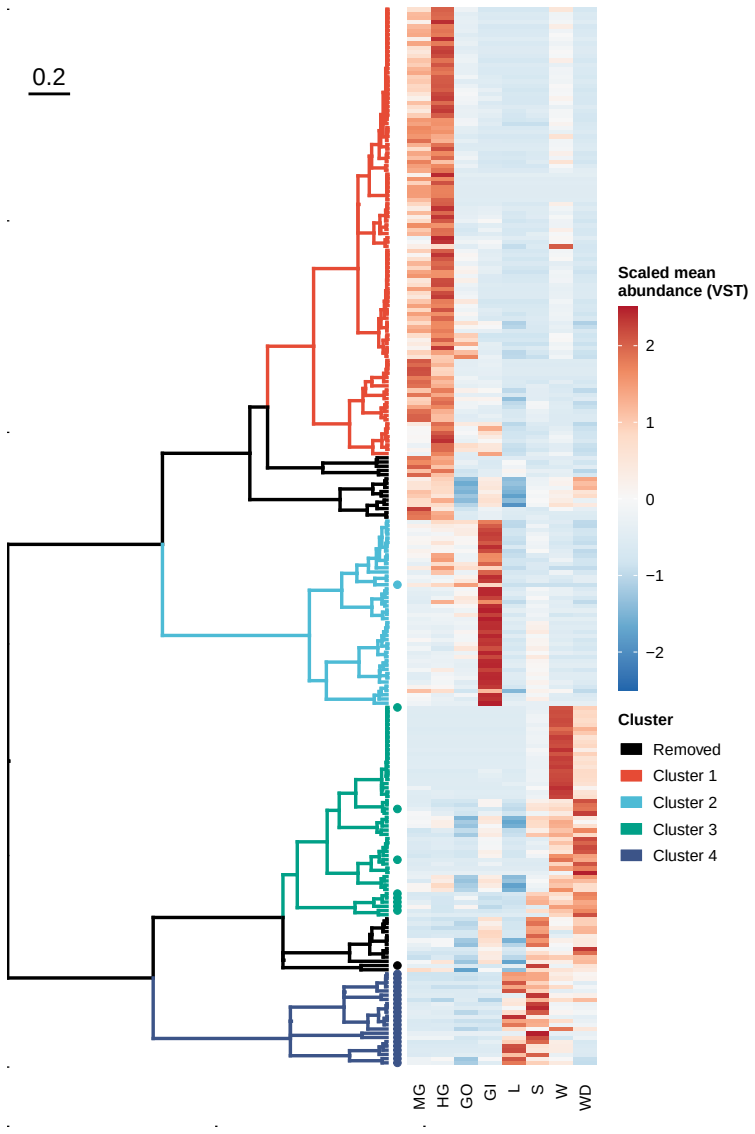
863

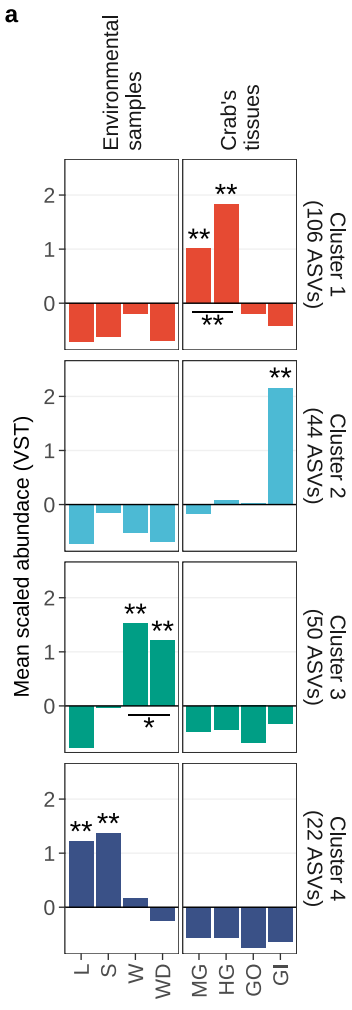
864

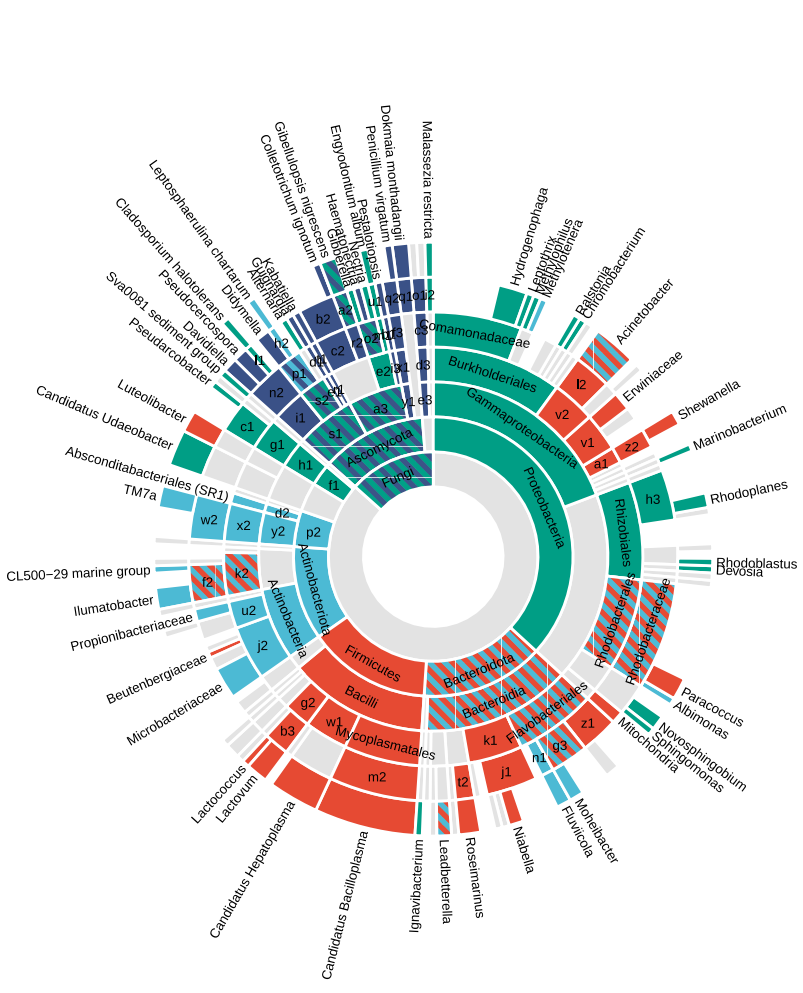

22.6
22.5
22.4
22.3
22.2
22.1
Longitude
113.9
114.1
114.3
114.5
Latitude
0
10
20km
N
Shui Hau
Sai Keng
To Kwa Peng


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

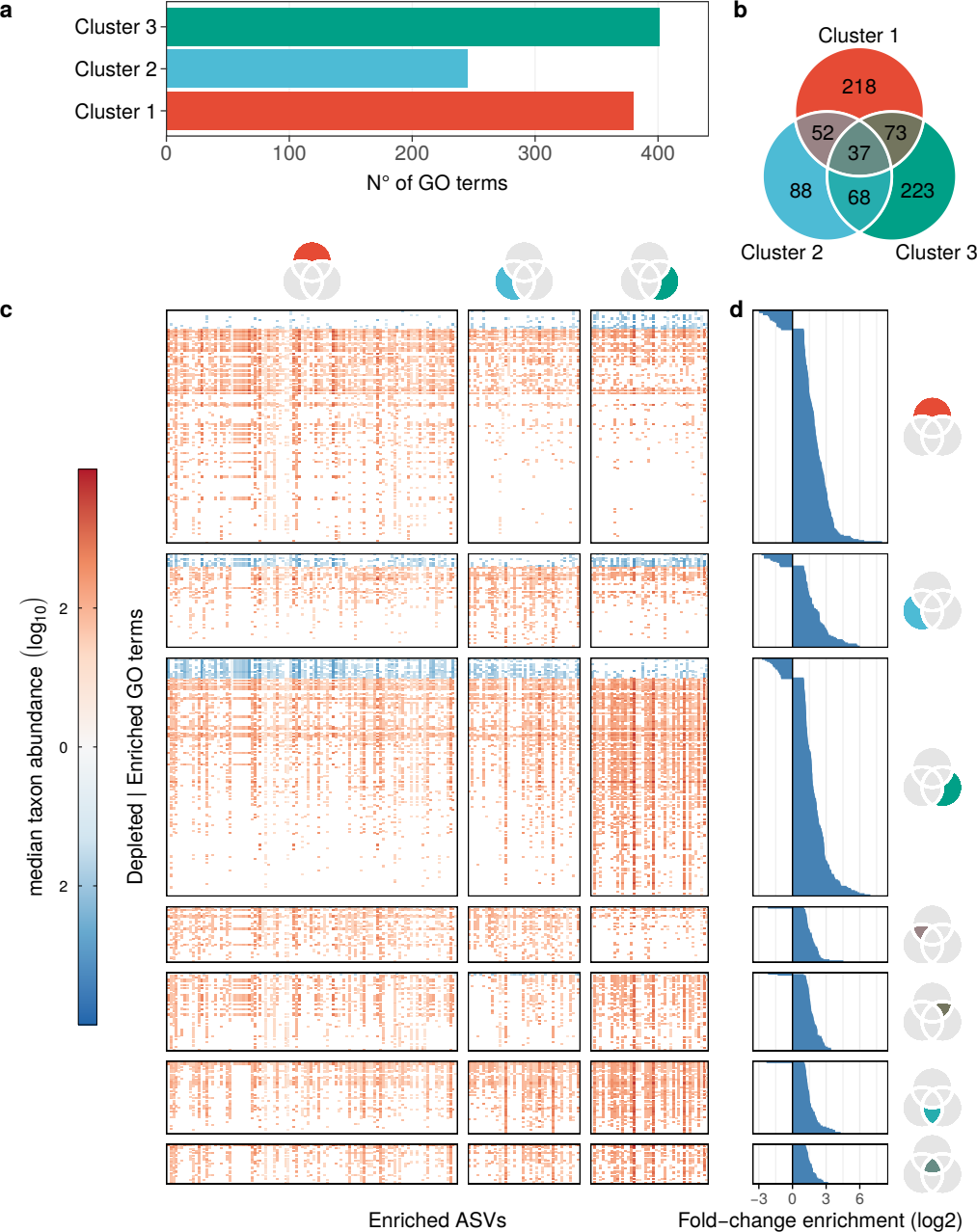

0.2
Scaled mean abundance (VST)
2
1
0
−1
−2
Cluster
Removed
Cluster 1
Cluster 2
Cluster 3
Cluster 4
MG
HG
GO
GI
L
S
W
WD





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)