RESEARCH

In silico prediction of potential indigenous microbial biomarkers in *Penaeus vannamei* identified through meta-analysis and genome-scale metabolic modelling

Neelakantan Thulasi Devika[1], Vinaya Kumar Katneni[1*], Ashok Kumar Jangam[1], Panjan Nathamuni Suganya[1], Mudagandur Shashi Shekhar[1] and Karingalakkandy Poochirian Jithendran[2]

Abstract

Background Understanding the microbiome is crucial as it contributes to the metabolic health of the host and, upon dysbiosis, may influence disease development. With the recent surge in high-throughput sequencing technology, the availability of microbial genomic data has increased dramatically. Amplicon sequence-based analyses majorly profile microbial abundance and determine taxonomic markers. Furthermore, the availability of genome sequences for various microbial organisms has prompted the integration of genome-scale metabolic modelling that provides insights into the metabolic interactions influencing host health. However, the analysis from a single study may not be consistent, necessitating a meta-analysis.

Results We conducted a meta-analysis and integrated with constraint-based metabolic modelling approach, focusing on the microbiome of pacific white shrimp *Penaeus vannamei*, an extensively cultured marine candidate species. Meta-analysis revealed that *Acinetobacter* and *Alteromonas* are significant indicators of "health" and "disease" specific taxonomic biomarkers, respectively. Further, we enumerated metabolic interactions among the taxonomic biomarkers by applying a constraint-based approach to the community metabolic models (4416 pairs). Under different nutrient environments, a constraint-based flux simulation identified five beneficial species: *Acinetobacter* spWCHA55, *Acinetobacter tandoii* SE63, *Bifidobacterium pseudolongum* 49 D6, *Brevundimonas pondensis* LVF1, and *Lutibacter profundi* LP1 mediating parasitic interactions majorly under sucrose environment in the pairwise community. The study also reports the healthy biomarkers that can co-exist and have functionally dependent relationships to maintain a healthy state in the host.

Conclusions Toward this, we collected and re-analysed the amplicon sequence data of *P. vannamei* (encompassing 117 healthy and 142 disease datasets). By capturing the taxonomic biomarkers and modelling the metabolic interaction between them, our study provides a valuable resource, a first-of-its-kind analysis in aquaculture scenario toward a sustainable shrimp farming.

Keywords *Penaeus vannamei*, Meta-analysis, 16S amplicon sequence analysis, Metagenomics, Genome-scale metabolic modelling, Flux balance analysis, Flux variability analysis, Microbial biomarker

*Correspondence:
Vinaya Kumar Katneni
Vinaya.Katneni@icar.gov.in
Full list of author information is available at the end of the article

Introduction

With the advent of high-throughput sequencing, metagenome datasets have become increasingly accessible. The culture-independent metagenomic approach has

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 2 of 13

facilitated extensive analysis of microbiome data for preventive and diagnostic purposes. Even in the aquaculture sector, microbial consortiums are crucial for improving sustainability and productivity of aquatic organisms [5, 43]. Several factors threaten the sustainability and yield of aquaculture species, one of which relates to the infection caused by pathogenic microorganisms [29]. Therefore, understanding the essential and significant microbial consortia would simplify the prediction of disease incidences. In aquatic animals, amplicon sequence based approach has been used to identify microbiomes, their composition, and functions [4, 30, 33, 54].

The current study focuses on *Penaeus vannamei*, one of the world's most widely cultured and traded marine candidate species. *P. vannamei* farming has increased rapidly to meet the growing edible high-quality protein demand. Despite this, *P. vannamei* farming is constantly hampered by abiotic and biotic factors [58]. One of the challenges faced by *P. vannamei* farming relates to emerging diseases caused by bacteria, viruses, and fungi affecting the sustainability of shrimp [1, 31, 32]. Antibiotics are used to prevent the growth of microbial infections, but the spread of microbial antibiotic resistance poses a hazard to human health [5]. Alternatively, microbes also promote the host's growth by acting as a probiotic to circumvent antimicrobials [50]. Thus, studying the potential, resident, and beneficial microbes is crucial for improving host health. To date, several amplicon sequence-based studies on *P. vannamei* have characterized the microbial communities in different habitats [6], developmental stages [6], disease conditions [59], and diet [47].

It is important to note that the conclusions drawn from a single study may not accurately reflect the microbial communities within the host [15]. Therefore, we performed a meta-analysis to gain a more robust and consistent understanding of health vs. disease states in *P. vannamei*. Previous microbiome-based meta-analyses targeted different ontogenic stages of the healthy and disease shrimp predicting taxonomic markers [56]. Another study revealed the role of different biological factors like environment and different life stages in shaping the microbiota [9]. Furthermore, these studies might not encapsulate the interrelationships between microbes for maintaining a stable microbial community. Therefore, it is necessary to profile and explore the microbial interactions that can mediate beneficial host functions. Constraint-based metabolic modelling with genome-scale metabolic models (GSM) has been a widely used approach to study microbial interactions between species in a community [13, 28, 49]. Moreover, reconstructing genome-scale metabolic networks from amplicon sequence data has also been employed to infer resource competition and metabolic cooperation potential in microbial consortia [36, 60]. Hence, employing, flux balance analysis on the genome-scale metabolic models allow assessment of nutritional requirements, interactions, and metabolic exchanges in a microbial consortia under diverse nutrient environments [3, 60]. Moreover, a microbial community's ability to coexist in different habitats depends on the cross-feeding of metabolites [57]. Natural metabolic products such as short-chain fatty acids have been used to control pathogenic bacterial growth [39, 48]. Likewise, using short-chain fatty acids in aquaculture also provides a growth advantage to *Penaeus vannamei* by suppressing the growth of pathogenic species [21].

Most microbiome-based studies have identified biomarkers associated with healthy and disease shrimp; however, this study is the first to integrate 16S amplicon sequence data and genome-scale metabolic modelling in aquaculture scenario to capture the metabolic interaction between the taxonomic biomarker species. The goal is to identify the most efficient taxonomic biomarkers that, in turn, should also limit the growth of pathogenic biomarker species. In summary, applying multi-pronged approach, we systematically characterized the natural indigenous species eventually coexist in the host or need to be artificially introduced to the host towards a sustainable health management.

Methods

Public data collection

The amplicon sequence data used for this meta-analysis were retrieved from the NCBI SRA database (Downloaded as on September, 2021). The studies collected were related to *P. vannamei* species associated with a disease, considering only the host tissue or intestine samples. Initially, the search yielded 13 studies with 838 datasets [Additional file 1: Table S1]. However, several studies were not considered due to lack of habitat information. The disease dataset included in this meta-analysis are infected with White feces disease events (WFD), White Spot Syndrome Virus (WSSV) and Acute Hepatopancreatic Necrotic Disease (AHPND). The available studies differed in the sub-regions of the 16S rRNA gene, such as V3-V4, V4, V1-V2, V2 mix, and V3 mix sequenced with different sequencing platforms. Due to the limited number of studies with proper metadata, a uniform sequencing technology or a common hypervariable region could not be implemented. Finally, six studies totaling 259 datasets were considered [Additional file 1: Table S2].

Devika *et al. Environmental Microbiome* (2023) 18:2

Page 3 of 13

Microbiome analysis

Initially, each study was processed separately, with single-end and paired-end sequences analyzed using the Quantitative Insights Into Microbial Ecology (QIIME2) pipeline [17]. The read quality assessment was conducted with the DADA2 plugin to weed out low-quality forward and reverse reads that did not meet the quality threshold of 20. For DADA2, the trimming and truncation parameters unique to each study were provided. The datasets were removed when the read quality dropped below the threshold. The filtered reads from each study were aligned and classified using the SILVA database [55]. Finally, Amplicon Sequence Variants (ASV), a higher resolution version of the Operational Taxonomic Unit (OTU) was generated. Next, we combined the dataset wherein individual studies were merged. We filtered out datasets that fall below 2000 reads from the combined dataset and removed low abundance features (removed features that appeared in less than ten datasets). In the end, 241 datasets from five studies were analysed. Next, the representative sequences obtained from the combined pre-processed data were assigned taxonomy using the pre-trained classifier SILVA (Silva release 138; 99% OTUs full-length sequences). Also, taxon archaea and eukaryota were not included as part of our analysis. In addition, chloroplast, mitochondria, and unassigned genus were removed from the feature table.

Alpha- and beta-diversity

The within-dataset difference was measured using alpha diversity in individual and combined dataset. For individual studies and combined dataset Shannon, Chao1, Observed Features, and Simpson were estimated to determine the community richness and evenness between the datasets. With the help of the vegan package in R (ggplot), beta-diversity was quantified using Bray–Curtis dissimilarity and visualized through nonmetric multidimensional scaling (NMDS). The beta-diversity was also measured with unweighted unifrac distance metrics. The statistical difference between the healthy and disease states was computed with the Kruskal–Wallis for alpha diversity metrics and Permutational Multivariate Analysis of Variance (PERMANOVA) for both the beta-diversity metrics. A p-value less than 0.05 was considered as significant.

Identification of biomarker

A taxonomic biomarker that can differentiate between a healthy and a disease state was found using the linear discriminant analysis Effect Size (LEfSe) method [46]. ASV tables derived from combined datasets were filtered for unassigned/uncultured genera before subjecting to LEfSe analysis. An effect size (LDA score) of > 2.0 with statistically significant p-value (< 0.05) was used for biomarker identification. A taxon was more accurately distinguished between its respective healthy and disease states when it has a large effect size and is statistically significant in a set of datasets.

Computing co-occurrence

The relative abundance data corresponding to healthy and disease biomarker identified with LEfSe were subjected to **Hi**gher-**Or**der **Co**-occurrence (HiOrCo) patterns in microbial samples for computing the highly co-occurring species [36]. The algorithm begins by considering the pairs of species that co-occur in samples and proceeds to a group of larger sizes and generate to a default of 100 communities in each size. The algorithm evaluates such that the species co-occur in at least twice in 10 datasets and should pass the FDR-correction test.

Mapping ASV to prokaryotic database with complete genome

The 16 s rRNA sequences corresponding to the healthy/disease biomarkers were retrieved from the combined dataset. All the retrieved sequences were mapped to the complete bacterial genomes downloaded from NCBI RefSeq (23,764 complete genomes as on November 11, 2021). We carried out standalone BLASTP with an e-value cut off 10^{-6} and with a percent identity and query coverage of 97%. Species with hits matching the above criteria were selected for the downstream analysis such as for building metabolic models.

Reconstruction of genome-scale metabolic model

The protein sequence that mapped to their closed reference genome were retrieved based on percent identity and query coverage [Additional file 1: Table S3]. A genome scale metabolic model was built separately for each of the species using CarveMe v1.2.2 [35]. Using top-down approach CarveMe constructs organism specific models after removal of reactions and metabolites that are absent in the target organism. A total of 123 individual models were reconstructed gap-filled, grown in M9 minimal media with glucose as the carbon source, assigned with biomass components specific to Gram-positive and Gram-negative bacteria [Additional file 1: Table S4]. The individual species assigned as disease biomarker were then merged pairwise with all possible pairs with the species assigned as healthy biomarker resulting in 4416 pairs.

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 4 of 13

Simulation of GSM using flux balance and flux variability analysis

CobraPy [16], a constraint-based modelling package in python with *cplex* solver for solving the optimization problems was used for performing simulations. The in silico growth prediction of individual and pairwise models were performed with highly reliable flux balance analysis (FBA). FBA is a constraint-based modelling method that estimates the fluxes of reactions in a metabolic network to capture the metabolic capabilities of an organism [40]. FBA solves linear system of equations derived from the stoichiometric matrix $S_{m \times r}$, expressed mathematically as follows:

$$Objective :$$
$$\text{Max } v_{bio}$$
$$s.t \quad S.v = 0$$
$$Lj < vj < Uj$$

where m is the number of metabolites and r is the number of reactions, v represents the flux through all reactions, Lj and Uj are the lower and upper bound flux of each reaction j. For the in silico growth prediction, 17 different carbon sources [Additional file 1: Table S5] were used and allowed the uptake of single carbon source at a time with maximization of biomass as the objective function. The in silico growth simulation on different nutrient environments were performed by setting the lower bound of each of the carbon sources to -10 mmol/gDw/h and setting the lower bound of other carbon sources to 0. The lower bounds of amino acid exchange reactions and other essential components was set as -1 mmol/gDW/h [Additional file 1: Table S4]. The in silico growth rate of single and paired species are compared and observed for a 10 percent increase or decrease in growth of healthy/disease species in the presence of another [24].

Flux variability analysis (FVA) was used to predict the acetate production/consumption in the community under diverse nutrient environments. FVA computes the maximum and minimum flux range through each reaction with biomass reaction constrained to the maximum growth rate achieved [38].
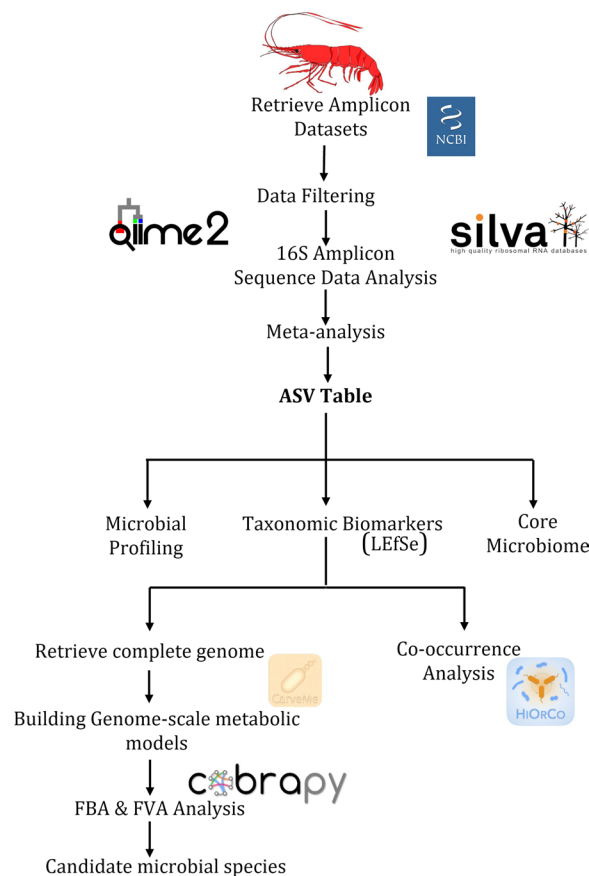
$$Maximize, Minimize \ vj$$
$$S.t. \ S.v = 0$$
$$v_j^{min} \leq v \geq v_j^{max}$$

where v represents the maximum and minimum flux through each reaction j. Since acetate is considered as an indicator which suppresses the growth of pathogenic species, we conducted FVA on acetate reaction. In a pairwise community, acetate was considered secreted by an organism, if the flux of acetate exchange reaction was positive, and consumed by the organism, if the flux of acetate exchange reaction was negative.

Results

The six studies included data spanning 259 datasets (117 Healthy & 142 Disease). The inclusion criteria for this meta-analysis were amplicon sequence-based studies on *P. vannamei* in healthy and disease states. For identifying a potential taxonomic biomarker, a single study may not be sufficient; hence we performed a meta-analysis that could serve as a representative and biologically meaningful biomarker. Additionally, we conducted constraint-based metabolic modelling approach to capture the metabolic capabilities of taxonomic biomarkers under different nutritional environments and inferred the interspecies interactions. Figure 1 depicts the workflow that outlines the key steps followed in this study.

Fig. 1 Outline of the study integrating meta-analysis and genome-scale metabolic modelling for identifying potential microbial species

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 5 of 13

Diversity analysis for individual datasets

The data from each study were systematically processed and reanalyzed using the QIIME2 pipeline. Diversity measures, such as alpha and beta indices, were calculated based on taxonomic profiles obtained from each study. Based on the Shannon index [Additional file 2: Figure S1], healthy microbial datasets exhibited significantly higher richness and evenness (Kruskal–Wallis test, $p < 0.05$, four out of six studies) than disease microbial datasets [Additional file 1: Table S6]. Similarly, the beta diversity index unweighted unifrac distance tested with PERMANOVA revealed a significant difference ($p < 0.05$, four out of six studies) between the healthy and disease datasets [Additional file 1: Table S6]. We also assessed beta diversity metrics based on Bray–Curtis distance and visualized with NMDS plot [Additional file 2: Figure S2]. Briefly, four out of six studies demonstrated typical patterns of richness and diversity among the healthy and disease dataset. The phylum-level taxonomic profile exhibited a high abundance of *Proteobacteria* in five studies, while *Firmicutes* was abundant in one study.
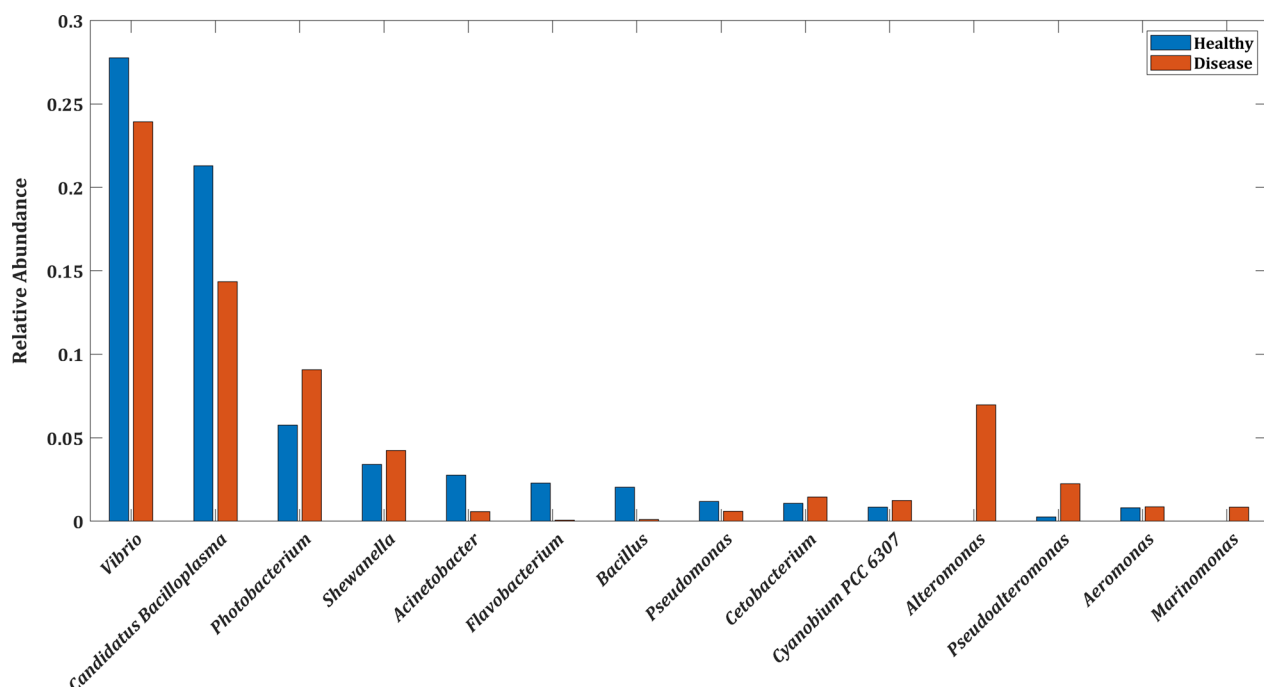
Diversity analysis for combined datasets

The datasets from six individual studies were combined and datasets with reads count below 2000 was removed, resulting in 241 datasets (one study was excluded due to low read counts). We computed alpha and beta diversity measures to investigate similarities within and between datasets. The alpha diversity metrics such as Shannon, Simpson, Chao1, ACE, and observed features computed for the combined dataset were presented in Additional file 1: Table S7. Shannon and Simpson index values revealed no significant difference (Kruskal–Wallis test, $p > 0.05$) between the healthy and disease states. With beta diversity metrics namely unweighted unifrac distance and Bray–Curtis distance, consistent results were observed, revealing a significant difference (PERMANOVA test, $p < 0.05$) between healthy and disease states. The Bray–Curtis distance which considered both species presence/absence and abundance was visualized through an NMDS plot [Additional file 2: Figure S3].

Microbial abundance at phylum and genus level in healthy and disease states

We examined the top five dominant phylum and genera by computing the mean relative abundance across healthy and disease datasets. A phylum or genera with a mean abundance of ≥ 0.01 was considered abundant. At the phylum level, *Proteobacteria* (62% in healthy vs. 72% in disease state) and *Firmicutes* (30% in healthy vs. 20% in disease state) are the major representatives followed by *Bacteroidota*, *Actinobacteriota*, and *Cyanobacteria* [Additional file 2: Figure S4] in both the states based on the mean relative abundance. Four genera namely, *Vibrio*, *Candidatus Bacilloplasma*, *Photobacterium*, and *Shewanella* dominated both healthy and disease states in

Fig. 2 Boxplot depicting the mean relative abundance of top 10 genus-level taxa in healthy and disease states

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 6 of 13

descending order of magnitude. However, the mean relative abundance of these genera was marginally different in both states, as shown in Fig. 2. Interestingly, the genus *Alteromonas* appeared only in the disease state, and the genus *Acinetobacter*, on the other hand, was more abundant in the healthy state. The disease state, however, showed a lower abundance of *Acinetobacter*.

Identifying the core microbiome

We then analyzed the core microbiome (refers to a taxon that must appear across dataset) to identify the key genera associated with healthy and disease states. The genus *Vibrio* of *Proteobacteria* phylum was universally prevalent across healthy and disease state. However, owing to heterogeneous nature of the dataset, our study considered the microbes that are present in at least 50% of the dataset while looking across the healthy and disease state. Among the phyla, *Firmicutes, Proteobacteria, Actinobacteriota, Bacteroidota, Cyanobacteria*, and *Planctomycetota* were highly abundant in both the states, making them core phyla in *P. vannamei* regardless of the health state of the host. The core microbiome analysis at the genus level identified 24 and 19 genera at a prevalence of 50% in the healthy and disease state, respectively. Among these genera, 17 were common in both the states and comprise the core genera associated with *P. vannamei* irrespective of the health state of the host. Out of the 43 core genera present in both states, seven (*Gemmobacter, Chryseomicrobium, Stenotrophomonas, LD29, Sva0081_sediment_group, PLTA13, SZB30*) and two (*PeM15, Pseudoalteromonas*) genera were prevalent exclusively in the healthy and disease state, respectively [Additional file 2: Figure S5]. Interestingly, the genus *Acinetobacter* was present at a sample prevalence of 80% in healthy state, while 65% in disease state.

Microbial biomarker detection in healthy and disease states

LEfSe analysis was carried out on the combined dataset to estimate whether there was a significant difference in the relative abundance between the healthy/disease state. LEfSe identified 32 beneficial and 73 disease genera [Additional file 1: Table S8] as prospective biomarkers with an effect size greater than two and a *p*-value < 0.05 [Additional file 2: Figure S6]. The top five healthy biomarker belonged to the phylum *Firmicutes*, and *Proteobacterium*, including genus, *Candidatus Bacilloplasma, Acinetobacter, Exiguobacterium, Lactobacillus*, and *Shimia*. On the other hand, the top five disease biomarker belonged to phylum *Proteobacteria* including genera *Alteromonas, Photobacterium, Pseudoalteromonas, Halomonas*, and *Marinomonas*. The mean relative abundance corresponding to the dominant

taxonomic biomarkers in the healthy and disease state is depicted in Fig. 3. Further, it was interesting to note that most of the disease biomarkers identified belonged to the phylum *Proteobacteria*.
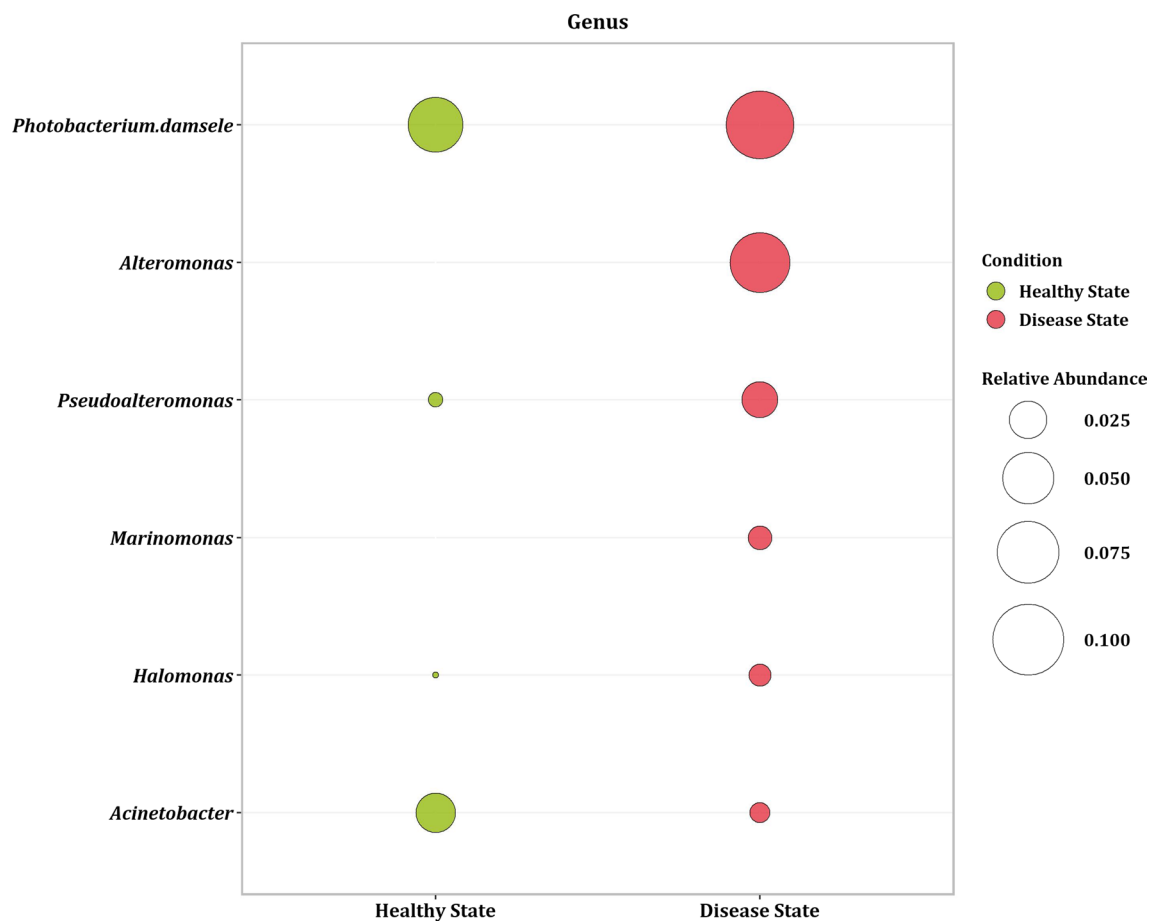
Microbial biomarker co-occurrence

The significantly co-occurring microbial biomarkers essential in maintaining host health were determined using HiOrCo. The HiOrCo algorithm computed the co-occurring communities found together in at least ten datasets more often than expected by chance in healthy [Additional file 1: Table S9a] and disease state [Additional file 1: Table S9b]. Overall, the healthy biomarkers were co-occurring to a consortia size of 15 (genera up to 15). The genera, namely *PLTA13* (identified as *Thiohalobacter thiocyanaticus* strain Hrh1 based on BLAST similarity analysis—92%), *Chryseomicrobium, Lactobacillus, Bifidobacterium, Phormidesmis, Rubrobacter, Lutibacter, Exiguobacterium, Legionella,* and *Acinetobacter* were co-occurring in a consortium of maximum size and can be considered as the representative consortium indicative of health. Whereas, the disease-specific biomarkers co-occur in up to 26 genera. The genera, namely *Marinomonas, Pseudoalteromonas Candidatus, Aestuariibacter, Alcanivorax, Seonamhaeicola, PeM15* (identified as *Geodermatophilus ruber* strain DSM 45317 based on BLAST similarity analysis—94%), and *CL500-3* (identified as *Mucisphaera calidilacus* strain Pan265 based on BLAST similarity analysis -88%) were some dominant co-occurring disease biomarker communities, which should be steered effectively to control the pathogenic state of the host.

Pairwise interactions between the taxonomic biomarkers

The 16 s rRNA sequences corresponding to the healthy/disease biomarkers identified with LeFSe were mapped to their closest reference genomes (prokaryotic database of bacterial species with a complete genome availability was created). Mapping with BLASTP based on percent identity and query coverage (97% similarity) retrieved 64 healthy (corresponding to 26 different genera) and 69 disease species (corresponding to 19 different genera).

A genome-scale metabolic model was built with CarveMe for the retrieved species and subjected to in silico phenotypic growth predictions on the single and pairwise (4416 pairs) species in a minimal media supplemented with 17 different carbon sources. Each of the paired models was analysed for an increase in growth rate for the healthy species (10% increase in in silico growth compared to the single) with a concomitant growth limit for the disease species (10% decrease in silico growth rate compared to single) in each of the environments.

Fig. 3 Plot depicting the mean relative abundance of dominant biomarker at genus level in healthy and disease state. Size of the circle represent the mean relative abundance
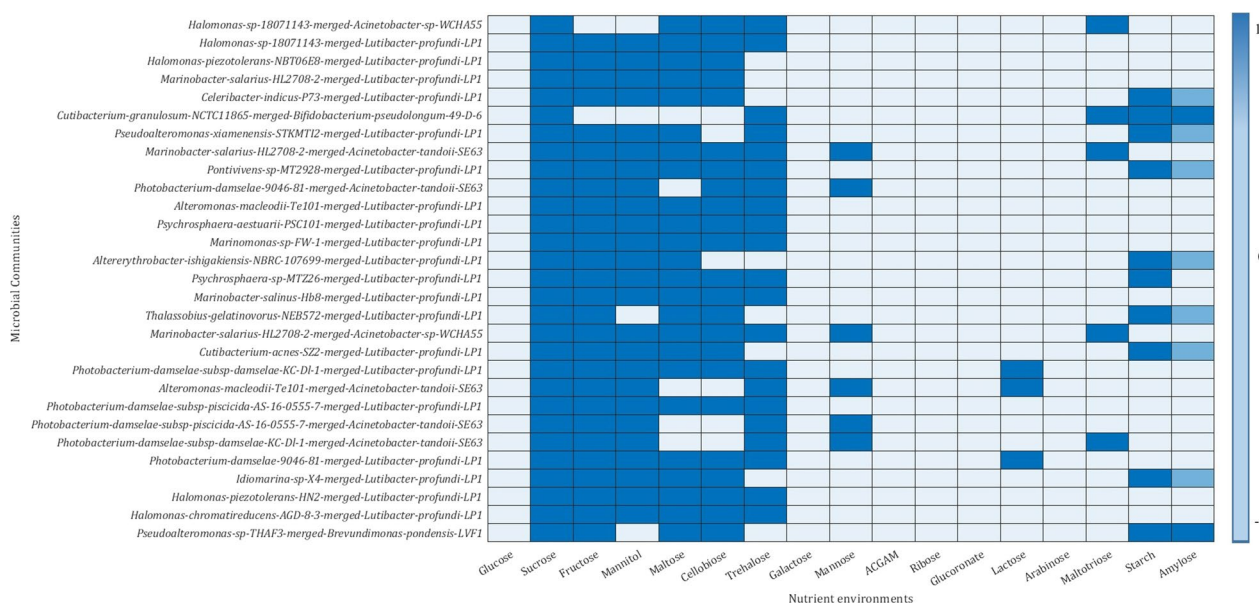
The healthy 69 species exhibited simulated growth capability in all the nutrient environments used in this study. On the other hand, the disease species exhibited less growth preference in sucrose, mannitol, and fructose environment. Further, out of the 4416 paired communities generated, only 794 pairs showed significant growth change and growth limitation on the healthy and disease species, respectively, in at least one of the nutrient environments. A total of 47 healthy species comprising 12 different genera form part of this 794 pairwise communities that limited the disease counterpart. These 47 healthy species limited the growth of 63 out of the 64 disease species except the strain *Synechococcus* CBW 1004. These growth limitations were majorly (711 paired communities) observed under sucrose environment, followed by trehalose (103 pairwise communities) and maltose (75 pairwise communities). Moreover, 29 out of the 794 pairwise microbial biomarker communities limited the growth of disease species in at least five of the nutrient environments [Fig. 4]. The five healthy species, namely

Acinetobacter sp WCHA55, *Acinetobacter tandoii* SE63, *Bifidobacterium pseudolongum* 49 D 6, *Brevundimonas pondensis* LVF1, and *Lutibacter profundi* LP1 form part of 29 communities which limited the growth of 22 disease species in at least five of the nutrient environments. Among the five strains, *Lutibacter profundi* LP1 limited the growth of maximum number of disease species (20 different species). Altogether, these results short-listed four genera indigenous in *P. vannamei*, limiting the growth of disease species that naturally resides on the host.

Cross-feeding of acetate in the pairwise communities

Flux variability analysis performed on the pairwise communities indicated that the production/consumption of acetate depends on the species with which it was paired and the nutrient environments.

The single and paired healthy species, namely, *Acinetobacter* sp WCHA55, *Acinetobacter tandoii* SE63, *Brevundimonas pondensis* LVF1, and *Lutibacter profundi*

Fig. 4 Heatmap depicting the growth suppression of disease species in a pairwise community. The dark blue represents the environment in which healthy species growth rate is 10% increase and growth rate of disease species is 10% decreased in pairwise community model
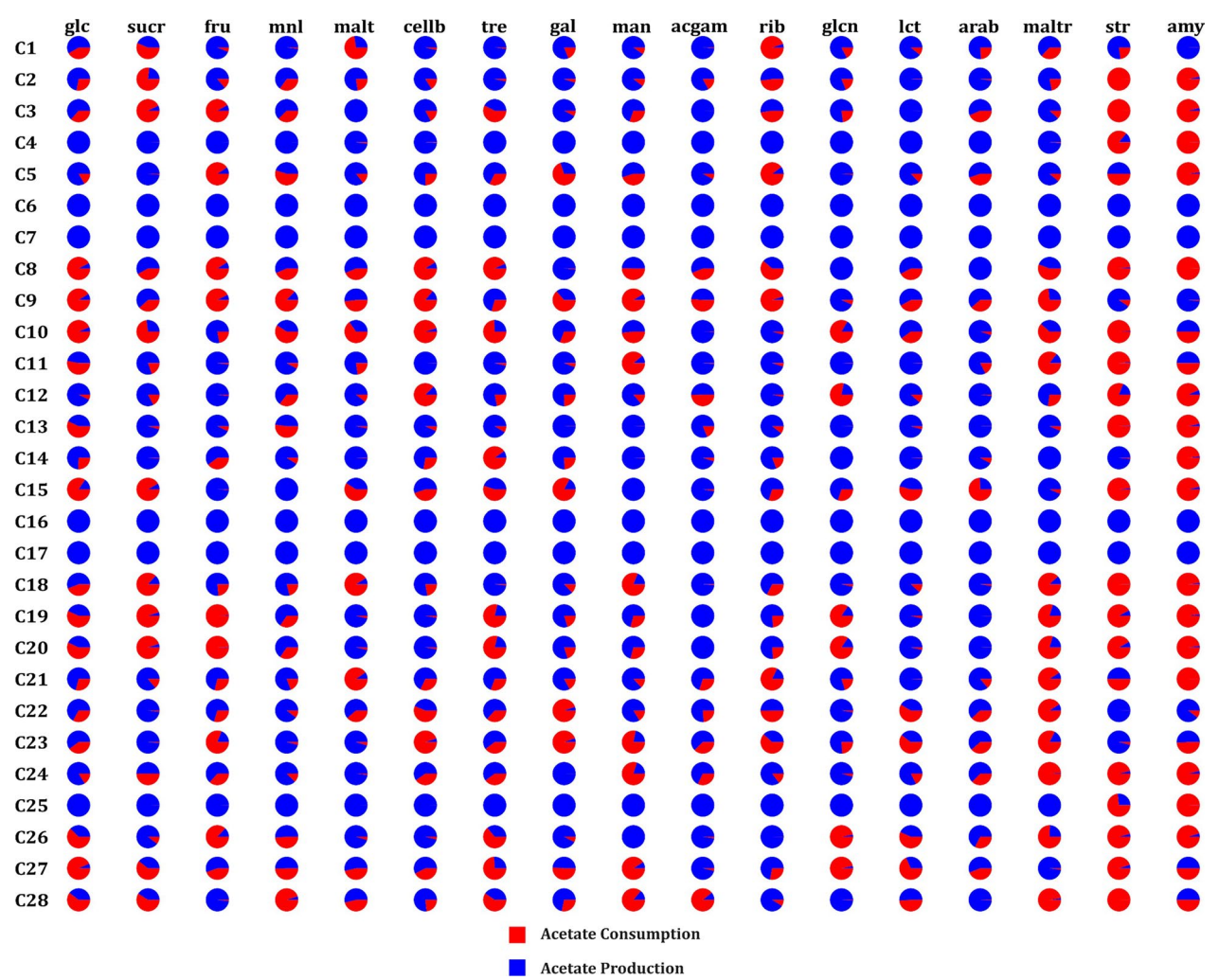
LP1 produced acetate under all nutrient environments. In the case of disease species, *Marinobacter salarius* HL27082 lost the ability to produce acetate upon paired with *Acinetobacter tandoii* SE63 and with *Acinetobacter* sp WCHA55 under all environments. While upon pairing with *Lutibacter profundi* LP1, *Marinobacter salaries* HL2708_2 could produce acetate in all the nutrient environments. Among the nutrient environments, glucose, fructose, and mannitol were associated with the production of acetate (as the acetate exchange reaction carry a positive flux) by the healthy species (Fig. 5). On the other hand, the disease species were consumers of acetate (as the acetate exchange reaction carry a negative flux) under these environments. Conversely, disease species produced acetate under maltotriose, amylose, and starch environments (as the acetate exchange reaction carry a positive flux).

Discussion

Naturally existing microbial communities perform several vital tasks, such as degradation of organic matter, recycling the nutrients, and controlling the development of microbial infections [5]. Hence, a fundamental understanding of the health and disease-associated microbial communities for maintaining host health is required. Numerous 16S amplicon sequence-based studies have attempted to unravel the role of microbes, but findings drawn from one study may not decipher the full spectrum of microbes responsible for health/disease states. Hence, analyzing multiple studies in a meta-analysis has the potential to make inferences about common features

associated with host health across different habitats. Although, several works were reported on human health perspective, very few studies are conducted on aquatic species, but did not extend to understand the metabolic interactions among them [9, 56]). Identifying taxonomic biomarkers in aquaculture is not new,however, most research focuses on single-study-based taxonomic biomarkers [27, 44], implying the need for a much comprehensive study. In this context, we performed a meta-analysis to identify the core taxonomic biomarkers that drive differences in the health/disease states in *P. vannamei* and determine the inter-species interactions exhibited by the microbial biomarkers under different nutrient environments. As an augmentation to the sustainable disease management, this study aims to identify beneficial indigenous species in *P. vannamei* essential for maintaining a good healthy state and controlling the growth of pathogenic species.

The study identified *Acinetobacter* as one of the prevalent and dominant genera in the healthy state. *Acinetobacter* was previously detected and reported in healthy shrimp to aid in suppressing harmful bacteria [18]. Furthermore, the leading genera revealed in this study's healthy state, namely *Acinetobacter* and *Candidatus Bacilloplasma,* were reported to act as a central hub connecting the significant bacterial population in the network of healthy shrimp hepatopancreas [51]. Further, *Exiguobacterium* genera in the healthy state have improved growth and survival in *P. vannamei* and provided probiotic advantages [7, 11]. *Lactobacillus* members identified in this study are known probiotics and

Fig. 5 Pie chart depicting the production (denoted in blue) and consumption (denoted in red) of acetate under different nutrient environments in the shortlisted microbial communities. (Name corresponding to the pairwise microbial communities are provided in Additional file 1: Table S10)

have been extensively studied for their abilities to maintain animal health [50]. *Shimia sp.*, another genus identified in healthy state in this study, has been reported to produce beneficial metabolites and degrade toxins [14]. Despite the fact that *Vibrio* was found in both states supporting previous studies [12, 22], it is considered floras opportunistic bacteria and cause disease when animal is under stress. Overall, the healthy genera captured in our study are comparable with earlier studies, confirming the correctness of our approach.

We also found several notable biomarkers in the disease state, including *Alteromonas*, *Photobacterium*, *Marinomonas* and *Pseudoalteromonas* which have previously been reported to cause WFD [1, 2, 26, 59]. Interestingly, these WFD related biomarkers are also associated with AHPND [8], a disease included in this study. In sum, the disease taxonomic biomarkers identified in this study

correlate with previous studies, indicating their potential value in diagnosing disease onset.

Identifying the core microbes in the healthy and disease state reveals the essential microbes for multiple facets of microbiome-associated host functions. Several meta-analysis studies identified the core microbiome based on species presence in at least 10% of the samples [41]; however, we used a sample prevalence of 50% in our study. *Chryseomicrobium* and *Stenotrophomonas*, the healthy core genera identified in this study, are known for generating bioactive (anti-microbial and anti-enzyme) chemicals and, further, the glucosidase inhibitors produced by these genera give an additional advantage to survive in a competitive environment [42]. *Gemmobacter*, another core microbe identified in this study, was also reported to be present in the healthy conditions, though the functional status of this organism is not discussed

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 10 of 13

[27]. Overall, the core genera catalogued in this study are associated with specific functional roles contributing to the growth of *P. vannamei.*

Microbes naturally coexist as a community rather than as individuals [10]. Hence, dysbiosis in the healthy consortia can favor the growth of opportunistic pathogens. Despite this, we lack a comprehensive understanding of the microbial communities that coexist and cooperate to help prevent bacterial infections. In our study, we cataloged the healthy and disease taxonomic biomarker, which frequently co-occurs and is critical for developing shrimp's particular facets. The knowledge of co-existing healthy genera catalogued in this study [Additional file 1: Table S9a] might be potentially applied for the management of aquaculture environment towards sustainable disease control.

Despite the advantages of amplicon-based sequencing, which includes microbial profiling and biomarker determination, it does not highlight the metabolic interactions [20]. With the vast number of genome sequences available, a whole genome-based approach was imperative to explore the metabolic interactions between healthy and disease states. This enabled the possibility of integrating genome-based metabolic modelling with an amplicon sequencing approach. Such genome-based approach provide insights into the taxonomic and functional interactions among the microbial communities [45, 53]. Hence, we further extended and enhanced our understanding of the identified taxonomic biomarkers by simulating the growth of biomarker species in diverse nutrient environments. Simulation through FBA allows the identification of healthy taxonomic biomarkers, which limit the growth of disease biomarkers. The constraint-based approach revealed five species: *Acinetobacter* sp WCHA55, *Acinetobacter tandoii* SE63, *Bifidobacterium pseudolongum* 49 D 6, *Brevundimonas pondensis* LVF1, and *Lutibacter profundi* LP1 limited the growth of the pathogen in a higher number of environments. The healthy biomarker species shortlisted with the metabolic modelling approach help steer the proliferation of pathogenic microorganisms and thereby control disease progression. The species from the genus *Acinetobacter* have been previously reported as a potential probiotic and believed to be safe for human health and could help replace antibiotics by controlling the pathogenic microorganisms and improve water quality in aquaculture ponds [18]. Similarly, the genera *Bifidobacteria* is another widely used probiotic in humans [23] and have also been reported to be present in healthy shrimp. Another shortlisted species, *Brevundimonas pondensis,* appears in various habitats, including aquatic environments [19], and has also been reported to be used for water pollutant treatment [34]. It should

be interesting to observe the role of *Lutibacter profundi* LP1, which limited the growth of several disease-specific taxonomic markers used in our study. Belonging to the family *Flavobacteriaceae*, the members of this family are widely used in food and dairy products and are also associated with the degradation of organic matter in marine, seawater, and freshwater [52]. In summary, genome-scale simulation identified potential and novel candidate species that can be utilized as supplements in *P. vannamei* farming.

Analyzing the flux variability of pairwise communities revealed the role of *Acinetobacter* sp. in acetate production. The ability of *Acinetobacter* to produce acetate in all pairwise communities might explain its potential role as a health indicator. It is well established that organic acids are used in food preservation, as feed additives, and to control pathogens [39]. Although many beneficial or probiotic species are used in aquaculture, the lack of consistency and performance under different conditions is a major concern. Since microbial species interactions vary in different nutrient environments, it is crucial to capture the favorable environment that controls disease species' growth. As a result, our findings add value by demonstrating that sucrose facilitates parasitic interactions (i.e., limits disease growth), which is consistent with previous research that highlighted the role of sucrose and beneficial species in improved water quality, *P. vannamei* growth, and microbial composition.

Conclusion

For the first time, 16S amplicon sequence data and genome-based metabolic modelling were combined for aquaculture application to find native biomarkers that may be best utilized to build probiotic formulations that leverage beneficial microorganisms for *P. vannamei* farming while limiting pathogen growth. Amplicon sequence-based analysis combined with metabolic modeling provided insights regarding metabolic interactions and the impact of nutrient environments, and finally, shortlist potential beneficial species to expedite experiments. We have employed this combination strategy for probiotic application in aquaculture, which was applied previously to human gut microbiota [25, 37, 60]. The key genera identified in this study could be used to prepare a formulation to replenish the healthy microbial consortia. In order to foster healthy shrimp farming, these essential genera can reduce the severity of disease conditions by removing opportunistic pathogenic bacteria and enhancing residential beneficial associations. Moreover, even though shotgun metagenomics can give a better taxonomic resolution and functional profile, we restricted our

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 11 of 13

analysis to the amplicon dataset due to the limited availability of shotgun data with *P. vannamei*. Nevertheless, we can further enhance our understanding with curated genome scale metabolic models, followed by an experimental validation for the direct application of beneficial microbes in aquaculture farming.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40793-022-00458-6.

Additional file 1. Table S1. Initial Dataset collected for the study. **Table S2**. Final Dataset considered for the analysis. **Table S3**. List of healthy and disease assigned species obtained from BLASTP analysis and subjected to FBA. **Table S4**. Minimal media components used for building Genome scale metabolic models. **Table S5**. List of carbon sources used for simulation of models. **Table S6**. Alpha and Beta diversity indices for Individual datasets. **Table S7**. Alpha and Beta diversity indices for combined datasets. **Table S8**. List of Healthy and Disease biomarkers identified with LEfSe. **Table S9a**. List of Healthy Co-occurring genera identified with HiOrco. **Table S9b**. Co-ocurring genera identified with HiOrco in disease state. **Table S10**. List of short-listed communities with growth advantage of healthy species in at least five nutrient environments

Additional file 2. Figure S1. Comparison of Shannon diversity between healthy and disease state. The black line indicate the median value for each state. **Figure S2**. NMDS plot computed on the individual dataset based on Bray-Curtis distance for comparing microbial composition between healthy and disease dataset. **Figure S3**. NMDS plot on the combined dataset based on Bray Curtis to compare the microbial composition between healthy and disease state. **Figure S4**. Comparison of top 5 dominant phyla present in healthy and disease state. **Figure S5**. Venn diagram showing the number of genus-level taxa shared, and unique among healthy and disease state at 50% sample prevalence. **Figure S6**. LEfSe analysis depicting genus level biomarkers with a LDA sore > 2 at P < 0.05. The disease biomarkers are depicted with a negative score (red) and a positive LDA score (green) for healthy biomarkers.

Declarations

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

Author details
¹Nutrition Genetics and Biotechnology Division, Indian Council of Agricultural Research - Central Institute of Brackishwater Aquaculture, Chennai, India. ²Aquatic Animal Health and Environment Division, Indian Council of Agricultural Research - Central Institute of Brackishwater Aquaculture, Chennai, India.

References

1. Alfiansah YR, Peters S, Harder J, Hassenrück C, Gärdes A. Structure and co-occurrence patterns of bacterial communities associated with white faeces disease outbreaks in Pacific white-leg shrimp *Penaeus vannamei* aquaculture. Sci Rep. 2020;10(1):1–14. https://doi.org/10.1038/s41598-020-68891-6.
2. Amin M, Kumala RRC, Mukti AT, Lamid M, Nindarwi DD. Metagenomic profiles of core and signature bacteria in the guts of white shrimp, *Litopenaeus vannamei*, with different growth rates. Aquaculture. 2022;550(December 2021):737849. https://doi.org/10.1016/j.aquaculture.2021.737849.
3. Ankrah NYD, Bernstein DB, Biggs M, Carey M, Engevik M, García-Jiménez B, Lakshmanan M, Pacheco AR, Sulheim S, Medlock GL. Enhancing microbiome research through genome-scale metabolic modeling. MSystems. 2021. https://doi.org/10.1128/msystems.00599-21.
4. Baldo L, Riera JL, Tooming-Klunderud A, Albà MM, Salzburger W. Gut microbiota dynamics during dietary shift in eastern African cichlid fishes. PLoS ONE. 2015;10(5):1–23. https://doi.org/10.1371/journal.pone.0127462.
5. Bentzon-Tilia M, Sonnenschein EC, Gram L. Monitoring and managing microbes in aquaculture—towards a sustainable industry. Microb Biotechnol. 2016;9(5):576–84. https://doi.org/10.1111/1751-7915.12392.
6. Cheng Y, Ge C, Li W, Yao H. The intestinal bacterial community and functional potential of *Litopenaeus vannamei* in the coastal areas of china. Microorganisms. 2021. https://doi.org/10.3390/microorganisms9091793.
7. Cong M, Jiang Q, Xu X, Huang L, Su Y, Yan Q. The complete genome sequence of *Exiguobacterium arabatum* W-01 reveals potential probiotic functions. MicrobiologyOpen. 2017. https://doi.org/10.1002/mbo3.496.
8. Cornejo-Granados F, Lopez-Zavala AA, Gallardo-Becerra L, Mendoza-Vargas A, Sánchez F, Vichido R, Brieba LG, Viana MT, Sotelo-Mundo RR, Ochoa-Leyva A. Microbiome of Pacific Whiteleg shrimp reveals differential bacterial community composition between wild, aquacultured and AHPND/EMS outbreak conditions. Sci Rep. 2017;7(1):1–15. https://doi.org/10.1038/s41598-017-11805-w.
9. Cornejo-Granados F, Gallardo-Becerra L, Leonardo-Reza M, Ochoa-Romo JP, Ochoa-Leyva A. A meta-analysis reveals the environmental and host factors shaping the structure and function of the shrimp microbiota. PeerJ. 2018;6:e5382. https://doi.org/10.7717/peerj.5382.
10. de Bruijn I, Liu Y, Wiegertjes GF, Raaijmakers JM. Exploring fish microbial communities to mitigate emerging diseases in aquaculture. FEMS Microbiol Ecol. 2018;94(1):1–12. https://doi.org/10.1093/femsec/fix161.
11. de Mello Júnior CC, Shizuo Owatari M, Dias Schleder D, Angel Poli M, Ramon Rodrigues Gelsleichter Y, Postai M, Elize Krüger K, de Carvalho FG, Priscila Pereira Silva B, Letícia Teixeira B, Leclercq G, Luiz Pedreira Mouriño J, Alejandro Vinatea L. Identification and characterization of microorganisms potentially beneficial for intensive cultivation of *Penaeus vannamei* under biofloc conditions: highlighting *Exiguobacterium acetylicum*. Aquac Res. 2021;52(8):3628–38. https://doi.org/10.1111/are.15207.
12. de Souza Valente C, Wan AHL. *Vibrio* and major commercially important vibriosis diseases in decapod crustaceans. J Invertebr Pathol. 2021;181(May 2020):107527. https://doi.org/10.1016/j.jip.2020.107527.
13. Devika NT, Jangam AK, Katneni VK, Patil PK, Nathamuni S, Shekhar MS. In silico prediction of novel probiotic species limiting pathogenic *Vibrio* growth using constraint-based genome scale metabolic modeling. Front Cell Infect Microbiol. 2021;11(September):1–10. https://doi.org/10.3389/fcimb.2021.752477.
14. Duan Y, Wang Y, Liu Q, Dong H, Li H, Xiong D, Zhang J. Changes in the intestine microbial, digestion and immunity of *Litopenaeus vannamei* in response to dietary resistant starch. Sci Rep. 2019;9(1):1–10. https://doi.org/10.1038/s41598-019-42939-8.
15. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun. 2017. https://doi.org/10.1038/s41467-017-01973-8.
16. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-based reconstruction and analysis for python. BMC Syst Biol. 2013. https://doi.org/10.1186/1752-0509-7-74.
17. Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolek T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, Dillon MR, Bolyen E, Caporaso JG, Knight R. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 12 of 13

publicly available data. Curr Protoc Bioinform. 2020;70(1):1–46. https://doi.org/10.1002/cpbi.100.

18. Farzanfar A. The use of probiotics in shrimp aquaculture. FEMS Immunol Med Microbiol. 2006;48(2):149–58. https://doi.org/10.1111/j.1574-695X.2006.00116.x.

19. Friedrich I, Klassen A, Neubauer H, Schneider D, Hertel R, Daniel R. Living in a puddle of mud: isolation and characterization of two novel Caulobacteraceae Strains *Brevundimonas pondensis* sp. Nov. and *Brevundimonas goettingensis* sp. nov. Appl Microbiol. 2021;1(1):38–59. https://doi.org/10.3390/applmicrobiol1010005.

20. Frioux C, Singh D, Korcsmaros T, Hildebrand F. From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes. Comput Struct Biotechnol J. 2020;18:1722–34. https://doi.org/10.1016/j.csbj.2020.06.028.

21. Gao S, Pan L, Huang F, Song M, Tian C, Zhang M. Metagenomic insights into the structure and function of intestinal microbiota of the farmed Pacific white shrimp (*Litopenaeus vannamei*). Aquaculture. 2019;499(April 2018):109–18. https://doi.org/10.1016/j.aquaculture.2018.09.026.

22. He Z, Pan L, Zhang M, Zhang M, Huang F, Gao S. Metagenomic comparison of structure and function of microbial community between water, effluent and shrimp intestine of higher place *Litopenaeus vannamei* ponds. J Appl Microbiol. 2020;129(2):243–55. https://doi.org/10.1111/jam.14610.

23. Heinken A, Thiele I. Systematic prediction of health—relevant human-microbial co-metabolism through a computational framework. Gut Microbes. 2015;6(2):85–92. https://doi.org/10.1080/19490976.2015.1023494.

24. Heinken A, Thiele I. Anoxic conditions promote species-specific mutualism between gut microbes in silico. Appl Environ Microbiol. 2015;81(12):4049–61. https://doi.org/10.1128/AEM.00101-15.

25. Henson MA, Orazi G, Phalak P, O'Toole GA. Metabolic modeling of cystic fibrosis airway communities predicts mechanisms of pathogen dominance. MSystems. 2019;4(2):1–20. https://doi.org/10.1128/msystems.00026-19.

26. Holt CC, Bass D, Stentiford GD, van der Giezen M. Understanding the role of the shrimp gut microbiome in health and disease. J Invertebr Pathol. 2021;186(April):107387. https://doi.org/10.1016/j.jip.2020.107387.

27. Huang Z, Zeng S, Xiong J, Hou D, Zhou R, Xing C, Wei D, Deng X, Yu L, Wang H, Deng Z, Weng S, Kriengkrai S, Ning D, Zhou J, He J. Microecological Koch's postulates reveal that intestinal microbiota dysbiosis contributes to shrimp white feces syndrome. Microbiome. 2020;8(1):1–13. https://doi.org/10.1186/s40168-020-00802-3.

28. Ibrahim M, Raman K. Two-species community design of lactic acid bacteria for optimal production of lactate. Comput Struct Biotechnol J. 2021;19:6039–49. https://doi.org/10.1016/j.csbj.2021.11.009.

29. Lafferty KD, Harvell CD, Conrad JM, Friedman CS, Kent ML, Kuris AM, Powell EN, Rondeau D, Saksida SM. Infectious diseases affect marine fisheries and aquaculture economics. Ann Rev Mar Sci. 2015;7:471–96. https://doi.org/10.1146/annurev-marine-010814-015646.

30. Larsen AM, Mohammed HH, Arias CR. Characterization of the gut microbiota of three commercially valuable warmwater fish species. J Appl Microbiol. 2014;116(6):1396–404. https://doi.org/10.1111/jam.12475.

31. Lee D, Yu Y-B, Choi J-H, Jo A-H, Hong S-M, Kang J-C, Kim J-H. Viral shrimp diseases listed by the OIE: a review. Viruses. 2022;14(3):585. https://doi.org/10.3390/v14030585.

32. Li J, Jiang H, Li L, Zhang X, Chen J. The effect of disease and season to hepatopancreas and intestinal mycobiota of *Litopenaeus vannamei*. Front Microbiol. 2019;10(APR):1–13. https://doi.org/10.3389/fmicb.2019.00889.

33. Li T, Long M, Gatesoupe FJ, Zhang Q, Li A, Gong X. Comparative analysis of the intestinal bacterial communities in different species of carp by pyrosequencing. Microb Ecol. 2015;69(1):25–36. https://doi.org/10.1007/s00248-014-0480-8.

34. Liu L, Feng Y, Wei L, Zong Z. Genome-based taxonomy of *Brevundimonas* with reporting *Brevundimonas huaxiensis* sp. nov. Microbiol Spectr. 2021. https://doi.org/10.1128/spectrum.00111-21.

35. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 2018;46(15):7542–53. https://doi.org/10.1093/nar/gky537.

36. Machado D, Maistrenko OM, Andrejev S, Kim Y, Bork P, Patil KR, Patil KR. Polarization of microbial communities between competitive and

37. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol. 2017;35(1):81–9. https://doi.org/10.1038/nbt.3703.

38. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5(4):264–76. https://doi.org/10.1016/j.ymben.2003.09.002.

39. Ng WK, Koh CB. The utilization and mode of action of organic acids in the feeds of cultured aquatic animals. Rev Aquac. 2017;9(4):342–68. https://doi.org/10.1111/raq.12141.

40. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245–8. https://doi.org/10.1038/nbt.1614.What.

41. Palanisamy V, Gajendiran V, Mani K. Meta-analysis to identify the core microbiome in diverse wastewater. Int J Environ Sci Technol. 2022;19(6):5079–96. https://doi.org/10.1007/s13762-021-03349-4.

42. Pandey S, Sree A, Dash SS, Sethi DP, Chowdhury L. Diversity of marine bacteria producing beta-glucosidase inhibitors. Microb Cell Fact. 2013;12(1):1–7. https://doi.org/10.1186/1475-2859-12-35.

43. Rajeev R, Adithya KK, Kiran GS, Selvin J. Healthy microbiome: a key to successful and sustainable shrimp aquaculture. Rev Aquac. 2021;13(1):238–58. https://doi.org/10.1111/raq.12471.

44. Reyes G, Betancourt I, Andrade B, Panchana F, Román R, Sorroza L, Trujillo LE, Bayot B. Microbiome of *Penaeus vannamei* larvae and potential biomarkers associated with high and low survival in shrimp hatchery tanks affected by acute hepatopancreatic necrosis disease. Front Microbiol. 2022;13(May):1–14. https://doi.org/10.3389/fmicb.2022.838640.

45. Sandhu M, Paul AT, Jha PN. Metagenomic analysis for taxonomic and functional potential of polyaromatic hydrocarbons (PAHs) and polychlorinated biphenyl (PCB) degrading bacterial communities in steel industrial soil. PLoS ONE. 2022;17(4 April):1–22. https://doi.org/10.1371/journal.pone.0266808.

46. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60. https://doi.org/10.1186/gb-2011-12-6-r60.

47. Shao J, Jiang K, Wang L. *Litopenaeus vannamei* fed diets with different replacement levels of fish meal by fish silage: a molecular approach on intestinal microbiota. Aquac Nutr. 2019;25(3):721–8. https://doi.org/10.1111/anu.12900.

48. Sun Y, O'Riordan MXD. Regulation of bacterial pathogenesis by intestinal short-chain fatty acids. Adv Appl Microbiol. 2013;85:93–118. https://doi.org/10.1016/B978-0-12-407672-3.00003-4.

49. Thommes M, Wang T, Zhao Q, Paschalidis IC, Segrè D. Designing metabolic division of labor in microbial communities. MSystems. 2019;4(2):1–21. https://doi.org/10.1128/msystems.00263-18.

50. Thompson J, Weaver MA, Lupatsch I, Shields RJ, Plummer S, Coates CJ, Rowley AF. Antagonistic activity of lactic acid bacteria against pathogenic vibrios and their potential use as probiotics in shrimp (*Penaeus vannamei*) culture. Front Mar Sci. 2022;9(February):1–9. https://doi.org/10.3389/fmars.2022.807989.

51. Wang H, Wan X, Xie G, Dong X, Wang X, Huang J. Insights into the histopathology and microbiome of Pacific white shrimp, *Penaeus vannamei*, suffering from white feces syndrome. Aquaculture. 2020;527(May):735447. https://doi.org/10.1016/j.aquaculture.2020.735447.

52. Wissuwa J, Bauer SLM, Steen IH, Stokke R. Complete genome sequence of *Lutibacter profundi* LP1T isolated from an Arctic deep-sea hydrothermal vent system. Stand Genom Sci. 2017;12(1):1–11. https://doi.org/10.1186/s40793-016-0219-x.

53. Wu X, Rensing C, Han D, Xiao K-Q, Dai Y, Tang Z, Liesack W, Peng J, Cui Z, Zhang F. Genome-resolved metagenomics reveals distinct phosphorus acquisition strategies between soil microbiomes. MSystems. 2022. https://doi.org/10.1128/msystems.01107-21.

54. Ye L, Amberg J, Chapman D, Gaikowski M, Liu WT. Fish gut microbiota analysis differentiates physiology and behavior of invasive Asian carp and indigenous American fish. ISME J. 2014;8(3):541–51. https://doi.org/10.1038/ismej.2013.181.

55. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "all-species Living Tree

Devika *et al. Environmental Microbiome*   (2023) 18:2

Page 13 of 13

Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014;42(D1):643–8. https://doi.org/10.1093/nar/gkt1209.

56. Yu W, Wu JH, Zhang J, Yang W, Chen J, Xiong J. A meta-analysis reveals universal gut bacterial signatures for diagnosing the incidence of shrimp disease. FEMS Microbiol Ecol. 2018;94(10):1–10. https://doi.org/10.1093/femsec/fiy147.

57. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci USA. 2015;112(20):6449–54. https://doi.org/10.1073/pnas.1421834112.

58. Zhang W, Zhu Z, Chen J, Qiu Q, Xiong J. Quantifying the importance of abiotic and biotic factors governing the succession of gut microbiota over shrimp ontogeny. Front Microbiol. 2021. https://doi.org/10.3389/fmicb.2021.752750.

59. Zheng Y, Yu M, Liu J, Qiao Y, Wang L, Li Z, Zhang XH, Yu M. Bacterial community associated with healthy and diseased Pacific white shrimp (*Litopenaeus vannamei*) larvae and rearing water across different growth stages. Front Microbiol. 2017;8(JUL):1–11. https://doi.org/10.3389/fmicb.2017.01362.

60. Zorrilla F, Buric F, Patil KR, Zelezniak A. MetaGEM: reconstruction of genome scale metabolic models directly from metagenomes. Nucleic Acids Res. 2021. https://doi.org/10.1093/nar/gkab815.

Publisher's Note