Microbiome determinants of productivity in whiteleg shrimp aquaculture

Xiaoyu Shan[1], Kunying Li[1], Patrizia Stadler[1], Martha Borbor[2], Guillermo Reyes[2], Ramiro Solórzano[2], Esmeralda Chamorro[3], Bonny Bayot[2,4], Otto X. Cordero[1*]
1 Department of Civil and Environmental Engineering, MIT
2 Centro Nacional de Acuicultura e Investigaciones Marinas, CENAIM, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador
3 Universidade Federal de Santa Catarina, Brazil.
4 Facultad de Ingeniería Marítima y Ciencias del Mar, FIMCM, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador
* ottox@mit.edu

Abstract

Aquaculture holds immense promise for addressing the food needs of our growing global population. Yet, a quantitative understanding of the factors that control its efficiency and productivity has remained elusive. In this study, we address this knowledge gap by focusing on the microbiome determinants of productivity, more specifically animal survival and growth, for one of the most predominant animal species in global aquaculture, whiteleg shrimp (*Penaeus vannamei*). Through analysis of the shrimp-associated microbiome from 610 aquaculture samples across Asia and Latin America, we established the presence of core phylogenetic groups, widely prevalent across aquaculture conditions in disparate geographic locations and including both pathogenic and beneficial microbes. Focusing on early stages of growth (larval hatcheries), we showed that the composition of the microbiome alone can predict approximately 50% of the variation in shrimp larvae survival rates. Taxa associated with high survival rates share recently acquired genes that appear to be specific to aquaculture conditions. These genes are involved in the biosynthesis of growth factors and in protein degradation, underscoring the potential role of beneficial microorganisms in nutrient assimilation. In contrast, the predictability of the microbiome on the adult shrimp weight in grow-out farms is weaker (10-20%), akin to observations in the context of livestock. In conclusion, our study unveils a novel avenue for predicting productivity in shrimp aquaculture based on microbiome analysis. This paves the way for targeted manipulation of the microbiome as a strategic approach to enhancing aquaculture efficiency from the earliest developmental stages.

Introduction

In aquaculture, and animal farming in general, animal survival and individual growth yield are key factors that affect global measures of performance[1–3]. Improving resilience against disease (i.e. survival) and the efficiency with which animals assimilate their feed (individual growth yield) lies at the heart of addressing the twin challenges faced by the future food systems: elevating production to satisfy the escalating global food demand while reducing environmental impacts to align with sustainable development goals. Within this framework, the host microbiome is widely acknowledged to play an important role[4–9]. Beneficial microbes have the potential to improve survival and growth by facilitating the digestion and assimilation of nutrients[10] or safeguarding against diseases[11]. Conversely, pathogenic microbes may compromise production by hindering growth or increasing mortality[12], thus reducing yield in relation to resource consumption.

Since the late 1980s, aquaculture witnessed a rapid increase, in contrast to the stable production of marine capture fisheries, providing high-quality edible proteins and setting a new record with earnings of USD 265 billion (only for aquatic animals) and employing 20.7 million people as of 2020[13]. Whiteleg shrimp (*Penaeus vannamei*) farming, in particular, leads in aquaculture animal production with an annual production of 5.8 million tonnes in 2020[13]. Despite its rapid expansion, productivity of shrimp aquaculture frequently faces setbacks due to feed wastage[14] and pathogen infections[15], underscoring the need for technological advancements to improve production through microbiome engineering. However, the variable outcomes associated with probiotic use in shrimp farming have generated both enthusiasm and skepticism[16,17]. This highlights a significant gap in our understanding caused by a lack of comprehensive, systematic studies on how the microbiome composition affects productivity in shrimp aquaculture. Addressing this gap is crucial for leveraging microbiome-based strategies to their fullest potential.

In this study, we investigate the linkage between the microbiome and production variables in the aquaculture of *P. vannamei*. We start by characterizing a shrimp-associated microbiome across hatcheries and farms distributed across the globe. Subsequently, we focus separately on the two major phases of shrimp production: the hatchery phase and the grow-out phase. In the hatchery phase where shrimp transition from nauplii to postlarvae, survival rates are critical for productivity as early-stage shrimp are particularly vulnerable to environmental

stressors and pathogenic threats. We investigate how the microbiome influences these survival rates, taking advantage of the tightly controlled abiotic conditions in hatcheries. In the grow-out phase, where the major goal is achieving optimal growth as shrimp advance to marketable sizes, we focus on body weight of shrimp as a direct indicator of growth efficiency and productivity under varying conditions of water quality and nutrient availability. Through this approach, we aim to shed light on the potential of microbiome interventions in enhancing the sustainability and productivity of shrimp aquaculture.

Results

Characterizing the global shrimp-associated microbiome

To better understand the role of the microbiome on shrimp production, we first compiled a global dataset of aquaculture shrimp-associated microbiome (Method). Briefly, we re-analyzed microbiome sequencing data of 610 samples[18–27] from 12 geographically distant shrimp cultures in Asia and Latin America, including 232 samples of shrimp larvae microbiome and 378 samples of intestine (gut) microbiome from juvenile or adult shrimp (Figure 1A, Table S1). Recognizing the coastal ocean as the natural source of microbes in marine aquaculture, we also incorporated 154 samples of coastal seawater microbiome[28–30]. This extensive effort yielded a database of 56,513 ASVs (Amplicon Sequence Variants), including 48,514 shrimp-associated ASVs and 7,999 coastal seawater ASVs, now publicly available for further aquaculture research.

With this dataset of global shrimp-associated microbiome, we begin by asking whether the aquaculture environment selects for specific taxonomic clades, or it simply reflects a random assortment from the surrounding seawater. We find that the mean pairwise phylogenetic distances (MPD) between shrimp-associated sequences and coastal seawater sequences are significantly higher than that expected by a null model (Nearest Relatedness Index = 9.7, $P < 0.001$, see Method), suggesting a selective adaptation of specific microbes to a farm-associated environment. Indeed, we identified taxonomic groups that are significantly enriched or diluted in shrimp-associated microbiome. Oligotrophic taxa such as SAR11 and SAR86 are significantly less prevalent (Figure S1, adjusted $P < 0.01$), while copiotrophic taxa such as Vibrionales, Rhodobacterales and Flavobacteriales are among the most abundant ones across global aquaculture shrimp microbiome (Figure 1B). Interestingly, these taxa were also among the most abundant taxa in microbial communities assembled on marine particles of polysaccharides such

as chitin[31,32]. Vibrionales are known for fast growth rate, especially in saline and estuarine environments (0.5~3% NaCl) and high temperature (25~30°C) featured by aquaculture conditions[33]. Flavobacteriales are noted for their capabilities in degrading high-molecular-weight polysaccharide and proteins[34,35], whereas Rhodobacterales are specialized in scavenging metabolic byproducts such as organic acids and amino acids[34,35].

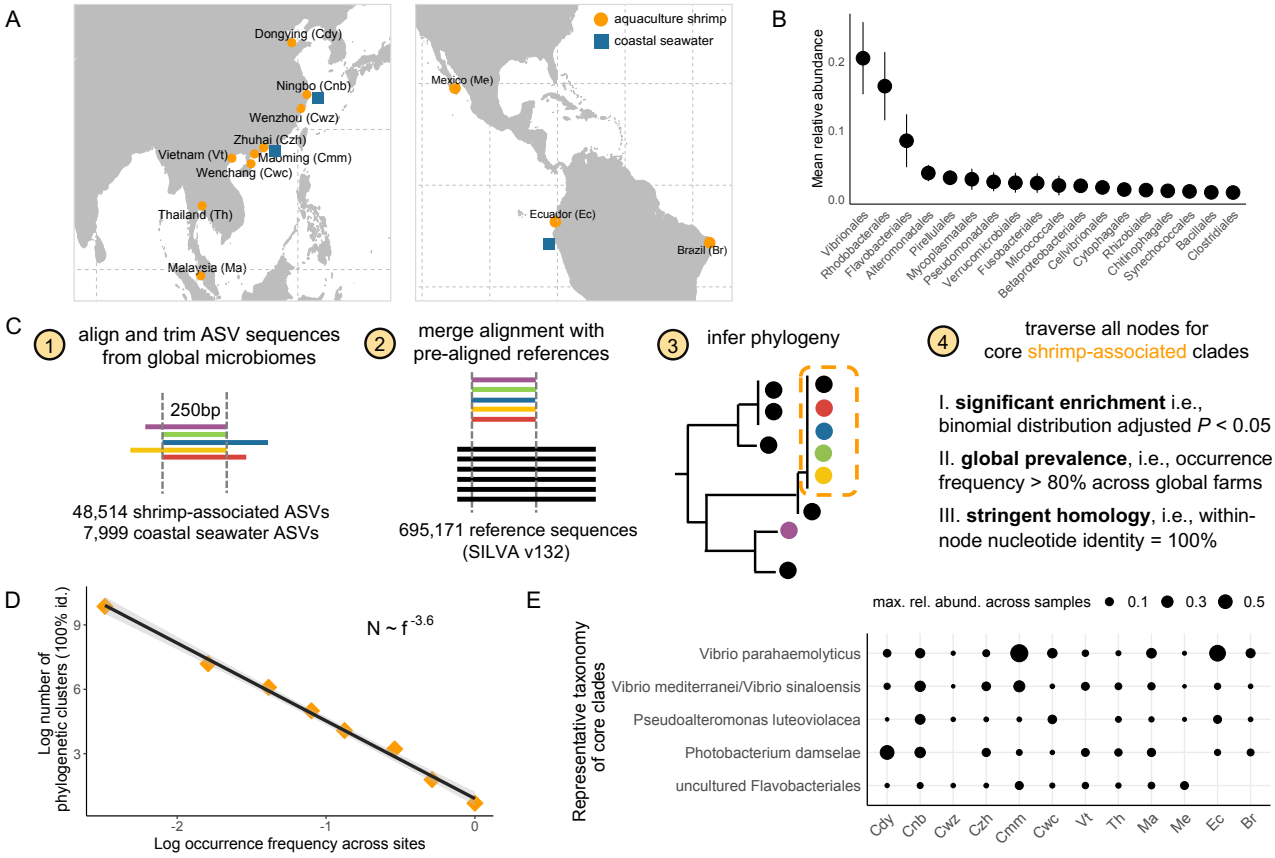

Figure 1 Identifying core clades of shrimp-associated microbiome from global aquaculture. (A) Geographic distribution of microbiome datasets compiled in this study. Orange circle indicates shrimp-associated microbiome. Blue square indicates coastal seawater microbiome. (B) Ranked mean relative abundance of bacterial taxa across global shrimp-associated microbiomes. (C) A bioinformatic pipeline is developed to identify core clades of shrimp-associated microbiome. See Methods for full descriptions. (D) A power law relationship between the number of phylogenetic clusters N and the occurrence frequency of phylogenetic clusters f. (E) Representative taxonomy of the identified core clades of global shrimp-associated microbiome. Size of dots indicates maximal relative abundance of a clade in a microbiome.

Having elucidated the overall taxonomic composition of shrimp-associated microbiome, we further ask if there are core clades at finer phylogenetic scale that are conservative and significantly enriched across global shrimp cultures (Figure 1C, see Method). By grouping ASV ribotypes into phylogenetic clusters, we identify a power law relationship between the number of phylogenetic clusters (*N*) and their occurrence frequency (*f*): $N \sim f^{-a}$, with $a = -3.6$ (Figure 1D). This power law exponent is larger in magnitude than those observed in other systems[36], including the global ocean microbiome (where a = -1.7)[37]. This indicates a more rapid decline in the occurrence frequency as we move from the most to the least ubiquitous cluster, suggesting a highly uneven distribution of phylogenetic clusters across different shrimp farms. Nonetheless, we were able to identify five phylogenetic clades that are significantly enriched across shrimp cultures worldwide (Figure 1E). This includes three clades related to Vibrionales, represented by *Vibrio parahaemolyticus*, *Vibrio mediterranei*/*sinaloensis* and *Photobacterium damselae*, all of which are commonly known as aquaculture pathogens of fish or shrimp[38–40]. In particular, *V. parahaemolyticus* (Vp), the causative agent of acute hepatopancreatic necrosis disease (AHPND), is notorious for causing enormous economic losses to global shrimp farms every year[12]. In addition to these pathogens, a clade in the order of Alteromonadales, represented by *Pseudoalteromonas luteoviolaceae*, is also part of the core shrimp farm microbiome. Members of this clade have been shown to inhibit shrimp pathogens by producing antimicrobial compounds[41], suggesting that some of the core microbiome members may confer beneficial effects to the host.

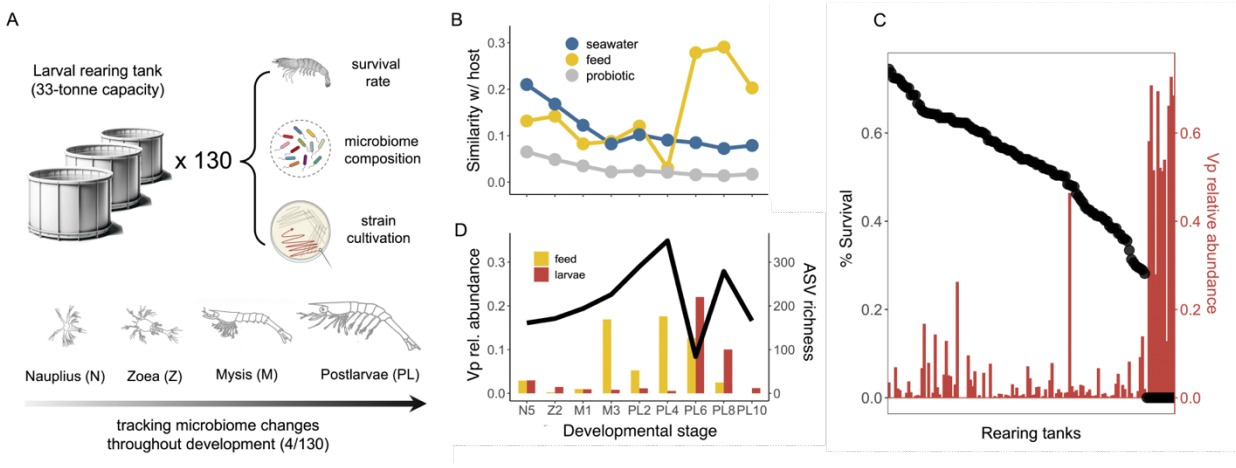

Figure 2 Microbiome characterization in a shrimp hatchery. (A) Shrimp larvae were sampled from 130 33-tonne capacity tanks where the survival rate of shrimp larvae was counted and recorded. The sampled shrimp larvae were further used for microbiome sequencing and bacterial strain isolation. For four selected

tanks, we sampled throughout the developmental stages from nauplius, zoea, mysis to postlarvae. The potential sources of microbiome such as seawater, feed (microalgae at the nauplius and zoea stage and *Artemia* at the mysis and postlarvae stage) as well as probiotics were also sampled and sequenced. (B) Similarity between shrimp larvae-associated microbiome composition and microbiome composition of potential sources. Similarity is calculated as 1- Bray-Curtis dissimilarity based on relative abundance. (C) Black dots show the ranked survival rate of shrimp larvae across 130 tanks. Red bar shows relative abundance of *V. parahaemolyticus* in the corresponding shrimp larvae-associated microbiome. (D) Black line indicates richness of shrimp larvae-associated microbiome across developmental stages. Red bar and yellow bar indicate relative abundances of *V. parahaemolyticus* across developmental stages in shrimp larvae microbiome and microbiome in the feed.

Microbiome composition predicts survival in shrimp larvae hatcheries

We first probe into the impacts of microbiome composition on productivity at the hatchery stage. Here, shrimp larvae are nurtured from their earliest nauplii (N) stage to the postlarvae (PL) phase, preparing them for transfer to large-scale farms. We tracked the microbiome succession through developmental stages (nauplii, zoea, mysis and postlarvae) in a shrimp hatchery over a period of 18 days (Figure 2A, see Method). Initially, the microbiome composition of the larvae is most similar to the surrounding seawater, especially during the early nauplii and zoea stages (Figure 2B). This resemblance gradually diminishes as the larvae mature towards the PL stage (from ~30% shared ASVs to ~8% shared ASVs, ~ 20% compositional similarity to ~ 10% compositional similarity). At this stage, the larvae-associated microbiome became most similar to that of the larval feed, which was based on the aquatic crustacean *Artemia franciscana* (~ 40% shared ASVs, ~ 30% compositional similarity; Figure 2B, Figure S2). Such opposite trends of microbiomes similarities with seawater and feed indicates a transition of microbial influences from environmental to dietary sources. Interestingly, the bacteria administered through probiotics (> 80% composed of Lactobacillaceae) appear to be unable to colonize the host, never reaching relative abundance above 1% throughout all larval developmental stages. This result highlights the need to develop new microbiome therapies based on microorganisms native to the conditions of seawater aquaculture.

To map microbiome composition to productivity, we sampled extensively across 130 tanks of 33-tonne capacity from this hatchery (Figure 2A, see Method). These tanks represent well-controlled replicate environments that undergo identical feeding regimes, therefore the

168 differences in animal survival represents a major cause of variability in productivity and feed
169 efficiency. Consistent with observations of Vp causing large scale losses in production, we found
170 in this hatchery that 10% (13 out of 130) of tanks showed signs of collapse (no harvest) and that
171 the microbiome of the larvae in these tanks was made of over 50% Vp (Figure 2C). Our
172 microbiome development data showed that the rise of Vp in the larval microbiome was a sudden
173 event, with an increase from nearly zero to ~ 20% relative abundance from day 4 to day 6 at the
174 PL stage (Figure 2D). The observed surge in Vp abundance first appeared in the feed (i.e.,
175 Artemia) before being mirrored in the larvae (Figure 2D, Figure S3), aligning with the
176 developmental phase when the larvae microbiome is primarily shaped by their feed rather than
177 the ambient seawater. This pattern indicates that *Artemia* can be a primary vector for Vp
178 transmission. Interestingly, excluding these collapsed tanks, the remaining 90% of tanks had
179 variable larvae survival rates, from ~30% to ~80% (Figure 2C). These rates were only poorly
180 correlated to Vp's abundance (Pearson's r = 0.007, $P = 0.98$), indicating that other microbiome
181 components may instead determine productivity.
182 Given this large variation in larvae survival across tanks, we asked to what extent it can
183 be attributed to differences in microbiome composition. To address this question, we employed a
184 Random Forest model to predict the survival rate of shrimp larvae based on their microbiome
185 composition. We utilized a leave-one-out (LOO) cross-validation approach, allowing us to
186 maximize the use of all available samples while ensuring an independent test set to prevent
187 overfitting. With all samples, we find that microbiome composition can predict nearly 70% of
188 the observed variation in larvae survival (LOO-$R^2 = 0.68$, Figure 3A), which is as expected since
189 a single ASV of Vp is known to be a strong indictor. We then ask how predictable the survival
190 rate is if we exclude all the collapsed tanks. Remarkably, we found that microbiome composition
191 alone still predicts nearly 50% of variation in survival rate (LOO-$R^2 = 0.46$, Figure 3B). To test
192 the robustness of this prediction, we conducted a series of random tests by reshuffling the
193 survival rates across tanks 100 times and reassessing predictability of the model. These
194 randomization tests consistently resulted in a significantly reduced predictive power (average
195 LOO-$R^2 < 0.05$, Figure 3A), demonstrating that specific microbiome compositions are indeed
196 closely linked to survival outcomes in shrimp larvae, instead of due to chance. We further show
197 that the prediction power of the model linearly increases with the number of samples (2.5% net
198 increase in LOO-R^2 per 10 samples added, linear regression $R^2 = 0.94$, $P < 0.001$, Figure 3C),

suggesting that the predictability of larvae survival by microbiome composition can be further elevated by expanding the sample space (i.e., sampling more tanks).

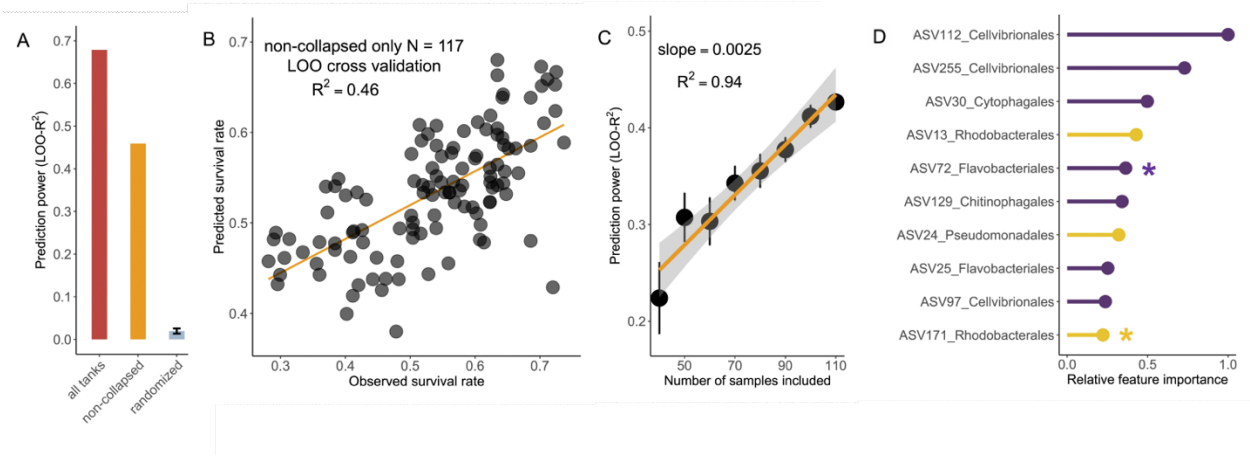

Figure 3 Microbiome composition as a strong predictor of shrimp larvae survival. (A) Statistical learning of microbiome composition to predict survival rate of shrimp larvae. Prediction power is indicated by leave-one-out cross-validation R^2. (B) Observed survival rate of shrimp larvae and that predicted by microbiome composition through leave-one-out cross-validation. The 13 tanks where a bloom of *V. parahaemolyticus* led to collapse of all shrimp larvae were excluded. (C) Prediction power of microbiome composition linearly increases as the number of samples increase. On average, increasing every 10 samples leads to 2.5% increase in prediction power. (D) Taxa showing highest feature importance in statistical learning to predict survival rate of shrimp larvae. Purple lines indicate potential primary degraders. Yellow lines indicate potential scavengers.

We further seek to better understand the taxa that are important in predicting larvae survival, especially their ecological roles and genomic features. We calculated relative feature importance of each ASVs, finding that a majority of the top important ones (7 out of top 10) are from taxonomic groups well known as polysaccharide degraders[35] (e.g., Cellvibrionales, Cytophagales and Flavobacteriales, Figure 3D). To better understand the genomic features of these important taxa, we isolated a total of 401 strains from the larvae samples, out of which 2 strains had 100% 16S rRNA gene sequence identity with the top 10 important ASVs from our Random Forest model (Figure 3D, Figure S4), one in the Rhodobacterales clade (ASV171, *Cribrihabitans* sp.) and the other in the Flavobacteriales clade (ASV72, *Psychroserpens* sp.). In addition to these two isolates, we isolated other seven strains within these two clades that are abundant (at least 1% relative abundances) in the shrimp larvae microbiome (Table S2, Figure

223 S4), suggesting that they may have fitness advantages in the shrimp larvae-associated
224 environment. We therefore leverage the genome of these isolates (we refer to as "abundant
225 isolates" thereafter) to identify genes that are specifically adaptive to a shrimp-associated niche.
226 Briefly, we filter for genes that are shared by *all* our shrimp-derived abundant isolates but are
227 rarely present (< 10%) across non-shrimp derived close relatives (i.e., in the same genus with the
228 shrimp-derived abundance isolates) (Figure 4A, see Methods for full details). This led to the
229 discovery of recent horizontal gene transfer events involving these abundant shrimp larvae-
230 derived strains in the Rhodobacterales or Flavobacteriales clades (Figure 4B, C, see
231 Supplementary Files 1,2 for sequences). Among the Rhodobacterales strains, we observed recent
232 horizontal transfer of genes involved in the biosynthesis of growth factors, such as vitamin B6[42]
233 and lysine[43] (Figure 4B, 4C, 100% nucleotide identity between homologs in *Cribrihabitans* sp.
234 ASV171 and *Phaeobacter italicus* ASV17), suggesting a potentially beneficial role of nutritional
235 supplementation for the host. Among the Flavobacteriales strains, we identified a recent
236 horizontal transfer of amidohydrolase genes, crucial for protein degradation, as well as a gene
237 encoding a TonB-dependent receptor (TBDR) (Figure 4B, 4D, 100% nucleotide identity between
238 homologs in *Psychroserpens* sp. ASV72 and *Meridianimaribacter flavus* ASV7). An exhaustive
239 search of public database further revealed homologs of this TBDR, such as an isolate genome in
240 Zhejiang Province of China (*Maribacter aurantiacus* KCTC 52409, GCF_005780245.1), as well
241 as environmental metagenomes in Shandong Province of China (ERR2094176), both showing
242 100% amino acid identity with the TBDR swept across our strains isolated from the shrimp
243 hatchery in Ecuador (Figure 4D). Both the Chinese genome and metagenome are also derived
244 from seawater aquaculture, underscoring global ecological relevance of this gene that is
245 specifically adaptive to aquaculture environment. Despite the functional versatility of TBDR,
246 previous work on marine microbiology has highlighted an important role of TBDR in
247 polysaccharide transport for Flavobacteriales[34]. Previous studies have also shown a TBDR with
248 exactly the same domain organization (i.e., a carboxypeptidase regulatory like domain along
249 with an outer membrane channel) functions in collagen degradation in coral-reef associated
250 ecosystems[44], which could inspire further biochemical studies confirming the function of this
251 TBDR.
252

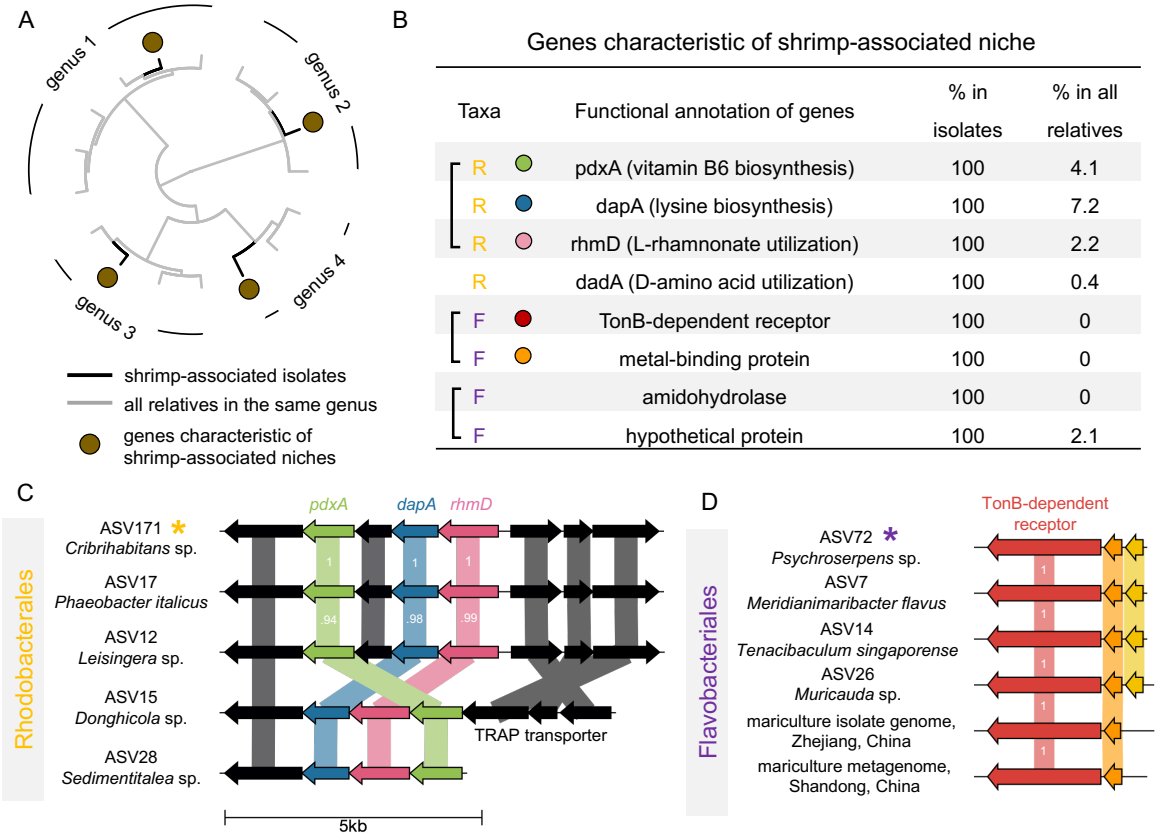

Figure 4 Horizontal transfer of genes that are specifically adaptive to a shrimp-associated niche. (A) Schematic illustration of comparative genomic analysis for identifying genes that are specifically adaptive of a shrimp-associated niche. The selected genes are shared by isolates that are abundant in the shrimp larvae microbiome but are present in none or very few (< 10%) of all the other publicly available genomes in the same taxonomic clades. Tips colored in black indicate strains isolated from the shrimp larvae samples. Tips colored in gray indicate publicly available genomes in the same genus with the shrimp-associated isolates. The brown circle indicates a gene that is only present in the genome of shrimp-associated isolates but not in any other genomes from the same genus. (B) A total of 8 genes are identified from the shrimp larvae-associated microbiome. Genes marked with a yellow 'R' are shared by all five abundant isolates in the Rhodobacterales clade. Genes marked with a purple 'F' are shared by all four abundant isolates in the Flavobacteriales clade. All these genes are rarely present in other genomes of the same genera from public databases (< 10%). The 16S rRNA sequences of all those nine isolates are 100% identical to ASVs in the sequenced shrimp larvae microbiome whose maximal relative abundances across tanks are at least 1%. Square brackets indicate that genes are next to each other in the genome. (C) Gene map of adjacent genes (*pdxA*, *dapA* and *rhmD*) experiencing horizontal transfer across five isolates from distinct genera in the order of Rhodobacterales. Numbers in the gene map indicates level of amino acid sequence identity between

homologs. (D) Gene map of two adjacent genes (TBDR and a metal-binding protein gene) experiencing recent horizontal transfer among four isolates from distinct genera in the order of Flavobacteriales isolated from shrimp larval hatchery in Ecuador, as well as their homologs in publicly available genome and metagenome derived from aquaculture in different regions of China. All the six TBDR genes have 100% identical sequence with each other.

Water quality is a stronger predictor of adult shrimp growth than the microbiome.

Having established the significant predictive power of microbiomes for animal survival in the hatchery phase, our investigation extends to the impact of microbiome on the growth of adult shrimp in shrimp grow-out ponds. In contrast to the well-controlled conditions of hatchery tanks, adult shrimp grow in dugout ponds that extend over several hectares of water – at least 100 times larger than a hatchery tank. These are complex ecosystems that farmers manage on an individual basis to achieve maximal productivity. Despite the expected heterogeneity of these systems, we asked to what extent the microbiome of individual animals could predict the individual phenotypes, in particular, body weight. To this end, we sampled 226 adult shrimps across 6 grow-out ponds, recording the body weight of each individual. We sequenced the microbiome of the hepatopancreas for all sampled shrimps, considering that it is a primary focus of Vp infections, as well as the intestine microbiome for a subset of 76 shrimps.

The hepatopancreas data revealed a relatively simple community structure dominated by three clades: Vibrionaceae, Entomoplasmatales, and Rhizobiaceae (Figure 5A). Notably, the representative sequences for Entomoplasmatales and Rhizobiaceae differed significantly from any known organisms (92% and 87% maximum sequence identity over the 16S rRNA V3-V4 region), and likely correspond to endosymbiontic bacteria. For the Entomoplasmata, we were able to assemble a genome from metagenomic reads, which was classified as Candidatus *Hepatoplasma crinochetorum*, an endosymbiont in non-insect arthropods[45]. Again, this genome showed less than 80% average nucleotide identity with any previously sequenced organism, highlighting the presence of unknown bacterial taxa within the shrimp hepatopancreas. In contrast to the hepatopancreas microbiome, the shrimp intestine microbiome exhibits significantly greater diversity (Figure 5A, ASV richness Wilcox test $P < 0.001$). Despite the distinct compositions of the intestine and hepatopancreas microbiomes, a significant correlation exists between them (Mantel test $P < 0.001$), suggesting a systemic linkage in microbial compositions across different digestive tract sections.

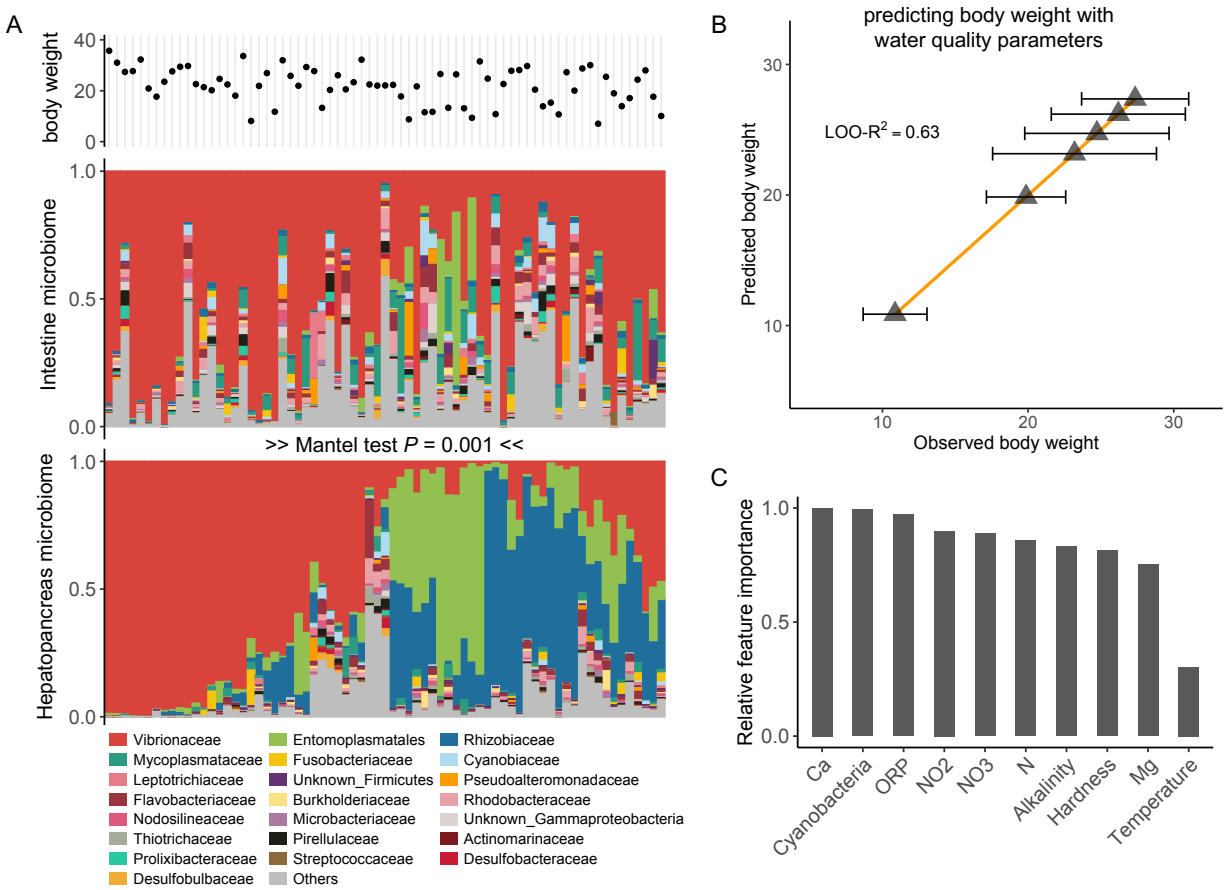

Figure 5 Adult shrimp body weight is better predicted by water quality than microbiome composition. (A) Adult shrimp individuals were sampled for body weight measurement and microbiome sequencing of hepatopancreas and the intestine. The microbiome composition in the intestine and that in the hepatopancreas are significantly correlated with each other. (B) Body weight of adult shrimp can be accurately predicted by water quality parameters. (C) Relative feature importance of water quality parameters in predicting adult shrimp body weight. ORP, oxidoreductive potential.

Based on this characterization of the hepatopancreas and intestine microbiomes, we sought to determine to what extent microbiome composition across the adult shrimp digestive tract could predict body weight in grow-out ponds. Our data revealed that the intestine and the hepatopancreas microbiome compositions were weak, although significant predictors of shrimp body weight (LOO-R^2 = 0.16 for 226 samples with hepatopancreas microbiome; LOO-R^2 = 0.18 for 76 samples with intestine microbiome). The observed R^2 were similar to the levels observed in other livestock gut ecosystems (average cross-validated $R^2 < 0.2$)[46] and significantly different

317 from those obtained with randomized data (LOO-R^2 = 0.03 ± 0.01 for 100 random tests, $P <$
318 0.001). In contrast to the high degree of unexplained variance across individuals, our analysis of
319 the metadata showed that the average body weight of individual ponds correlates strongly with
320 water quality parameters (LOO-R^2 = 0.63 for 226 samples, Figure 5B), indicating that the
321 average animal weight is primarily controlled by water chemistry. Among the key factors
322 revealed by the data alkalinity, hardness, oxidoreductive potential, nitrogen concentration, and
323 cyanobacteria density emerge as key predictors (Figure 5C). Alkalinity, indicating buffer
324 capacity of the pond water, has been shown to closely correlated productivity in shrimp farms[47].
325 Hardness is determined by the concentration of dissolved calcium and magnesium, which are
326 important for exoskeleton development and osmoregulation[47,48]. Cyanobacterial blooms can have
327 adverse effects due to toxin production and eutrophication[49]. These factors highlight the complex
328 dynamics between water chemistry, eutrophication and animal health that control productivity,
329 and thus feed-efficiency, in large-scale aquaculture facilities.
330
331 **Discussion**
332 In this study, we discovered that the microbiome composition alone accounts for 50% of shrimp
333 larvae survival, highlighting the potential of microbiome as an early indicator for host health.
334 The microbiome associated with these larvae likely affects host production by aiding in the
335 breakdown of polymeric organic materials or by synthesizing beneficial growth factors such as
336 vitamins, as suggested by our comparative genomics analysis. While prior research has
337 demonstrated that mapping the structure to the function of microbial communities can facilitate
338 functional predictions[50,51], these studies were primarily based on simplified microcosms in
339 laboratory settings. Our investigation extends these findings by establishing a robust structure-
340 function relationship within a complex, real-world ecosystem. Notably, our study reveals that
341 tree-based models, like Random Forest and Gradient Boosting, largely outperform linear models
342 in predicting functional output in this intricate real-world context (Figure S5). This suggests that
343 the functional landscape of natural microbiomes might be more complex than that synthetic
344 microbial communities with low biodiversity.
345 With respect to adult shrimps, which develop in less-controlled, complex ecosystems like
346 grow-out ponds, our analysis shows that water quality parameters are strong predictors of the
347 average shrimp weight per pond. However, there is a large degree of variance in individual body

weight, which the microbiome can only partially explain (< 20%) likely due to variable chemical and biological conditions in these environments. This suggests that hatcheries may be a better target to design microbiome-based interventions during early developmental stages, especially considering that, as shown in our data, currently used probiotics based on non-native bacteria display low engraftment rates.

Our study uncovers an understudied taxonomic and functional diversity within the shrimp-associated microbiome, opening avenues for microbiome engineering. For example, a gene cluster involved in cofactor biosynthesis, prevalent across five genera of Rhodobacterales, shows strong signs of adaptive potential. Similarly, the widespread horizontal transfer of the TBDR gene among our isolates in Ecuador, as well as those from Asian aquacultures, suggests a universal fitness advantage for various strains of Flavobacteriales in widely distributed aquaculture systems. This points to the potential of engineering polysaccharide degradation-assisting probiotics (e.g., Flavobacteriales) that are adaptable across global shrimp aquaculture settings. These findings provide promising targets for downstream genetic and biochemical studies, paving the way for innovative strategies to optimize shrimp production by microbiome engineering.

Methods

Sampling in the hatchery

Microbiome sampling was performed within commercial hatchery at the coast of Ecuador from January 27 to February 19, 2021. A total of 130 tanks with *P. vannamei* was sampled at the end of the production cycle, when the survival rate of shrimp larvae was also recorded. There were 13 tanks presenting an outbreak disease and samples of larvae of these tanks were collected. Furthermore, for a subset of 4 tanks, we sampled the shrimp larvae every two days throughout the developmental stages from nauplius (N), zoea (Z), mysis (M) and postlarvae (PL). This results in 9 timepoints spanning 18 days from N5 to PL10 stages for each tank. The shrimp larvae were fed with microalgae by the zoea stage (N5 and Z2 data points in our samples), and then *Artemia* since the mysis stage (M1 to PL10 data points in our samples). The feed microalgae, *Artemia*, together with the ambient seawater and administrated probiotics were also sampled for microbiome sequencing to track the potential source of microbiome in the shrimp larvae.

379 **Sampling in shrimp grow-out ponds**

380 A total of 224 adult shrimp were collected from six different pools from a large-scale shrimp
381 farm located in the coast of Ecuador and stored in cryovials with DMSO. The work was carried
382 out wearing KN95 masks and nitrile gloves in a sterile area to prevent contamination. Adult
383 shrimps were collected before feeding began, starting at 7 AM and finishing by 10 AM. Animals
384 were kept in buckets with aeration and water from their respective pools until their turn for
385 collection. Each animal was weighed and photographed, after which they were taken to a sterile
386 area for dissection of the intestine and the hepatopancreas. Half of the hepatopancreas was used
387 for DNA sequencing and the other half for microscopic observation, to record the degree of lipid
388 vacuole fullness and changes in color that could indicate sickness. The remaining half of the
389 hepatopancreas and intestine were immediately frozen in separate cryovials with liquid nitrogen
390 and kept at -20°C until shipment.

391 Water quality was analyzed by measuring pH, salinity, temperature, total ammonium
392 nitrogen, nitrite, nitrate, total nitrogen, phosphate, phosphorus, N:P ratio, alkalinity, calcium,
393 calcium hardness, magnesium, magnesium hardness, potassium, Ca:Mg:K ratio, hydrogen
394 sulfide (H2S), Redox potential and turbidity. Algae type quantification and identification were
395 also performed by counting in a Neubauer chamber.

396 **DNA extraction and amplicon sequencing**

397 DNA extraction of the shrimp larvae was performed with Gentra Puregene Tissue Kit (QIAGEN)
398 following the protocol provided by the manufacturer, except that bead beating was used to
399 increase the extraction efficiency for 60 seconds at 5000 rpm. In brief, the extraction procedure
400 includes cell lysis (1:1 cell lysis solution), RNA removal (4µL of RNAse A), protein
401 precipitation (250 µL Protein Precipitation Solution), DNA precipitation (100% isopropanol) and
402 purification (70% ethanol). Purified DNA was shipped to Argonne National Laboratory
403 (Lemont, IL) for amplicon sequencing on an Illumina MiSeq targeting the V4 region of 16S
404 rRNA using the 515F and 806R primers. Sequence-specific Peptide nucleic acid (PNA) clamps
405 were used to block the amplification of host-derived mitochondrial 16S sequences at V4 region.

406 Shrimp hepatopancreas and intestine samples were torn into small pieces using tweezers.
407 These fragments were further homogenized using a TissueLyser III (Qiagen) for 5 minutes at 25
408 Hz with PowerBeads Pro plates (Qiagen). Following this, DNA extraction was carried out on the
409 resulting turbid liquid using the DNAdvance Kit (Beckman Coulter), according to the

410 manufacturer's protocol. Briefly, the extraction process included cell lysis using a 1:1 cell lysis
411 solution, DNA binding to magnetic beads, and purification with 70% ethanol. Amplicon
412 sequencing were performed at Seqcenter (Pittsburg, PA) with Zymo Quick-16S Plus Library
413 Prep methods to amplify the V3/V4 region with 341F and 806R primers. Illumina
414 sequencing was performed on an Illumina NovaSeq X Plus sequencer in one or more
415 multiplexed shared-flow-cell runs, producing 2x151bp paired-end reads.
416 Amplicon sequences analyses were performed on the MIT Engaging computing cluster,
417 where we used QIIME2 v2019.2[52] to demultiplex the raw reads, and DADA2 plugin[53] to
418 generate Amplicon Sequence Variants (ASVs). Taxonomy of representative ASVs were assigned
419 with the classify-sklearn method by QIIME2 v2019.2.
420 **Strain isolation and genotyping**
421 We first isolated ~200 isolates using Difco™ Marine Broth 2216 media (BD). Colonies were
422 isolated by first growing on agar plates prepared with MBL medium[54] containing homogenized
423 Artemia as the sole carbon source, and then streaking to single colony purity on agar plates
424 prepared with 1/10 diluted Marine Broth 2216 medium (adjusted for salt) containing 0.2% N-
425 acetyl-glucosamine as the sole carbon source. A total of ~ 200 Rhodobacterales strains were
426 further isolated with a customized media recipe favoring growth of Rhodobacterales species
427 among other marine copiotrophic heterotrophs[35]. The media is MBL minimal media with 40mM
428 sodium succinate added as the carbon source. Marine Broth 2216 was spiked into the media with
429 a dilution factor of 1:80 to facilitate bacterial growth. Vp was selectively isolated by the
430 CHROMagar™ *Vibrio* (CaV) selective media, where the mauve colonies are picked for
431 genotyping. Genotyping of all isolates were performed by full-length 16S Sanger sequencing
432 (GENEWIZ at Azenta Life Sciences). 16S sequences of isolates were aligned to ASV sequences
433 from the shrimp larvae mirobiome with blastn in the blast suite v2.12.0. A maximal-likelihood
434 phylogenetic tree of representative 16S sequences was constructed with MEGA v11.
435 **Whole genome sequencing and genomic data processing**
436 Shotgun whole genome sequencing for representative isolates of Flavobacteriales and
437 Rhodobacterales were performed at the SeqCenter (Pittsburgh, PA), on an Illumina NextSeq
438 2000 platform (2x151bp pair-ended). Raw reads were trimmed to remove adaptors and low-
439 quality bases (-m pe -q 20) with Skewer v0.2.2[55]. The remaining paired reads were checked for

440 quality with FastQC v0.11.9. Quality-filtered reads were assembled into contigs with MEGAHIT
441 v1.2.9[56].
442 Nanopore long-read sequencing was combined with shotgun whole genome sequencing
443 also at the SeqCenter (Pittsburgh, PA) to close the genome of Vp. Closed genome was assembled
444 using Unicycler v0.4.9[57] by combining Illumina short reads and Nanopore long reads, resulting
445 in 2 chromosomes and 3 circular plasmids. The assembly graph in gfa format was visualized by
446 Bandage v0.8.1[58]. Coding sequences are predicted by prodigal v2.6.3, followed by functional
447 annotation by eggnog-mapper v2[59] (--go_evidence non-electronic --target_orthologs all --
448 seed_ortholog_evalue 0.001 --seed_ortholog_score 60).
449 **Metagenomic sequencing and processing**
450 We did shotgun metagenomic sequencing for 6 shrimp hepatopancreas samples (2 samples with
451 highest relative abundance of Vibrionaceae, 2 samples with highest relative abundance of
452 Entomoplasmatales and 2 samples with highest relative abundance of Rhizobiaceae). Shotgun
453 sequencing including library preparation was performed at SeqCenter (Pittsburgh, PA) on an
454 Illumina platform. We performed quality check for pair-ended fastq files with FastQC v0.11.9,
455 after which low-quality bases were trimmed by Skewer v0.2.2 [55] (-m pe -q 20). Quality-filtered
456 reads were first filtered with Bowtie2 v2.5.3 (--un-conc) to get rid of all the sequence aligned
457 with the *P. vannamei* genome (NCBI RefSeq accession GCF_003789085). All the unaligned
458 sequence was then used to assemble contigs with MEGAHIT v1.2.9. The reads that were
459 unaligned to the *P. vannamei* genome were mapped back to assembled contigs with Bowtie2[60]
460 2.5.3 to generate coverage table. Metagenomic assembled genomes (MAGs) were generated with
461 CONCOCT[61] v1.1.0 (-c 10000). Generated bins were evaluated for completeness and
462 contamination using CheckM[62] v1.1.2. Taxonomic classification of MAGs was performed by
463 GTDB-tk v2.1.0[63]. We were able to generate a MAG of *Vibrio fischeri* (completeness 100%,
464 contamination 0.5%) and a MAG of *Hepatoplasma crinochetorum* (completeness 87.5%,
465 contamination 0%) through this effort.
466 **Identifying core clades from global shrimp-associated microbiome**
467 We compiled a comprehensive global dataset of amplicon microbiome sequences, specifically
468 targeting the 16S rRNA V4 region with an approximate overlap of ~250 base pairs (bp). This
469 overlap facilitates the downstream analysis of core microbial clades. We retrieved amplicon
470 sequencing data in FASTQ format from the National Center for Biotechnology Information

471 (NCBI) database. Using DADA2[53] within QIIME2[52] (version 2019.2), we processed samples
472 from various studies, yielding 48,514 shrimp-associated amplicon sequence variants (ASVs) and
473 7,999 coastal seawater ASVs.

474 For sequence alignment, we utilized MAFFT[64] (version 7.245), followed by sequence
475 trimming to uniform lengths using TrimAl[65]. To ensure high-quality alignment of the vast
476 number of sequences, we initially conducted stringent alignments (--maxiterate 1000) focusing
477 on ASVs from the most abundant taxonomic groups, including Vibrionales, Rhodobacterales,
478 Flavobacteriales, and Alteromonadales. These aligned sequences were trimmed to uniform
479 lengths, creating a 'backbone' alignment. Subsequent sequences were integrated into this
480 backbone using MAFFT's --add function. We then merged this composite alignment with pre-
481 aligned reference sequences from the SILVA database[66] (release 132, Nr99) to construct
482 phylogenetic trees for each major clade. After a final trim with Trimal, the sequences were
483 standardized to approximately 250 bp. We constructed the phylogenetic trees using
484 FastTreeMP[67], employing the -nt, -gtr, and -gamma options, and visualized the trees with
485 iTOL[68]. Moreover, we conducted homologous clustering of all shrimp-derived sequences using
486 MMseqs2 (-s 7.5 –min-seq-id 1)[69]. This analysis supports the power-law distribution illustrated
487 in Figure 1D.

488 To identify core shrimp-associated clades, we applied three stringent criteria. First, a
489 clade must show significant enrichment (adjusted $P < 0.05$, with Bonferroni correction) of
490 shrimp-derived sequences, assuming a binomial distribution of these sequences across the entire
491 phylogenetic tree. Second, the clade must encompass shrimp-associated sequences that are
492 globally ubiquitous, present in at least 80% of sampled global shrimp cultures. Third, we
493 required complete homology within each clade, selecting only those with zero branch length,
494 indicating 100% sequence identity. Ultimately, five clades met these criteria for significant
495 enrichment, global prevalence, and stringent homology. We assigned representative taxonomy to
496 each clade based on the dominant taxonomy of the corresponding SILVA[66] reference sequences.
497 For the overarching phylogenetic analysis, we constructed a tree with all 56,513 ASV sequences
498 derived from both shrimp and coastal seawater. We labeled each tree tip according to its source.
499 The observed mean phylogenetic distance (MPD) between shrimp-associated and coastal
500 seawater sequences was calculated. To assess the significance of our findings, we randomized
501 the MPD calculations 1,000 times by reshuffling the tip labels. The Net Relatedness Index (NRI)

502 is calculated by the standardized effect size of the MPD, which is
503 $(oMPD - mean(rMPD))/sd(rMPD)$.

504 **Identifying characteristic genes of shrimp-associated niche**

505 We aligned the 16S rRNA gene sequences of our 501 isolates to the Flavobacteriales and
506 Rhodobacterales ASV sequences in the microbiome. This resulted in 4 Flavobacteriales isolates
507 and 5 Rhodobacterales isolates with 100% identical sequences with an *abundant* ASV in the
508 shrimp-larvae microbiome (at least 1% relative abundance in any sample). These nine isolates
509 spanned nine distinct genera, prompting us to retrieve all available genomes for these genera
510 from the RefSeq database. This search yielded 394 genomes for Flavobacteriales and 222 for
511 Rhodobacterales.

512 Subsequently, we conducted protein sequence clustering using MMseqs2[69] (-s 7.5 -min-
513 seq-id 0.5 -c 0.5). This analysis produced a matrix, organizing proteins by rows and genomes by
514 columns. We then refined our search to identify proteins universally present in our shrimp larvae
515 isolates yet scarcely found (< 10%) within other genomes of the same genera. This rigorous
516 filtering process highlighted eight genes meeting these specific criteria. We utilized EggNOG-
517 mapper[59] v2 for functional annotation the identified protein sequences. To visualize gene maps,
518 we employed Clincker[70], facilitating a comprehensive representation of the gene distribution.
519 Moreover, our extensive search within the public metagenomic database MGnify[71] led to the
520 discovery of a 100% homologous sequence in mariculture environments in China.

521 **Machine learning-based predictions**

522 We implemented a Random Forest (RF) model to predict outcomes like shrimp larvae survival
523 rate and adult shrimp body weight based on microbiome composition, using leave-one-out
524 (LOO) cross-validation for evaluation. In this method, the RF model, developed with the R
525 package randomForest[72], is trained on all but one sample to predict the outcome for the excluded
526 sample, a process repeated for each sample. Predictive accuracy is quantified by comparing
527 observed and predicted values through R^2 from linear regression, with robustness checked
528 against 100 randomized trials of the outcome vector. To assess how sample size affects
529 prediction power, we varied sample subsets (n= 40 to 110, in increments of 10) and conducted
530 LOO cross-validation 20 times for each size. We also compared RF performance against other
531 machine learning methods like gradient boosting and lasso regression, to gauge their efficacy (as

532 detailed in Figure S5), providing a comprehensive view of the predictive capabilities of various
533 approaches within our dataset.

534 **Other statistical analysis**

535 All other statistical analyses are implemented with R[73] v4.1.3. World map is visualized using the
536 default world map in R package ggmap[74] v3.0.0.

537

538 **Data availability**

539 All raw data generated by this study has been deposited on NCBI under BioProject
540 PRJNA1144440. This BioProject includes (1) amplicon sequencing data of shrimp larvae
541 including its associated feed and environmental samples, (2) amplicon sequencing data of
542 hepatopancreas and intestine samples of adult shrimp, (3) shotgun metagenomic sequencing data
543 of selected adult shrimp hepatopancreas samples, (4) draft genomes assembled for the 9
544 abundant isolates shown in Figure 4. The compiled global aquaculture shrimp microbiome
545 dataset is publicly available at 10.5281/zenodo.13241422 for free use. Detailed information and
546 sequences of the horizontal transferred genes are provided as supplementary file 1 and 2.

547

548 **Acknowledgement**

549 We thank Salvador Almagro-Moreno, and members of the Cordero lab, in particular Andreas
550 Sichert, Rachel Gregor, Gabriel Vercelli and Suryateja Jammalamadaka, for helpful discussions
551 and valuable assistance. X.S thanks Zhijian Huang, Shenzheng Zeng, Haipeng Guo and
552 Mengmeng Wang for their help in collecting the global microbiome datasets. X.S also thanks
553 Shiwei Wang and Victoria Marando in Laura Kiessling lab at MIT for their help on
554 understanding the interactions between aquaculture isolates.

555

556

557 1. Kankainen, M. *et al.* How to measure the economic impacts of changes in growth, feed
558 efficiency and survival in aquaculture. *Aquac. Econ. Manag.* **16**, 341–364 (2012).
559 2. Davidson, J., Bebak, J. & Mazik, P. The effects of aquaculture production noise on the
560 growth, condition factor, feed conversion, and survival of rainbow trout, Oncorhynchus
561 mykiss. *Aquaculture* **288**, 337–343 (2009).

562 3. Nedelec, S. L., Simpson, S. D., Morley, E. L., Nedelec, B. & Radford, A. N. Impacts of
563 regular and random noise on the behaviour, growth and development of larval Atlantic
564 cod (Gadus morhua). *Proc. R. Soc. B Biol. Sci.* **282**, (2015).
565 4. Fan, J. *et al.* Dynamics of the gut microbiota in developmental stages of Litopenaeus
566 vannamei reveal its association with body weight. *Sci. Reports 2019 91* **9**, 1–10 (2019).
567 5. Xiong, J. *et al.* The Underlying Ecological Processes of Gut Microbiota Among
568 Cohabitating Retarded, Overgrown and Normal Shrimp. *Microb. Ecol.* **73**, 988–999
569 (2017).
570 6. Wang, H., Huang, J., Wang, P. & Li, T. Insights into the microbiota of larval and
571 postlarval Pacific white shrimp (Penaeus vannamei) along early developmental stages: a
572 case in pond level. *Mol. Genet. Genomics* **295**, 1517–1528 (2020).
573 7. Rajeev, R., Adithya, K. K., Kiran, G. S. & Selvin, J. Healthy microbiome: a key to
574 successful and sustainable shrimp aquaculture. *Rev. Aquac.* **13**, 238–258 (2021).
575 8. Angthong, P., Chaiyapechara, S. & Rungrassamee, W. Shrimp microbiome and immune
576 development in the early life stages. *Dev. Comp. Immunol.* **147**, 104765 (2023).
577 9. Zhang, M. *et al.* Characterization of the intestinal microbiota in Pacific white shrimp,
578 Litopenaeus vannamei, fed diets with different lipid sources. *Aquaculture* **434**, 449–455
579 (2014).
580 10. Mizrahi, I., Wallace, R. J. & Moraïs, S. The rumen microbiome: balancing food security
581 and environmental impacts. *Nat. Rev. Microbiol. 2021 199* **19**, 553–566 (2021).
582 11. Sonnenschein, E. C., Jimenez, G., Castex, M. & Gram, L. The Roseobacter-Group
583 Bacterium Phaeobacter as a Safe Probiotic Solution for Aquaculture. *Appl. Environ.*
584 *Microbiol.* **87**, 1–15 (2021).
585 12. Kumar, R., Ng, T. H. & Wang, H. C. Acute hepatopancreatic necrosis disease in penaeid
586 shrimp. *Rev. Aquac.* **12**, 1867–1880 (2020).
587 13. Food and Agricultural Organization of the United Nations. *The State of World Fisheries*
588 *and Aquaculture 2022*. *The State of World Fisheries and Aquaculture 2022* (FAO, 2022).
589 doi:10.4060/CC0461EN
590 14. Bardera, G. *et al.* The importance of behaviour in improving the production of shrimp in
591 aquaculture. *Rev. Aquac.* **11**, 1104–1132 (2019).

592 15. Holt, C. C., Bass, D., Stentiford, G. D. & van der Giezen, M. Understanding the role of
593 the shrimp gut microbiome in health and disease. *J. Invertebr. Pathol.* **186**, 107387
594 (2021).
595 16. Ninawe, A. S. & Selvin, J. Probiotics in shrimp aquaculture: Avenues and challenges.
596 *Crit. Rev. Microbiol.* **35**, 43–66 (2009).
597 17. Butt, U. D. *et al.* Overview of the latest developments in the role of probiotics, prebiotics
598 and synbiotics in shrimp aquaculture. *Fish Shellfish Immunol.* **114**, 263–281 (2021).
599 18. Huang, F., Pan, L., Song, M., Tian, C. & Gao, S. Microbiota assemblages of water,
600 sediment, and intestine and their associations with environmental factors and shrimp
601 physiological health. *Appl. Microbiol. Biotechnol.* **102**, 8585–8598 (2018).
602 19. Xiong, J. *et al.* Integrating gut microbiota immaturity and disease-discriminatory taxa to
603 diagnose the initiation and severity of shrimp disease. *Environ. Microbiol.* **19**, 1490–1501
604 (2017).
605 20. Huang, L. *et al.* The bacteria from large-sized bioflocs are more associated with the
606 shrimp gut microbiota in culture system. *Aquaculture* **523**, 735159 (2020).
607 21. Huang, Z. *et al.* Microecological Koch's postulates reveal that intestinal microbiota
608 dysbiosis contributes to shrimp white feces syndrome. *Microbiome* **8**, 1–13 (2020).
609 22. Su, H. *et al.* Persistence and spatial variation of antibiotic resistance genes and bacterial
610 populations change in reared shrimp in South China. *Environ. Int.* **119**, 327–333 (2018).
611 23. Wang, Y. *et al.* Fine-scale succession patterns and assembly mechanisms of bacterial
612 community of Litopenaeus vannamei larvae across the developmental cycle. *Microbiome*
613 **8**, 1–16 (2020).
614 24. Zoqratt, M. Z. H. M., Eng, W. W. H., Thai, B. T., Austin, C. M. & Gan, H. M.
615 Microbiome analysis of Pacific white shrimp gut and rearing water from Malaysia and
616 Vietnam: implications for aquaculture research and management. *PeerJ* **6**, e5826 (2018).
617 25. Zeng, S. *et al.* Dissimilarity of microbial diversity of pond water, shrimp intestine and
618 sediment in Aquamimicry system. *AMB Express* **10**, 1–11 (2020).
619 26. Elizondo-González, R. *et al.* Changes on the intestinal bacterial community of white
620 shrimp Penaeus vannamei fed with green seaweeds. *J. Appl. Phycol.* **32**, 2061–2070
621 (2020).

622 27. Pilotto, M. R. *et al.* Exploring the Impact of the Biofloc Rearing System and an Oral
623 WSSV Challenge on the Intestinal Bacteriome of Litopenaeus vannamei. *Microorg. 2018,*
624 *Vol. 6, Page 83* **6**, 83 (2018).
625 28. Wang, K. *et al.* Bacterial biogeography in the coastal waters of northern Zhejiang, East
626 China Sea is highly controlled by spatially structured environmental gradients. *Environ.*
627 *Microbiol.* **17**, 3898–3913 (2015).
628 29. Zäncker, B., Cunliffe, M. & Engel, A. Bacterial community composition in the sea surface
629 microlayer off the peruvian coast. *Front. Microbiol.* **9**, 415419 (2018).
630 30. Mai, Y. zhan, Lai, Z. ni, Li, X. hui, Peng, S. yao & Wang, C. Structural and functional
631 shifts of bacterioplanktonic communities associated with spatiotemporal gradients in river
632 outlets of the subtropical Pearl River Estuary, South China. *Mar. Pollut. Bull.* **136**, 309–
633 321 (2018).
634 31. Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F. & Cordero, O. X. Microbial interactions
635 lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* **7**, (2016).
636 32. Enke, T. N. *et al.* Modular Assembly of Polysaccharide-Degrading Marine Microbial
637 Communities. *Curr. Biol.* **29**, 1528-1535.e6 (2019).
638 33. Imhoff, J. F. "Vibrionales". *Bergey's Manual® Syst. Bacteriol.* 491–555 (2005).
639 doi:10.1007/0-387-28022-7_11
640 34. Buchan, A., LeCleir, G. R., Gulvik, C. A. & González, J. M. Master recyclers: features
641 and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* **12**,
642 686–698 (2014).
643 35. Gralka, M., Pollak, S. & Cordero, O. X. Genome content predicts the carbon catabolic
644 preferences of heterotrophic bacteria. *Nat. Microbiol. 2023 810* **8**, 1799–1808 (2023).
645 36. Pang, T. Y. & Maslov, S. Universal distribution of component frequencies in biological
646 and technological systems. *Proc. Natl. Acad. Sci. U. S. A.* (2013).
647 doi:10.1073/pnas.1217795110
648 37. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science (80-.*
649 *).* (2015). doi:10.1126/science.1261359
650 38. Prado, P. *et al.* Presence of Vibrio mediterranei associated to major mortality in stabled
651 individuals of Pinna nobilis L. *Aquaculture* **519**, 734899 (2020).

652 39. Del, M. *et al.* Isolation and characterization of infectious Vibrio sinaloensis strainsfrom
653 the Pacific shrimp Litopenaeus vannamei (Decapoda: Penaeidae). *Rev. Biol. Trop.* **60**,
654 567–576 (2012).
655 40. Rivas, A. J., Lemos, M. L. & Osorio, C. R. Photobacterium damselae subsp. Damselae, a
656 bacterium pathogenic for marine animals and humans. *Front. Microbiol.* **4**, 63518 (2013).
657 41. RADJASA, O. K. *et al.* Antibacterial Property of a Coral-Associated Bacterium
658 Pseudoalteromonas luteoviolacea Against Shrimp Pathogenic Vibrio harveyi (In Vitro
659 Study). *HAYATI J. Biosci.* **12**, 77–81 (2005).
660 42. Tramonti, A. *et al.* Knowns and Unknowns of Vitamin B 6 Metabolism in Escherichia
661 coli . *EcoSal Plus* **9**, (2021).
662 43. Muduli, S., Karmakar, S. & Mishra, S. The coordinated action of the enzymes in the L-
663 lysine biosynthetic pathway and how to inhibit it for antibiotic targets. *Biochim. Biophys.*
664 *Acta - Gen. Subj.* **1867**, 130320 (2023).
665 44. Bhattacharya, S., Choudhury, J. D., Gachhui, R. & Mukherjee, J. A new collagenase
666 enzyme of the marine sponge pathogen Pseudoalteromonas agarivorans NW4327 is
667 uniquely linked with a TonB dependent receptor. *Int. J. Biol. Macromol.* **109**, 1140–1146
668 (2018).
669 45. Leclercq, S., Dittmer, J., Bouchon, D. & Cordaux, R. Phylogenomics of "Candidatus
670 Hepatoplasma crinochetorum," a Lineage of Mollicutes Associated with Noninsect
671 Arthropods. *Genome Biol. Evol.* **6**, 407–415 (2014).
672 46. John Wallace, R. *et al.* A heritable subset of the core rumen microbiome dictates dairy
673 cow productivity and emissions. *Sci. Adv.* **5**, 8391–8394 (2019).
674 47. Boyd, C. E., Tucker, C. S. & Somridhivej, B. Alkalinity and Hardness: Critical but
675 Elusive Concepts in Aquaculture. *J. World Aquac. Soc.* **47**, 6–41 (2016).
676 48. Galkanda-Arachchige, H. S. C., Roy, L. A. & Davis, D. A. The effects of magnesium
677 concentration in low-salinity water on growth of Pacific white shrimp (Litopenaeus
678 vannamei). *Aquac. Res.* **52**, 589–597 (2021).
679 49. Alonso-Rodríguez, R. & Páez-Osuna, F. Nutrients, phytoplankton and harmful algal
680 blooms in shrimp ponds: a review with special reference to the situation in the Gulf of
681 California. *Aquaculture* **219**, 317–336 (2003).

682 50. Oliveira, R. A. *et al.* Statistical design of a synthetic microbiome that clears a multi-drug
683 resistant gut pathogen. *bioRxiv* 2024.02.28.582635 (2024).
684 doi:10.1101/2024.02.28.582635
685 51. Skwara, A. *et al.* Statistically learning the functional landscape of microbial communities.
686 *Nat. Ecol. Evol. 2023 711* **7**, 1823–1833 (2023).
687 52. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data
688 science using QIIME 2. *Nat. Biotechnol. 2019 378* **37**, 852–857 (2019).
689 53. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon
690 data. *Nat. Methods 2016 137* **13**, 581–583 (2016).
691 54. Shan, X., Szabo, R. E. & Cordero, O. X. Mutation-induced infections of phage-plasmids.
692 *Nat. Commun. 2023 141* **14**, 1–10 (2023).
693 55. Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: A fast and accurate adapter trimmer for
694 next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 1–12 (2014).
695 56. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-
696 node solution for large and complex metagenomics assembly via succinct de Bruijn graph.
697 *Bioinformatics* **31**, 1674–1676 (2015).
698 57. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial
699 genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**,
700 e1005595 (2017).
701 58. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of
702 de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
703 59. Cantalapiedra, C. P., Herņandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J.
704 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
705 Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
706 60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
707 *2012 94* **9**, 357–359 (2012).
708 61. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat.*
709 *Methods 2014 1111* **11**, 1144–1146 (2014).
710 62. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
711 assessing the quality of microbial genomes recovered from isolates, single cells, and
712 metagenomes. *Genome Res.* **25**, 1043 (2015).

713 63. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory
714 friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316
715 (2022).
716 64. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
717 Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
718 65. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
719 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
720 (2009).
721 66. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data
722 processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
723 67. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-
724 Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
725 68. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
726 tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
727 69. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for
728 the analysis of massive data sets. *Nat. Biotechnol. 2017 3511* **35**, 1026–1028 (2017).
729 70. Gilchrist, C. L. M. & Chooi, Y. H. clinker & clustermap.js: automatic generation of gene
730 cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
731 71. Mitchell, A. L. *et al.* MGnify: the microbiome analysis resource in 2020. *Nucleic Acids*
732 *Res.* **48**, D570–D578 (2020).
733 72. Liaw, A. & Wiener, M. Classification and Regression by randomForest. **2**, (2002).
734 73. R Development Core Team, R. *R: A Language and Environment for Statistical*
735 *Computing*. *R Foundation for Statistical Computing* (2011). doi:10.1007/978-3-540-
736 74686-7
737 74. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2.
738