bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

3

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

6 Affiliations:
7 [1] Department of Biology, University of Florence, Sesto Fiorentino, IT50019, Italy

8 [2] The Swire Institute of Marine Science and the Division of Ecology and Biodiversity, The
9 University of Hong Kong, Hong Kong, Hong Kong SAR

10 [3] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, IT38098,
11 Italy

12 * These authors contributed equally to this work

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

14 ## Abstract

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

27

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

28 **Introduction**

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

41 In chronological order, the true crabs (Decapoda; Brachyura) were among the last groups of
42 arthropods to perform the sea-land transition (*5*), but they have been particularly successful
43 with more than twenty phylogenetically unrelated families that include species with
44 intertidal, semiterrestrial and truly terrestrial lifestyles (*5*, *7*). Adaptations to terrestrial
45 environments involved several morphological and physiological changes in sensory,
46 locomotory, respiratory, excretory, osmoregulatory, digestive and reproductive systems (*8*, *9*,
47 *18*, *10*–*17*). Interestingly, brachyuran gills have not proved to be fully functional for
48 respiration in terrestrial environment (*1*) and are involved in osmotic regulation, urine
49 dilution, nitrogen excretion and ion-regulation in most terrestrially adapted families (*10*, *19*,
50 *20*). Other essential adaptations include herbivory and immune system recognition of
51 potentially fungal pathogens, which are rare in marine environment but abundant on land,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

82 Results

83 Microbial composition of crab's organs and environmental samples

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

101 0.42% of the total ASVS detected). In general, microbial diversity was higher in the
102 environmental samples than in crab's organs, except for the gills where we found a high
103 diversity of 16S rRNA amplicons (Figure 2a and b, Table S3). The fungal diversity was low in
104 all crab's organs, whereas environmental samples were characterised by a highly diverse
105 fungal assemblage, with soil, water, and water debris reporting the highest diversities (Figure
106 2c & d). For each sample category and at each site, the difference between extrapolated
107 diversity—namely the inverse Simpson index calculated by simulating a higher sequencing
108 depth than the observed depth for each sample—and observed diversity was really low
109 (Table 1), meaning that our sampling effort was effective in assessing the natural microbial
110 diversity for each category. Even the Good's coverage estimator was higher than 99.9% in all
111 sample types highlighting that only 1‰ of clustered sequences came from ASVs detected only
112 once in the whole dataset (Table 1 and Table S3). Microbial diversity was rather uniform
113 across sampling sites, except for To Kwa Peng (TKP), which showed a significantly lower
114 diversity with respect to the other sites in terms of ITS1 data (Figure S3).

115 **Microbial distribution across sample types**

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

143 **Defining characteristic patterns across sample types**

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

161 Taxonomic and functional enrichments in crab's organs and environmental 162 samples

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

218 ## Discussion

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

250 the present integrated set of results support the novel idea that semi-terrestrial crabs, only
251 recently migrated onto the land, are associated with microbial communities that play specific
252 functional roles in response to the dramatic environmental pressures posed by the sea-to-
253 land transition. The next questions are where these microbes were acquired and where are
254 their reservoirs?

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

267 The digestive system of brachyuran crabs comprises different tracts according to their
268 embryological origin and functional role. We intentionally focused on the midgut, where
269 central digestion and absorption occur, and the hindgut, which has a role in water and ion
270 transport (*8*, *16*). In contrast with the intestinal bacterial communities found in the
271 freshwater Chinese mitten crab, *Eriocheir sinensis* (*40*), and notwithstanding the different
272 functions, the selected intestinal tracts of *C. haematocheir* were homogeneous in terms of
273 microbial communities, suggesting, at least for this terrestrial crab species, a similar role of
274 such traits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

275 Host-associated microbiota are known to play a crucial role in the digestive process of many
276 terrestrial arthropods (*41*). Thus, it is not surprising that the highly diverse microbial
277 community found in *C. haematocheir* gut is clearly distinguished from all the other internal
278 and environmental assemblages. This semi-terrestrial species mainly relies on leaf litter and
279 integrates its diet with small arthropod preys (*28*, *29*). The ability to cope with difficult-to-
280 digest vascular plant compounds (i.e., cellulose, lignin, and polyphenols) has been brought
281 forward as a fundamental trait in the adaptive processes related to terrestrialisation (*16*, *42*,
282 *43*). The termites' ability to degrade lignocellulose is classical example of high gut microbiome
283 specialization (*41*). Conversely to the termites' gut microbiome, however, the bacterial
284 assemblages inhabiting the gut of *C. haematocheir* were not enriched in known functions that
285 would help the host to selectively digest cellulose products. Indeed, no gene with cellulase
286 activity were enriched in the gut, possibly reflecting a more variable diet of these crabs, not
287 strictly specialized on plant material digestion (*28*). The gut microbiome shared limited
288 similarities with the gut core microbiome of a marine predatory brachyuran, the
289 commercially important mud crab *Scylla paramamosain* (*44*). Core Proteobacteria shared by
290 *S. paramamosain* and *C. haematocheir* include only *Shewanella*, while Firmicutes include
291 *Lactococcus, Candidatus Hepatoplasma* and *Candidatus_Bacilloplasma* (hereafter
292 *Hepatoplasma* and *Bacilloplasma*). *Hepatoplasma,* and *Bacilloplasma* are rare in marine
293 environments, while they were previously described as colonizers of the hepatopancreas of
294 the terrestrial isopod *Porcellio scaber* (*33*, *45*). *Hepatoplasma* was detected in *P. scaber*
295 specimens collected from very different geographical areas (Germany and western Canada),
296 proving the strong host-microbe association (*45*, *46*)., Other bacterial genera found in our
297 intestinal cluster (such as *Niabella*, *Paracoccus* and *Shewanella*) were also isolated from
298 terrestrial isopods (*42*). Of particular interest is also the presence of *Lactococcus*, which is a
299 genus of lactic acid bacteria known as homofermenters, producing a single product, lactic

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

300 acid, as the main, or only product of glucose fermentation. Lactic acid has a selective role since
301 it lowers the pH and selects the environmental microbes that can potentially thrive in the gut.
302 These microorganisms, common in the dairy industry, are also known to colonize the gut of
303 termites, *Lactococcus nasutitermitis* (*47*)and do not include marine species, although they are
304 occasionally described as pathogens of aquacultured species (*48*, *49*). Another microorganism
305 we found in the intestinal cluster that has never been described in the marine environment,
306 thus suggesting acquisition from terrestrial microorganisms, is *Lactovum*, a genus of bacteria
307 within the family Streptococcaceae. The genus includes a single species, *Lactovum miscens*, an
308 aerotolerant, anaerobic species originally isolated from soil of the Stiegerwald forest in
309 Germany (*50*). In addition, the intestinal cluster was enriched in *Erwiniaceae*, a family that
310 includes insect's symbionts such as *Buchnera aphidicola* (*51*). The gut microbiota we found in
311 the mid- and hindgut of *C. haematocheir* then shared similarities with the gut microbiota of
312 other terrestrial crustaceans and insects, suggesting the presence of functions/preferences of
313 those taxa that could help arthropods digestive systems, a hypothesis that deserves further
314 functional investigations.

315 The possibilities of a vertical transmission vs. acquisition from the environment of this
316 specialized gut microbiota deserves a deeper discussion and further experiments. First, land
317 and terrestrial crabs are neither social or gregarious and only few of them show some basic
318 degrees of parental care (*17*, *52*). Second, *C. haematocheir*, as well as the most land-adapted
319 family of brachyuran crabs, the Gecarcinidae, still retain an indirect development strategy and
320 release planktonic larvae in coastal waters. Consequently, *C. haematoicher*, as well as the land
321 crabs with planktonic larvae in general, may have developed intimate gut-specific symbiotic
322 associations only through pseudo-vertical transmission (sensu *53*), i.e., by acquiring and
323 selecting their gut microbiota directly from food or, most probably, through consumption of
324 adults' faeces (coprophagy) found in or around their burrows. The megalopae of various

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

325 terrestrial and land crabs are known to specifically recruit at their spawning grounds, as
326 showed by the iconic mass recruitment of the Christmas Island red crab *Gecarcoidea natalis*
327 (*54*), or in areas where they can chemically detect adult populations, such as the case of the
328 swamp ghost crab *Ucides cordatus* (*55*). In some species, this selective recruitment towards
329 populations of conspecific adults ultimately results in the occurrence of juveniles and sub-
330 adults within secondary branches excavated along the burrows of adult crabs (*56*–*58*). Indeed,
331 these behaviours maximise the chances of finding adult populations and, consequently, a
332 successful pseudo-vertical transmission of microbiota through coprophagy. Further
333 metagenomic analyses on faeces are needed to clarify this point.

334

335 The gonads host a low microbial diversity with respect to the other organs and environmental
336 matrices we analysed, as expected for an internal organ in no direct contact with the
337 environment and not morphologically connected to the gut and the gills. Their microbial
338 community, moreover, is shared with both the other organs and the environment,
339 determining the absence of a specific taxonomic and functional cluster associated with them.
340 Microorganisms are known to be associated with the reproductive system of arthropods, in
341 both insects (*59*) and crustaceans (*60*). Few studies, however, have specifically characterised
342 these gonad-associated microbiota and most of them were focused on specific pathogens of
343 sexually transmitted infections or reproductive parasites, with the most notable example
344 being *Wolbachia* (*61*). When present, the vertical transmission of specific microbiota during
345 oogenesis or at birth is known to determine a colonization of the gonads that ultimately
346 affects the microbial composition of the offspring (*60*). In this view, the absence of an
347 exclusive microbiota in *C. haematoicher* gonads can be explained, as for the gut microbiota, by
348 the lack of vertical transmission through the parental-offspring pathways, as a result of the
349 presence of planktonic stages.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

350 Although protected in the gill chambers, *C. haematocheir* gills are in direct communication
351 with the external environment, but they do show a unique resident microbiota, which is
352 different from the microbial communities detected in the environmental matrices. These gill-
353 associated microbial assemblages are only enriched in prokaryotes, unlike the soil/litter and
354 the water communities, which are composed almost entirely by fungi and a mix of
355 Proteobacteria and fungi, respectively. The prokaryotic cluster associated with the gills is
356 characterised by a strong uniformity in terms of both taxonomy and functions. Actinobacteria
357 associated to the gills include Microbacteriacea and *Illumatobacter*, which are known to play
358 an important role in marine organisms by producing bioactive compounds crucial in the
359 defence against pathogens (*62*). Some of the bacteria detected on the gills of *C. haematocheir*,
360 such as *Ilumatobacter* and *Albimonas*, were also found in the gills of the Chinese mitten crab
361 *Eriocheir sinensis* (*40*), but at this stage their metabolic functions are not clear. In contrast to
362 the gut, genes associated with cellulose synthase activity were found in the microbiota
363 associated with these organs (GO:0016760). This activity is typical of biofilm-forming bacteria
364 that use cellulose both as a physical barrier against armful molecules such as antibiotics but
365 also biocides and metallic cations, and as a molecular glue to help their interaction with the
366 host (*31*). Since the gills of *C. haematocheir* are exposed to external perturbations, the
367 presence of biofilm-forming functions may help to boost the resilience of bacteria stabilizing
368 host-microbiome interactions. Genes related to the reduction of nitric compounds (namely:
369 nitric oxide reductase activity, GO:0016966, and nitrous-oxide reductase activity,
370 GO:0050304) were enriched in both the gills and environmental water and water debris
371 samples. Besides their role in the anaerobic metabolism of nitrogen, these functions are
372 involved into pathogenesis and antibiotic resistance in bacteria (*63*) and may help tissue
373 colonization. In addition, the covalent incorporation of a nitric oxide molecule (nitrosylation)
374 of cytochrome c and quinol oxidases inhibits cellular respiration acting as an antimicrobial

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

378 Our results show that fungi are strongly depleted in all crab's organs and that gills should be
379 considered an efficient selective filter between the environment and the host. In our opinion,
380 the substantial absence of fungal associations can be explained by different hypotheses that
381 all converge on the existence of defences from potential pathogens. We may hypothesize that
382 some physiological and anatomical characteristics of specific organs, such as gut and gills, may
383 inhibit adhesion and development of fungi. In insects, for instance, physiochemical
384 characteristics of the cuticle, or feeding habits, can counteract the entomopathogenic fungi
385 cuticular adhesion, germination of spores and hyphal growth (*64*). Moreover, prokaryotes-
386 fungi direct competition through bacterial production of antifungal molecules may limit the
387 growth of the latter. Events of fungal exclusion by direct competition mediated by symbiotic
388 bacteria have been extensively described in beetles (*65*). *Streptomyces* sp. SPB74, symbiont of
389 the beetle *Dendroctonus frontalis*, produces an active compound, mycangimycin, which
390 specifically inhibits antagonist fungi without acting on the mutual ones (*65*). This protective
391 role may be extended to the offspring. Antifungal molecules produced by the symbiotic
392 bacterium *Burkholderia gladioli* are known to protect the eggs of the beetle *Lagria villosa*
393 against pathogenic fungi (*35*, *66*). Pathogenic fungi have been reported in intertidal and
394 terrestrial crabs, such as the case of the swamp ghost crab, *U. cordatus*, infected by species of
395 black yeast, *Exophiala cancerae*, and *Fonsecaea brasiliensis*, which cause a condition called
396 "lethargic crab disease" (*21*, *22*). The role of prokaryotic associations in the defence against
397 pathogenic fungi can be of critical importance for terrestrial crabs, since the immune
398 molecules of crustaceans are less efficient than the ones of insect against fungal infections
399 (*67*).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

400 In conclusion, this is the first attempt to ascertain the role of host-microbiome associations in
401 the transition to land of brachyuran crabs, the most recent arthropod taxa to perform such an
402 evolutionary leap. Using a targeted metagenomic approach, we demonstrated persistent
403 organ-specific crab-microbiome associations which help a semi-terrestrial crab coping with
404 novel environmental challenges posed by its recently occupied habitats, i.e., the lowland
405 forests or East Asia. We found clear differentiations among the microbiota associated with
406 different crab's organs and the ones found in the environmental matrices, suggesting a
407 selective acquisition of the microbiota through "gain" and "loss" mechanisms. The differences
408 found among the organ-related and environmental clusters were not merely taxonomical
409 since the different clusters harbour different metabolic profiles. These results corroborate the
410 hypothesis that the recorded differences can be due to the presence of metabolic
411 complementation mechanisms that took place in those organs mostly impacted by the
412 challenges posed by the terrestrial environment. Our data also show possible evolutionary
413 convergences towards a uniform 'terrestrial intestinal microbiota' across different lineages of
414 arthropods, suggesting the presence of bacterial associations linked to terrestrial life. The
415 present study strongly supports the hypothesis that semi- and terrestrial crabs are an
416 appropriate model system to study the evolution of arthropod-microbe interaction under the
417 selective pressures posed by the sea-land transition, which is happing right now in Brachyura
418 (*23*).

419

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

420 # Material and methods

421 ## Study species

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

429 ## Sample collection and total DNA extraction

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

455 **16S (V3-V4) rRNA gene amplification and sequencing**

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

480 **ITS1 rDNA region sequencing**

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

503 **Amplicon sequence variant inference**

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

524 **Inferring functional content of amplicon variants**

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

531 **Statistical analyses**

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

539 **References**

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

549 5. F. Giomi, M. Fusi, A. Barausse, B. Mostert, H.-O. Pörtner, S. Cannicci, Improved heat
550 tolerance in air drives the recurrent evolution of air-breathing. *Proc. R. Soc. B Biol. Sci.*
551 **281**, 20132927 (2014).

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

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

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

560 9. C. A. Farrelly, P. Greenaway, Land crabs with smooth lungs: Grapsidae, Gecarcinidae,
561 and Sundathelphusidae ultrastructure and vasculature. *J. Morphol.* **215**, 245–260
562 (1993).

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

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

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

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

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

575 15. R. Simoni, F. Giomi, D. Spigoli, H.-O. Pörtner, S. Cannicci, Adaptations to semi-terrestrial
576 life in embryos of East African mangrove crabs: a comparative approach. *Mar. Biol.* **160**,
577 2483–2492 (2013).

578 16. E. S. Lindquist, K. W. Krauss, P. T. Green, D. J. O'Dowd, P. M. Sherman, T. J. Smith III, Land
579 crabs as key drivers in tropical coastal forest recruitment. *Biol. Rev.* **84**, 203–223
580 (2009).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

602 25. I. Zilber-Rosenberg, E. Rosenberg, Role of microorganisms in the evolution of animals
603 and plants: The hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723–735
604 (2008).

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

607 27. E. Rosenberg, O. Koren, L. Reshef, R. Efrony, I. Zilber-Rosenberg, The role of
608 microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* **5**, 355–362
609 (2007).

610 28. K. F. Abd El-Wakeil, Trophic structure of macro- and meso-invertebrates in Japanese
611 coniferous forest: Carbon and nitrogen stable isotopes analyses. *Biochem. Syst. Ecol.* **37**,
612 317–324 (2009).

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

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

619 31. H.-C. Flemming, J. Wingender, The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–633
620 (2010).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

671 51. X. Chen, M. D. Hitchings, J. E. Mendoza, V. Balanza, P. D. Facey, P. J. Dyson, P. Bielza, R.
672 Del Sol, Comparative Genomics of Facultative Bacterial Symbionts Isolated from
673 European Orius Species Reveals an Ancestral Symbiotic Association . *Front. Microbiol.* .
674 **8** (2017), p. 1969.

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

677 53. D. M. Wilkinson, The Role of Seed Dispersal in the Evolution of Mycorrhizae. *Oikos.* **78**,
678 394 (1997).

679 54. A. M. Adamczewska, S. Morris, Ecology and Behavior of Gecarcoidea natalis , the
680 Christmas Island Red Crab, During the Annual Breeding Migration. *Biol. Bull.* **200**, 305–
681 320 (2001).

682 55. K. Diele, D. J. B. Simith, Effects of substrata and conspecific odour on the metamorphosis
683 of mangrove crab megalopae, Ucides cordatus (Ocypodidae). *J. Exp. Mar. Bio. Ecol.* **348**,
684 174–182 (2007).

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

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

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

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

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

697 61. F. Chevalier, J. Herbinière-Gaboreau, D. Charif, G. Mitta, F. Gavory, P. Wincker, P. Grève,
698 C. Braquart-Varnier, D. Bouchon, Feminizing Wolbachia: a transcriptomics approach
699 with insights on the immune response genes in Armadillidium vulgare. *BMC Microbiol.*
700 **12 Suppl 1**, S1–S1 (2012).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

705 last universal common ancestor to modern bacterial pathogens. *Curr. Opin. Microbiol.*
706 **29**, 1–8 (2016).

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

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

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

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

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

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

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

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

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

732 73. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes,
733 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*.
734 **13**, 581–583 (2016).

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

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

740 76. V. Deshpande, Q. Wang, P. Greenfield, M. Charleston, A. Porras-Alfaro, C. R. Kuske, J. R.
741 Cole, D. J. Midgley, N. Tran-Dinh, Fungal identification using a Bayesian classifier and the
742 Warcup training set of internal transcribed spacer sequences. *Mycologia*. **108**, 1–5
743 (2016).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

754

755 Acknowledgments

756 We thank the friends of the Integrate Mangrove Ecology laboratory of the Swire Institute of
757 Marine Science, HKU for their help in sampling and lab work. We also thank Prof Colin Little
758 and Gray A. Williams for 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
770 public Github repository available at:

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

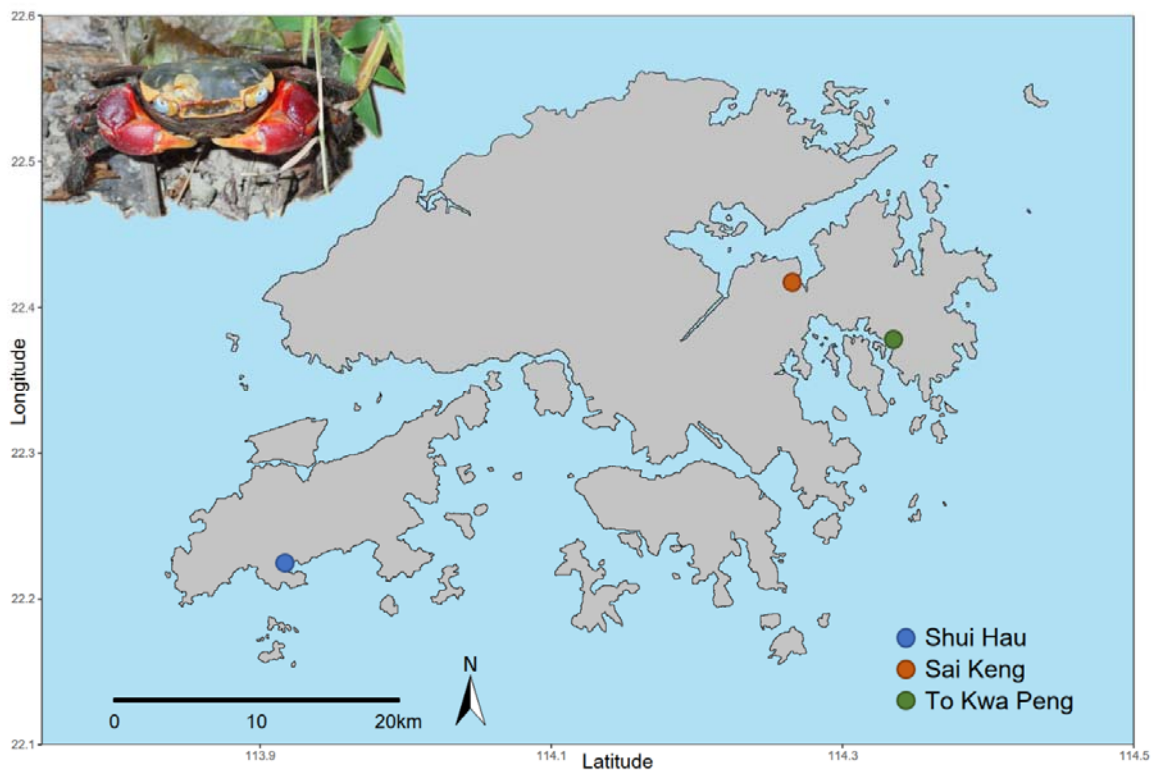
771 https://github.com/GiBacci/Chiromantes_haematocheir_microbiome. Sequencing data were
772 uploaded to the European Nucleotides Archive (ENA) under the project ID: PRJEB43930.

773 **Funding**: This work was supported by the project "The sky's the limit: the irresistible ascent
774 to land and trees by crabs", sponsored by TUYF Charitable Trust funds, Hong Kong, (HKU no
775 260008686.088562.26000.400.01) and by the Eighth Government Matching Grant, Hong
776 Kong Government (HKU no 207080320.088562.26020.430.01) to S.C. and by the RAE
777 Improvement Fund from the Faculty of Science, HKU (HKU no
778 000250449.088562.26000.100.01) to S.C and D.C.

779

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

780
Figures
781



782

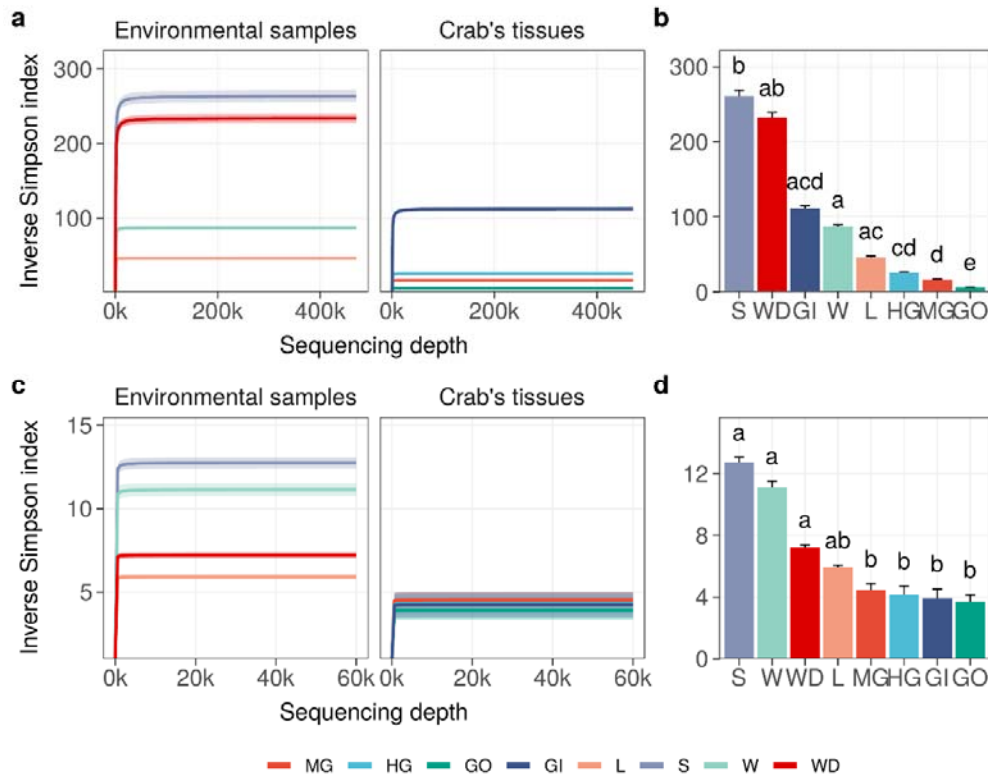
783 **Figure 1. Geographical distribution of the sampled populations of *C. haematocheir*.** The
784 map shows the names and location of the three sampling sites visited for the study in the New
785 Territories and on Lantau Island and an adult male *C. haematocheir* in its natural environment
786 (top left).
787

788

789

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

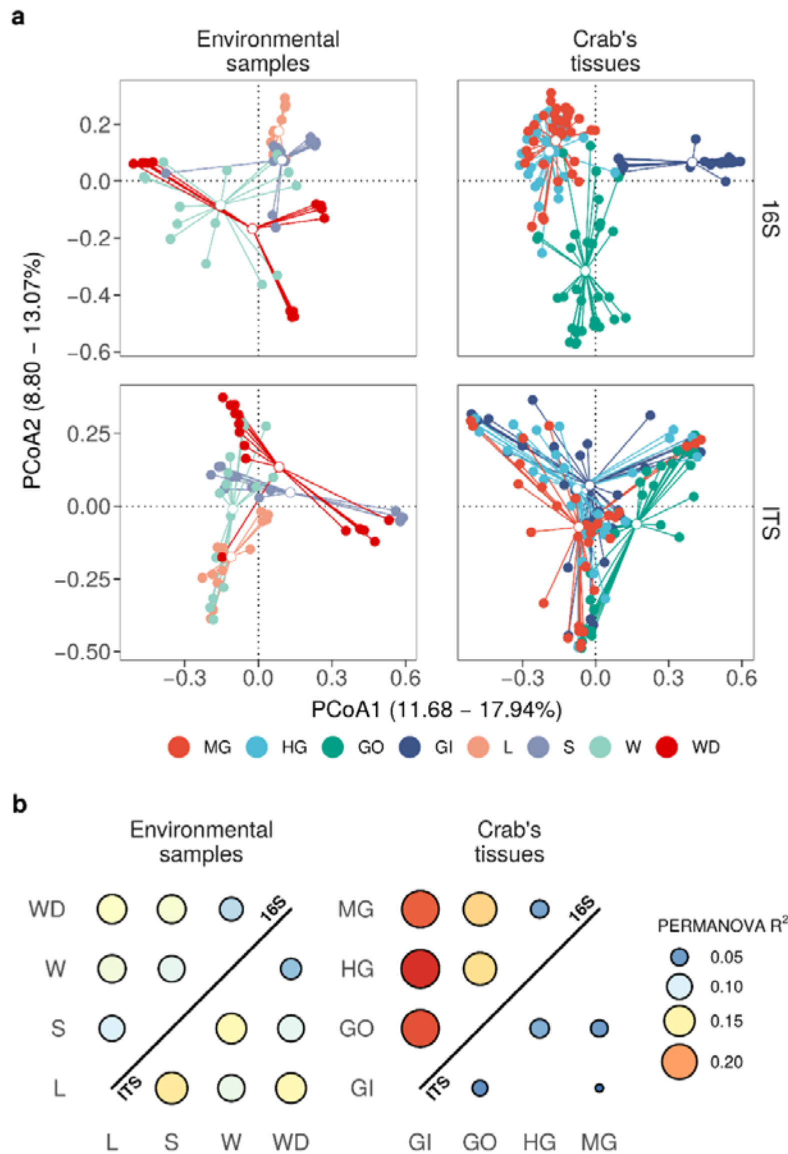
790



791

792 **Figure 2: Microbial diversity in crab's organs and environmental samples.** The average
793 of the inverse Simpson index was reported with increasing sampling effort for all types of
794 samples. Interpolated and extrapolated diversity was reported in panel a and c (16S rRNA
795 gene and ITS-1 region, respectively), whereas observed diversity was reported in panel b and
796 d (16S rRNA gene and ITS-1 region respectively). Significant differences in microbial diversity
797 (Wilcoxon non-parametric test) were reported using lowercase letters (panel b and d)
798 whereas colours and acronyms on the x-axis correspond to different sample types (MG,
799 midgut; HG, hindgut; GO, gonads; GI, gills; L, litter; S, soil; W, water; WD, water debris). If two
800 means were significantly different, all letters on top of the two boxes must be different; if two
801 means were equal, at least one letter must be the same.
802

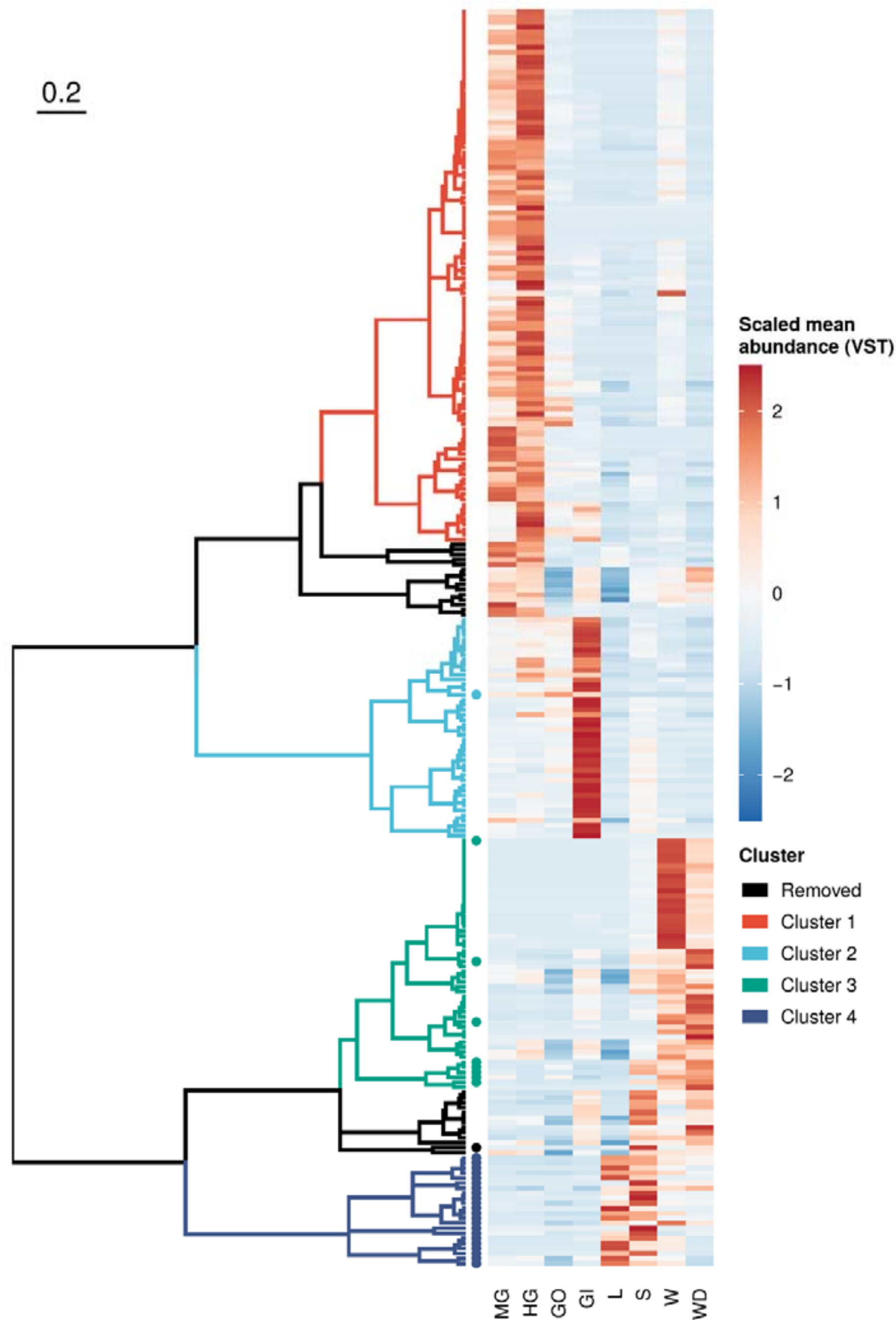
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



803

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



816
817 **Figure 4: Sequence variant clustering according to their abundance along sample types.**
818 Amplicon sequence variants reporting a different abundance pattern in one or more sample
819 types (loglikelihood ratio test of DESeq2) were clustered according to their mean variance-
820 stabilized abundance. Abundance values were reported using different colours after
821 clustering based on Kendall correlation (right side of the plot). Clusters were coloured
822 according to the scheme reported in the legend whereas removed clusters (namely those
823 composed with less than 10 variants) where reported in black. Sequence variants inferred
824 from ITS-1 amplicon sequencing were reported using a solid dot.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

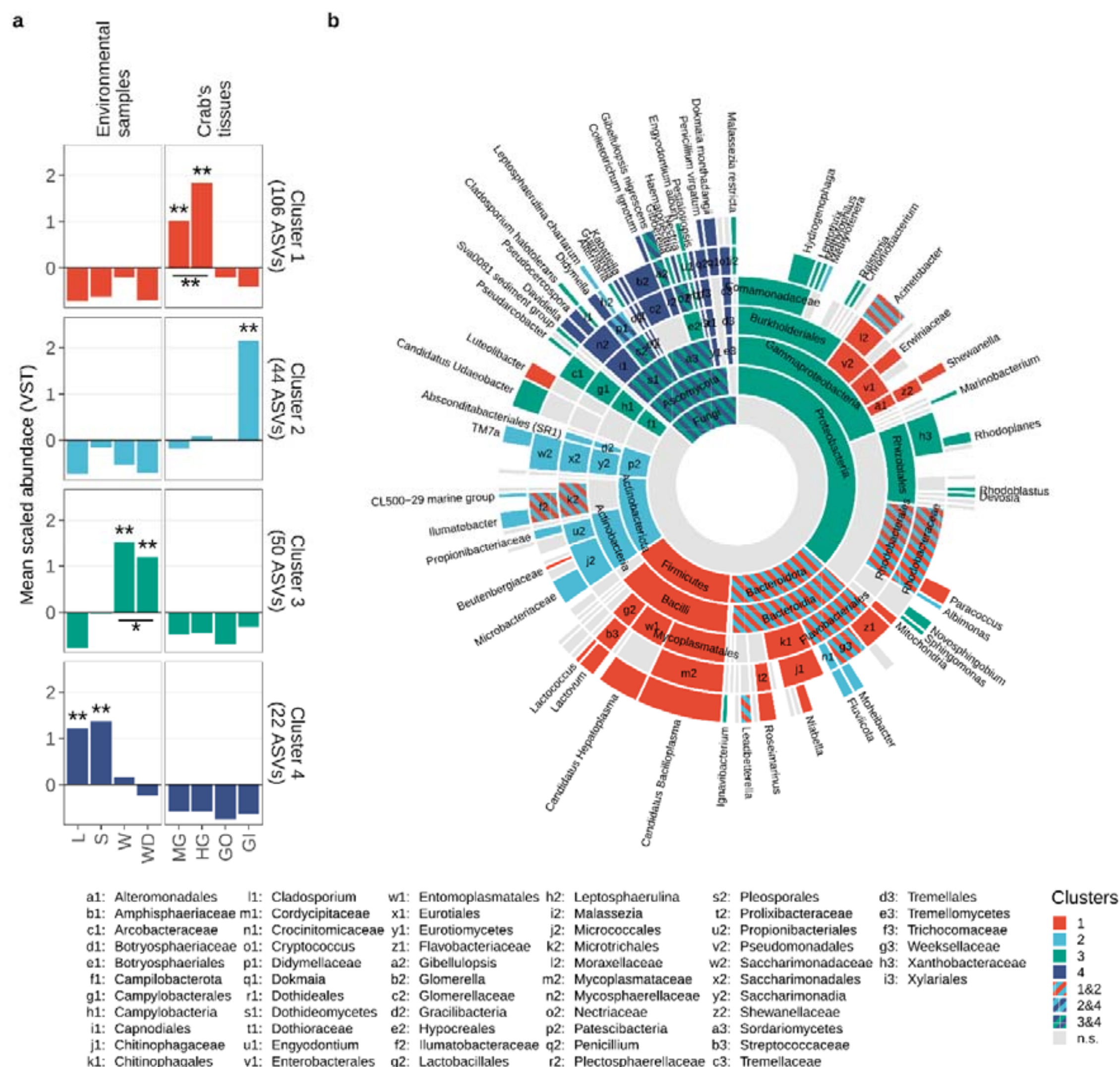
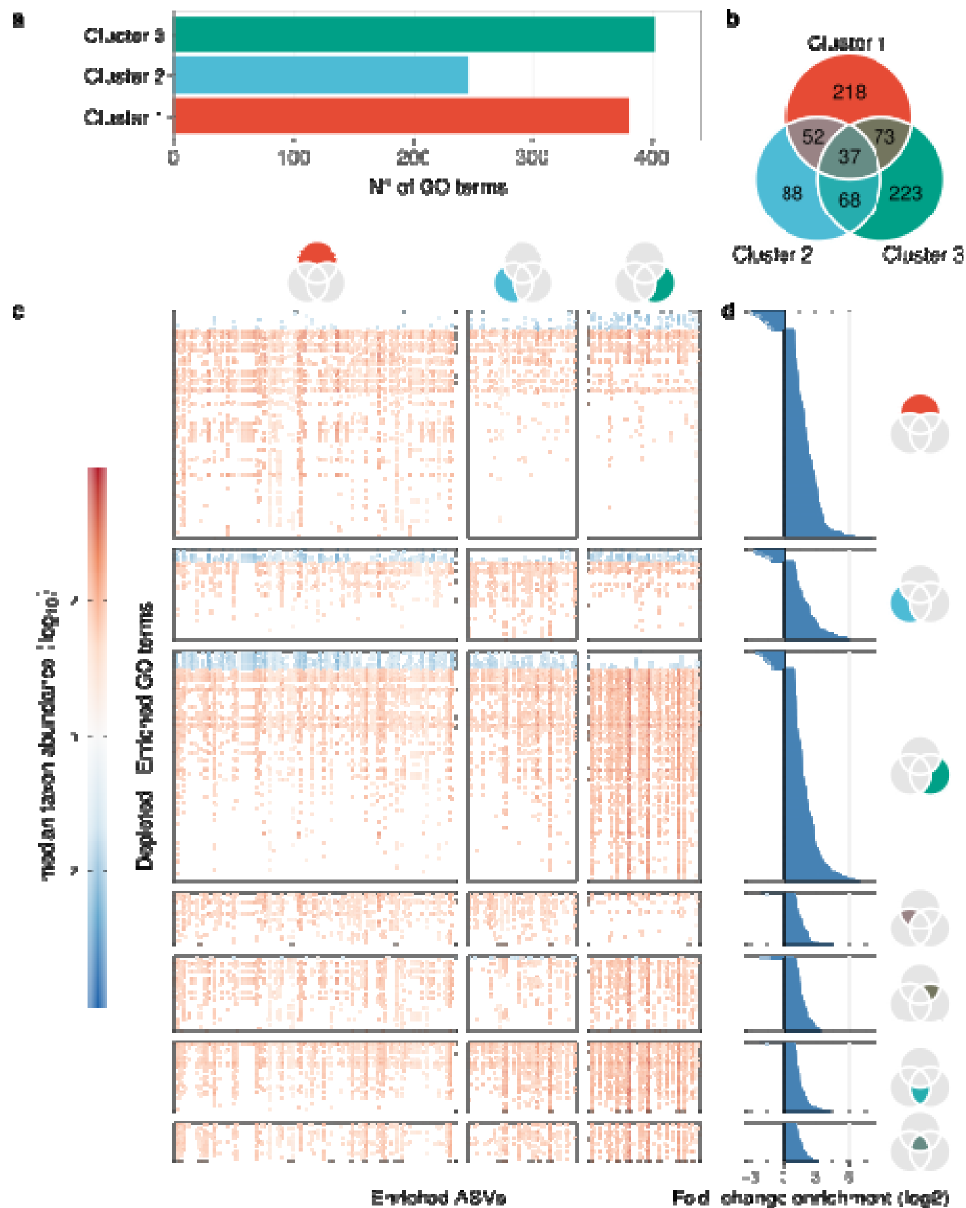


Figure 5: Taxonomic enrichment in clusters. a) Scaled variance-stabilized counts were tested for differences along sample types within all clusters detected. A pairwise Wilcoxon test was performed, and results were reported highlighting significant differences with one asterisk (p-value < 0.05) or two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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



837

838 **Figure 6: Enrichment analysis of "Gene Ontology" terms associated with detected**
839 **functions.** a) Number of GO terms associated to bacterial function detected in each cluster
840 (cluster 4 was not reported since it was entirely composed of Fungi). b) Venn diagram of GO
841 terms detected in the ASVs of each cluster. Diagram sections in common between two or more
842 clusters were coloured by interpolating colours reported in panel a. c) Median taxon
843 abundance of GO terms (y-axis) in all ASVs detected in each cluster (x-axis). The median taxon
844 abundance was reported using different shades of red—for enriched terms, namely those
845 detected with a higher frequency in respect to the whole population—and blue—for depleted
846 terms, namely those detected with a lower frequency than the rest of the population. The plot
847 was vertically divided according to ASV clusters whereas it was horizontally divided
848 according to venn diagram sections reported in panel b. d) Mean enrichment fold changes
849 associated to each term. Fold-changes were transformed using the logarithmic function (with
850 base equal two) to report enriched and depleted terms symmetrically around the zero.
851

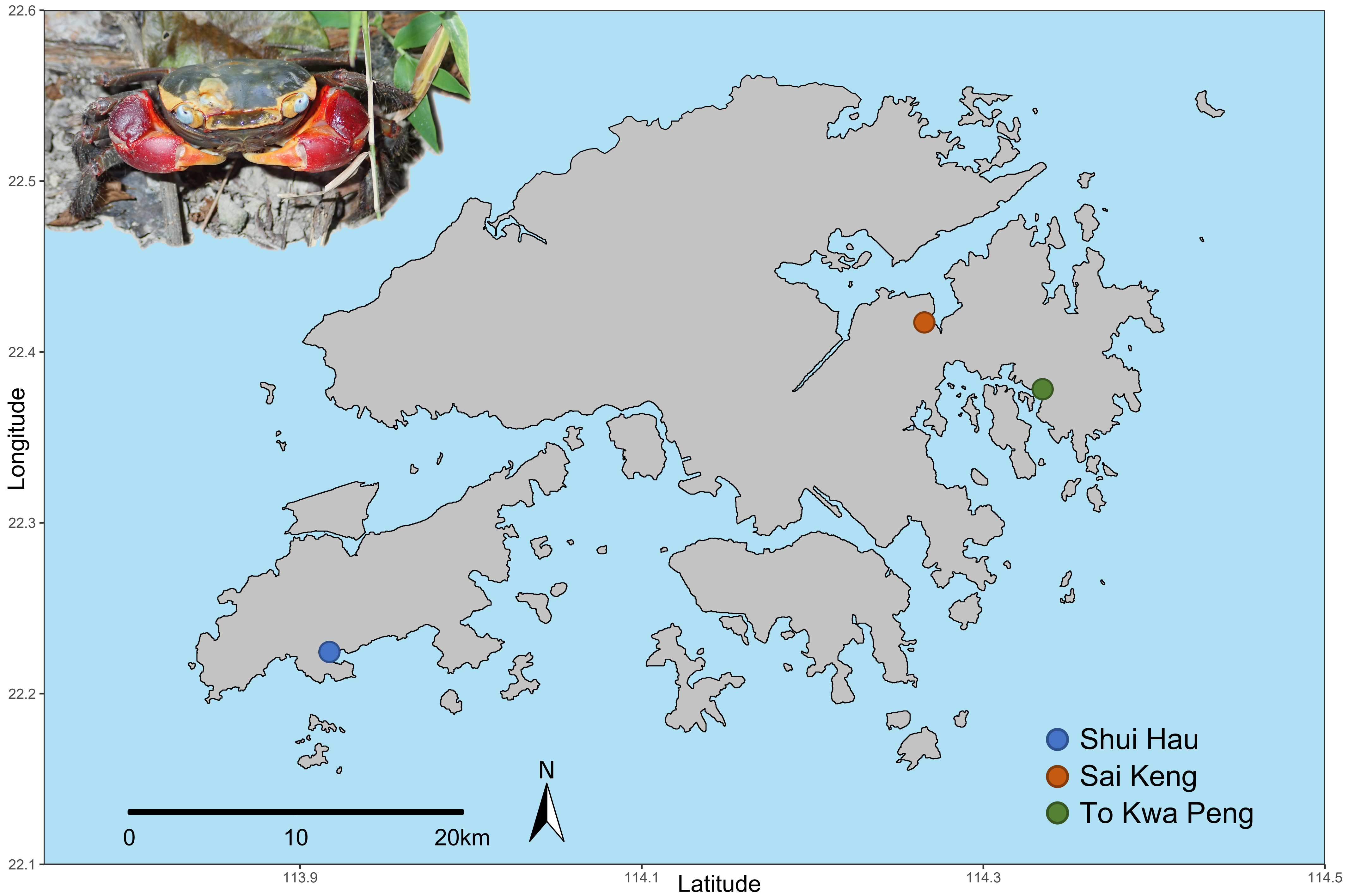
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.30.437674; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

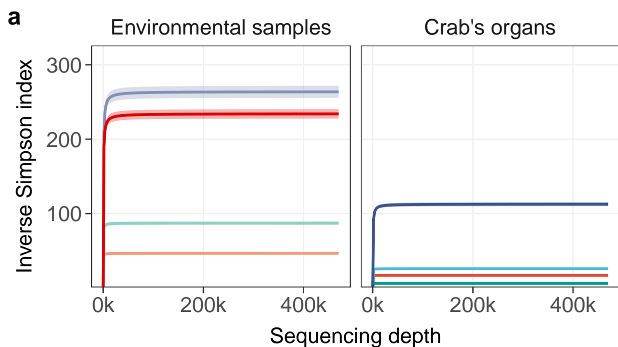
852 # Tables
853

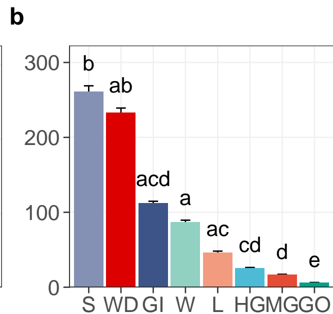
854 **Table 1: Mean differences between extrapolated diversity and observed diversity in**
855 **each sample type and site.** The differences between extrapolated Simpson diversity (Inverse
856 Simpson index computed for a sequencing depth higher than the observed one) and observed
857 diversity (Inverse 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

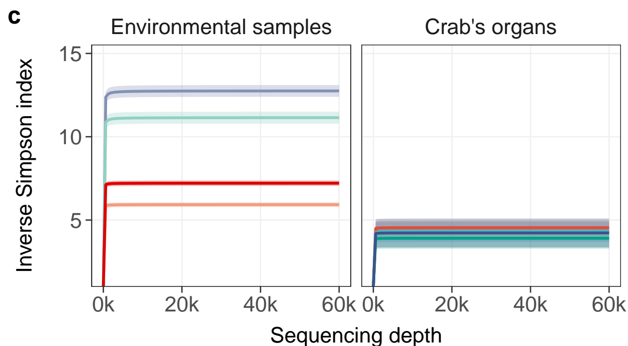
		Inverse Simpson (ext - obs)	Good's coverage estimator
Sample type			
	MG	0.051 ± 0.040	99.995 ± 0.001
	HG	0.132 ± 0.126	99.997 ± 0.001
	GO	0.085 ± 0.029	99.987 ± 0.005
	GI	0.575 ± 0.230	99.994 ± 0.002
	L	0.185 ± 0.114	99.963 ± 0.007
	S	2.741 ± 0.809	99.993 ± 0.001
	W	0.246 ± 0.066	99.997 ± 0.001
	WD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

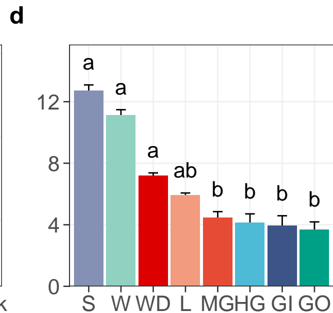
862
863
864


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
300
200
100
0
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
12
8
4
0
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

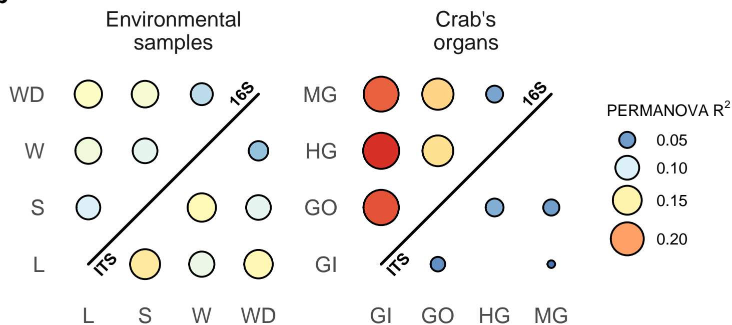
16S

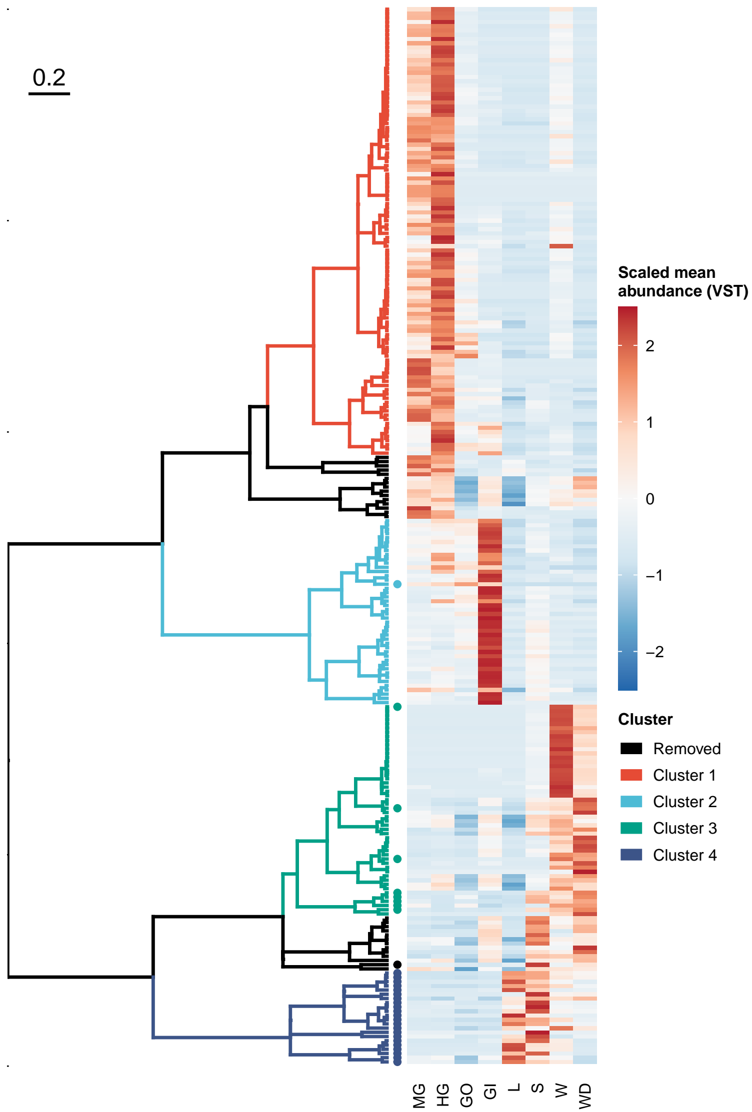
ITS

PCoA2 (8.80 – 13.07%)

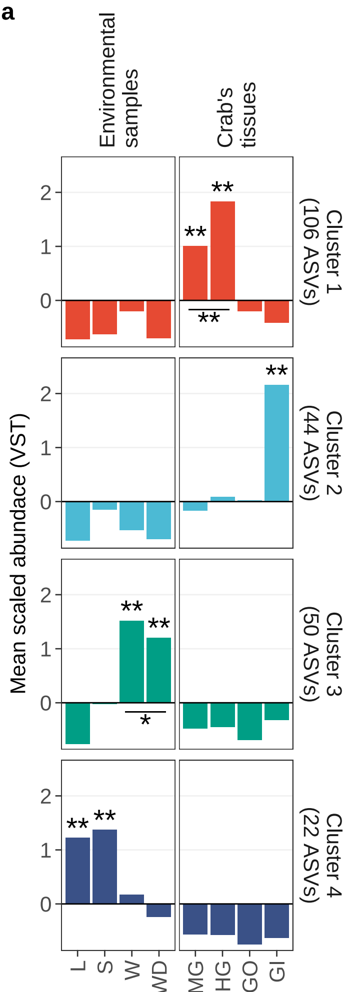
PCoA1 (11.68 – 17.94%)

MG HG GO GI L S W WD




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

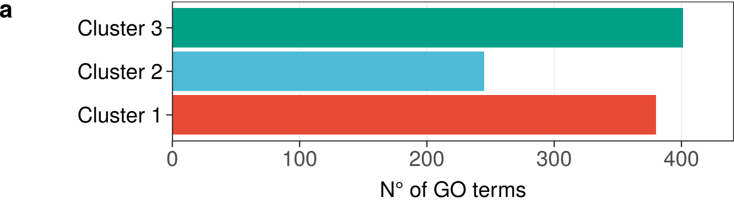
a

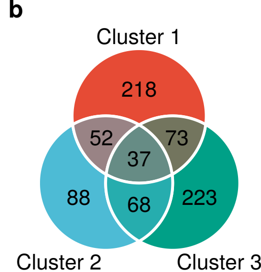


b

Proteobacteria
Gammaproteobacteria
Burkholderiales
Comamonadaceae
Rhizobiales
Rhodobacterales
Rhodobacteraceae
Bacteroidota
Bacteroidia
Flavobacteriales
Firmicutes
Bacilli
Mycoplasmatales
Actinobacteriota
Actinobacteria
Fungi
Ascomycota
Hydrogenophaga
Leptothrix
Methylophilus
Methylotenera
Ralstonia
Chromobacterium
Acinetobacter
Erwiniaceae
Shewanella
Marinobacterium
Rhodoplanes
Rhodoblastus
Devosia
Paracoccus
Albimonas
Novosphingobium
Sphingomonas
Mitochondria
Moheibacter
Fluviicola
Niabella
Roseimarinus
Leadbetterella
Ignavibacterium
Candidatus Bacilloplasma
Candidatus Hepatoplasma
Lactovum
Lactococcus
Microbacteriaceae
Beutenbergiaceae
Propionibacteriaceae
Ilumatobacter
CL500−29 marine group
TM7a
Absconditabacteriales (SR1)
Candidatus Udaeobacter
Luteolibacter
Pseudarcobacter
Sva0081 sediment group
Davidiella
Pseudocercospora
Cladosporium halotolerans
Didymella
Leptosphaerulina chartarum
Alternaria
Kabatiella
Colletotrichum ignotum
Gibellulopsis nigrescens
Haematonectria
Gibberella
Engyodontium album
Nectria
Pestalotiopsis
Dokmaia monthadangii
Penicillium virgatum
Malassezia restricta

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





c

median taxon abundance (log$_{10}$)

2

0

2

Depleted | Enriched GO terms

Enriched ASVs