Microbial diversity in selected Indonesian marine organisms

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium
(Dr. rer. nat.)

seit 1558

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät der
Friedrich-Schiller Universität Jena

von Diplom-Bioinformatiker Philipp Hennersdorf

geboren am 12.11.1986 in Berlin-Mitte

Gutachter:

Prof. Dr. Hans Peter Saluz, Friedrich Schiller Universität, Jena

Dr. Gianni Panagiotou, Hans-Knöll-Institut, Jena

Dr. Manfred Weidmann, University of Stirling, Stirling

Datum der Verteidigung:

11. April 2018

for my beloved wife

Contents

Abstract

Metagenomics, the study of genetic material taken directly from the environment, generally includes the analysis of a whole microbial community, from both a taxonomic and functional perspective. In contrast to traditional microbiological techniques, this approach makes the isolation and cultivation of single organisms unnecessary. This allows the investigation of non-cultivable microorganisms, which represent the majority of microorganisms. While most metagenomic studies have focused on human and model organisms, the metagenome of marine species is relatively unknown. One area in particular that could be of major interest for aquaculture, but which has not been sufficiently investigated, is the influence of the environment on the composition of the bacterial community.

Indonesia with its myriad of islands is located within the coral triangle, an area that has the highest marine biodiversity on earth. The natural richness of highly valued fish species and ideal aquaculture conditions have made fishery an important economic factor for the country. However, population growth and rapid economic development have led to an increased production of wastewaters, which is discharged virtually untreated into the coastal areas and greatly affecting the ecosystems found there.

This work deals with the effects that highly polluted and non-polluted environmental conditions have on bacterial communities. To do this, the microbial communities of three fish species (*Atule mate*, *Epinephelus sexfasciatus* and *Epinephelus fuscoguttatus*) and one shrimp species (*Penaeus monodon*) were investigated. Two different metagenomic techniques were applied. The results of a 16S amplicon-based approach did not reveal any differences between the investigated environments at a high taxonomic level. However, differences between the sampling locations could be found. A subsequent whole metagenome sequencing approach revealed, that predominant members of the microbiome living within a controlled environment of a mariculture facility are more stably distributed than those within free-living host species. Furthermore, the interactions between endoparasites and potential pathogenic bacteria were investigated. An analysis of the microbiome of shrimp shows only minor differences in the microbial community composition between free-living and aquacultured *Penaeus monodon*. In contrast, the microbial communities of the polluted environment showed significant differences.

Zusammenfassung

Metagenomik ist die Untersuchung von genetischem Material, welches direkt aus der Umwelt entnommen wurde und umfasst hauptsächlich die Untersuchung einer mikrobiellen Gemeinschaft im Ganzen, taxonomisch und funktional. Da im Gegensatz zur klassichen Genomik die Isolierung und Kultivierung von einzelnen Organismen nicht notwendig ist, ermöglicht dieses Forschungsgebiet die Untersuchung unkultivierbarer Mikroorganismen, welchen den überwiegenden Teil der Mikroorganismen bilden. Während viele metagenomische Fragestellungen in Modellorganismen und dem Menschen untersucht wurden, ist das Metagenom mariner Spezies noch relativ unbekannt. Insbesondere der Einfluss von Umweltfaktoren auf die Zusammensetzung mikrobieller Gemeinschaften ist eine interessante und noch nicht ausreichend aufgeklärte Fragestellung, welche in Bezug auf Aquakulturen aber von großen Interesse sein könnte.

Indonesien mit seiner endlosen Anzahl an Inseln liegt inmitten des Korallendreiecks, eine Region mit einer der größten marinen Biodiversität der Welt. Der natürliche Reichtum an hochwertigen Fischspezies und ideale Bedingungen für Aquakulturen haben die Fischerei zu einem wichtigen Wirtschaftsfaktor für das Land gemacht. Ein schnelles Bevölkerungswachstum und die enorme wirtschaftliche Entwicklung des Landes führen jedoch zu einer ansteigenden Produktion von Abwässern, welche fast unbehandelt in die Küstenregionen fließen und einen zunehmend negativen Einfluss auf marine Ökosysteme haben.

Die vorliegende Arbeit beschäftigt sich mit den Auswirkungen von stark verschmutzten und sauberen Wasserbedingungen auf die mikrobiellen Gemeinschaften. Hierzu wurden Proben von drei verschiedenen Fischen, *Atule mate*, *Epinephelus sexfasciatus* und *Epinephelus fuscoguttatus* sowie des Shrimps *Penaeus monodon* auf ihre mikrobiellen Gemeinschaften untersucht. Zur Anwendung kamen dabei zwei unterschiedliche metagenomische Techniken. Während die Ergebnisse einer 16S Amplikon-basierten Analyse auf einem hohen taxonomischen Level keinerlei Unterschiede zwischen den verschiedenen Wasserbedingungen, sondern eher geographische Unterschiede nahelegt, zeigte ein Next Generation Sequencing Ansatz, dass die mikrobiellen Gemeinschaften unter kontrollierten Bedingungen einer Marikultur eine stabilere Verteilung haben, als in frei lebenden Individuen. Zudem wurde das Zusammenspiel zwischen Endoparasiten und potentiell pathogenen Bakterien untersucht. Die Untersuchung der mikrobiellen Gemeinschaften des Shrimps zeigten hingegen, dass zwischen sauberen Wasserbedingungen und kontrollierten Wasserbedingungen einer Aquakultur nur marginale Unterschiede in der Zusammensetzung existieren. Die mikrobiellen Gemeinschaften verschmutzer Wasserbedingungen unterschieden sich hingegen deutlich.

1 Introduction

1.1 Marine environment of Indonesia

The equatorial archipelagic state of Indonesia in Southeast Asia is located inside the coral triangle, an area with a marine biodiversity exceeding that of any other place on earth [2, 3]. Consisting of a myriad of islands surrounded by shallow waters, as well as deep sea areas, the region provides the ideal conditions for a variety of tropical marine ecosystems, including mangroves, sea grass and coral reefs. Over 75% of known coral species can be found in the coastal zones of the coral triangle, thereby making it home to the greatest diversity of coral reef fish and other marine species. Due to the natural richness of high value fish species that exists there, fishery has been always an important economic factor for Indonesia [4]. In addition, the coastline of the country with its countless numbers of bays and shallow waters provides the ideal conditions for aquaculture. According to the Food and Agriculture Organization of the United Nations (FAO) [5], fishery production in Indonesia reached 8.9 million tons in 2012, of which inland and marine catch accounted for about 5.8 million tons and aquaculture for 3.1 million tons. Artisanal fishermen are responsible for 95% of fishery production. In terms of economic value, the fishing industry contributed 3% to the Gross Domestic Product (GDP) of Indonesia and

Figure 1: Satellite image of Southeast Asia. The map shows the Republic of Indonesia, including parts of neighboring countries. Each sampling location of this study is highlighted in red. The lighter blue areas represent shallow waters and the dark blue areas deep waters (modified by [1]).

21% to the agricultural economy in 2012.
With an estimated population of over 255 million people, Indonesia represents the fourth most populated country in the world. Over the last decades, the economy and quality of live have developed rapidly. Both of these factors have resulted in an enormous increase in environmental pollution, also with respect to marine ecosystems. The increase with respect to marine pollution can be attributed to the following: Firstly, land-based pollutants, including coastal and upstream agricultural discharges of pesticides, fertilizer and sediment runoff, as well as urban and industrial development [3], have resulted in the direct disposal of untreated or poorly purified wastewater into the aquatic environment. In addition, the release of waste and toxic materials through oil and gas-related activities and marine traffic accidents have also increased marine pollution [3]. Many species have lost critical habitats used as feeding or breeding grounds due to coral mining, the over-harvesting of mangrove timber [3] and the conversion of mangroves to aquaculture structures [6]. The depletion of fishing stocks due to overexploitation and the increasing number of aquacultures have also affected benthic communities as well as local fish communities and their environment [7, 8, 6, 9]. All of these factors play an important role for the region of Indonesia, where 140 million people live along the coast and fishery resources are a vital source of food and income for the local population [7].

1.2 Metagenomics

The investigation of microbial organisms to discover the organism itself, its metabolic products and pathways, and how it interacts with the environment has been the aim of microbiology from the beginning. However, the investigation of microorganisms using traditional microbiological techniques is limited due to the necessity of having to isolate and cultivate a specific microorganism. Today, it is widely known that only 1% of microbes can be isolated using the traditional techniques of microbiology, which means that the majority of microorganisms and their biological pathways remain inaccessible [10]. Metagenomics, which was first mentioned by Handelsmann *et al.* in 1998 [11], is the study of genetic material taken directly from the environment, e.g., soil, water or the gut of eukaryotic host organisms. In theory, metagenomics enables the investigation of non-cultivable microorganisms in order to access 100% of the genetic information inside an environment without having to isolate and cultivate single microorganisms [12]. The term metagenome, therefore refers to the collection of all genomes and genes from members of the microbiota in a defined environment, whereas the term microbiome refers to the metagenome and the surrounding environmental conditions.
With the advent of next-generation sequencing techniques, which has resulted in the availability of large-scale sequencing data at lower sequencing costs, the field

of metagenomics gained increased attention. So far numerous metagenomic surveys have been performed to elucidate the mechanism of the microbiomes of well-studied eukaryotic host organisms. For instance, the Human Microbiome Project [13, 14] established in 2008 aims to characterize the human microbiome and the influences on human health and diseases. In addition, Micro B3 (Marine Microbial Biodiversity, Bioinformatics, Biotechnology) [15] represents another large interdisciplinary project, in which metagenomics is an important key aspect. The main aim of the project is to develop innovative bioinformatical approaches to make large-scale genomic and metagenomic data accessible for marine ecosystems biology.

1.2.1 Profiling of environmental samples with phylogenetic marker genes

The investigation of microbial composition(i.e., the phylogenetic content of all microorganisms inside an environmental sample) has its roots in a pioneer study by Pace *et al.* (1985) [16], which describes the culture-independent retrieval of 16S rRNA genes [12]. The use of small subunit ribosomal RNA (rRNA) as a phylogenetic marker to detect community composition and diversity in an environment has become a standard technique for metagenomic profiling. In particular, 16S rRNA gene sequences of bacteria and archaea, as well as 18S rRNA gene sequences of eukaryotes, are used as stable phylogenetic markers to analyze the prokaryotic taxonomic composition within environmental samples [17]. The ubiquity and extreme sequence conservation of these markers allows a broad spectrum for the detection of non-cultivable members of microbial communities. In marine metagenomics, in particular, 16S rRNA sequences are frequently used to determine the taxonomical classifications of cultured and uncultured bacteria, such as important host organisms and samples taken directly from water under different environmental conditions.
The 16S rRNA gene, encoding for the small subunit of the bacterial and archaea ribosome, consists of nine variable regions (V1–V9) of highly conserved genes, whereby each region has a different evolution rate. The high level of sequence conservation and its universal presence in all bacteria and archaea makes the 16S rRNA gene the perfect phylogenetic marker. Furthermore, a variety of standard PCR primers for amplification of the variable regions as well as well maintained sequence databases for annotation exist. In addition, many well-developed bioinformatical pipelines for annotation and analysis of 16S data are freely available.
Complete 16S amplicon surveys are capable of investigating the bacterial composition of environmental samples, in order to show the differences between each sample in a given environment (alpha diversity) and also to compare the samples between different environments (beta diversity). In addition, it also allows the investigation of core microbiomes, which are defined as the bacterial composition shared over

each sample in an environment. The advent of next-generation sequencing techniques has led to the renaissance of culture-independent molecular surveys using 16S rRNA sequences due to the increased availability of more sequences at lower cost [18].

1.2.2 Whole metagenome shotgun sequencing approaches

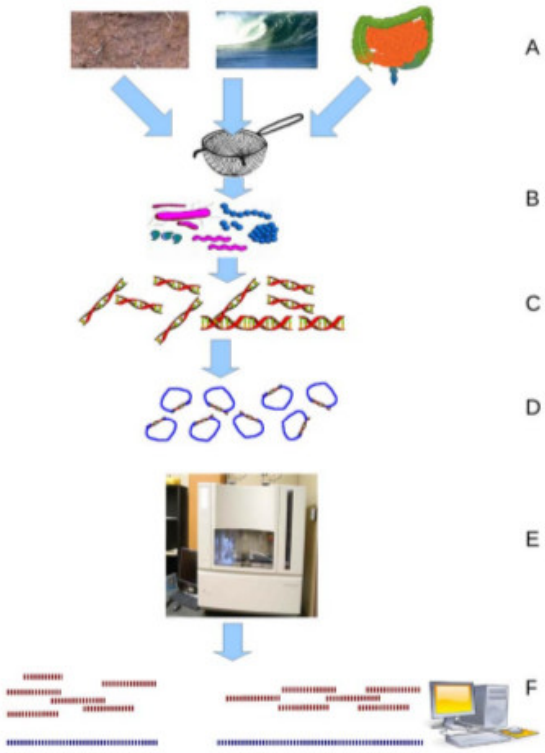


Figure 2: Process of whole metagenome shotgun sequencing. (A) Sampling from the environment; (B) filtering particles, typically by size; (C) extraction of DNA; (D) construction of metagenomic library; (E) sequencing; and (F) metagenomic sequence assembly (obtained from: [19]).

The analysis of whole metagenome shotgun sequencing approaches is gaining importance with the availability of large-scale sequencing data at lower costs for sequencing resulting from the emergence of next-generation sequencing techniques. Metagenomic, metatranscriptomic and other whole-community functional assays provide new ways of studying complex ecosystems involving host organisms, biogeochemical environments, pathogens, biochemistry and metabolism, as well as the interaction between them [20]. In addition, the increased taxonomical resolution provides a more detailed identification of the organisms in a community due to the availability of complete genomes instead of only small single fragments, like 16S rRNA or 18S rRNA, for sequence comparisons.
This is done by performing whole metagenomic shotgun sequencing based on extraction of DNA and RNA directly from the environment, followed by preparation

of a metagenomic library and short-read sequencing using next-generation sequencing techniques. The current standard technique for whole metagenomic shotgun sequencing approaches is the Illumina platform, which offers read lengths of up to 150 bp for HiSeq and 300 bp for MiSeq [21]. After sequencing, the millions of short random sequences that result can then be assembled to gain longer fragments (contigs) or be used as markers for specific organisms or metabolic functions [20].
In contrast to single genome assembly, metagenomic assembly introduces specific problems, which are caused by the presence of multiple genomes inside metagenomic sequencing datasets. Over the last few years a variety of metagenomic assemblers have become available through the use of different strategies. The majority of these use a graph-based reconstruction [20] to solve these specific problems. In most instances, however, a whole genome assembly from a metagenomic dataset remains impossible. Most metagenomic assemblers try to provide the largest reliable and useful contig achievable from input sequences [20].
After the metagenomic assembling step, identification of organisms within an environmental sample can be achieved with a higher taxonomic resolution than with 16S amplicon sequencing. This step can be done either by using *de novo* binning approaches (intrinsic sequence properties) or genome databases (extrinsic information). However, one thing that has to be taken into consideration is that all these methods are biased towards model organisms and pathogens. Intrinsic binning approaches train taxonomic classifiers from reference genomes and then use this sequence-free classifier to bin the metagenomic data with statistical methods. This approach gives great results in environments with insufficient prior information and sequence-based studies, but require large computational resources and long running times. However, extrinsic approaches compare metagenomic reads directly against reference databases to identify taxonomic units. This is mainly done using homology searches with varying sensitivities, i.e, blastn [22] or mapping approaches. The results of these approaches can be highly ambiguous and are more difficult to interpret. In addition, phylogenetic approaches, like least common ancestor algorithms, are needed. Moreover, extrinsic approaches also need a huge amount of computational resources.
One benefit of using a whole metagenome shotgun sequencing approach is the possibility of identifying genes and pathways in environmental samples. In addition, metatranscriptomics can also measure the expression of certain functions. Most functional metagenomic approaches use assembled metagenomic sequences to identify protein-coding genes (CDS). The resulting full CDS are then assigned to functional categories using well-maintained functional databases, like NCBI nr [23], KEGG ontology [24, 25, 26], PFAM [27] or COGs [28]. Broader biological functions are constructed by hierarchical ontologies based on these low level annotations [20]. However, like the sequence databases for taxonomic annotation, these databases are biased towards model organisms and pathogens.

1.2.3 Functional metagenomic-based approaches

Functional metagenomic approaches are based on the screening of DNA library clones directly for a phenotype, whereby genes are recognized by their functions. In contrast to sequence-based metagenomic approaches, a prior annotation step based on the sequence similarities of known genes is not needed. This way, incorrect annotations or biased databases are avoided. However, the results of functional metagenomic approaches are unambiguous.
Following this basic strategy, an environmental sample is first collected and filtered for DNA. Subsequently, the total community DNA is extracted and a metagenomic library is generated from the isolated DNA using a suitable cloning vector. After transferring the library to a suitable host strain, the individual clones can be screened for the presence of enzymatic or other bioactivities originating from the environmental DNA fragment.
Functional metagenomic approaches introduce potential methods for the identification of novel genes and functions, however they are limited in terms of the availability of suitable heterologous expression hosts [10].

1.3 Overview of investigated marine host species

1.3.1 *Atule mate*

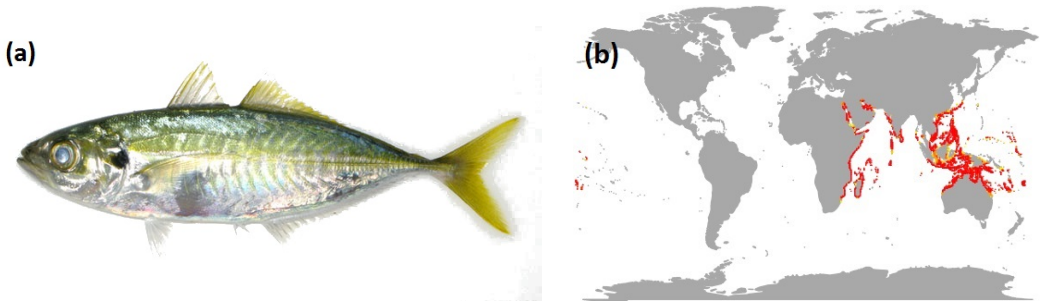


Figure 3: Caught specimen of *Atule mate* (a) and occurrence of the species (b). a: The body is oblong, moderately compressed and olive green dorsally with shading to white or silvery ventrally and greenish yellow caudal and dorsal fins [29]. **b:** The red and yellow areas show the probability of occurrence of *Atule mate* within a range of 95% to 100% (modified from [30, 1]).

Atule mate, commonly known as yellowtail scad (order Perciformes, family Carangidae), is a small pelagic carangid occurring mainly in the tropical and subtropical coastal waters of the Indo-Pacific region [31]. It appears mainly in inshore waters, i.e., mangroves, coastal bays and coral reefs and feeds on cephalopods, crustaceans and planktonic invertebrates. The body of *Atule mate* is oblong and moderately compressed, with an almost evenly convex dorsal and ventral profile. Dorsally the

body is olive green, with shading to white or silvery ventrally. Adult specimens develop an adipose eyelid that completely covers the eye except for a vertical slit over the pupil. The characteristics of *Atule mate* include the greenish yellow caudal and dorsal fins and a black spot posteriorly on the opercle at the level of the upper eye [32]. The school forming *Atule mate* lives at a depth of 80 m and is an important component of fishery. In Southeast Asia, where it represents a prized food fish, *Atule mate* makes up a high proportion of the pelagic catch.

1.3.2 *Epinephelus sexfasciatus*

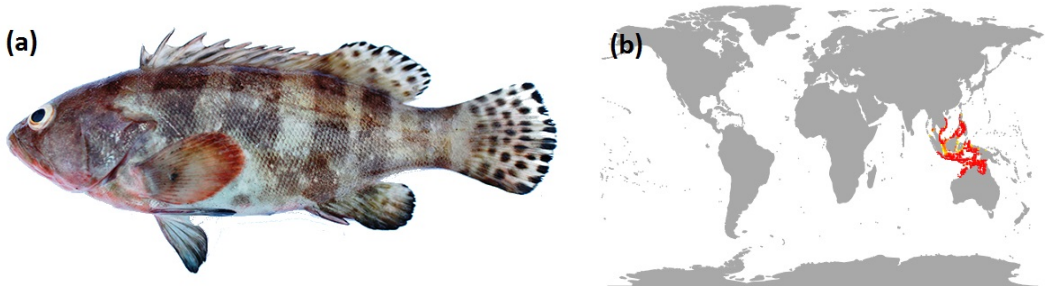


Figure 4: Picture of *Epinephelus sexfasciatus* (a) and occurrence of the species (b). a: The body usually has five brown bars, a grayish head and scattered pale spots [29]. **b:** The red and yellow areas show the probability of occurrence of *Epinephelus sexfasciatus* within a range of 95% to 100% (modified from [30, 1]).

The sixbar grouper, scientifically known as*Epinephelus sexfasciatus* (order Perciformes, family Serranidae), is a marine, reef-associated fish species endemic to the Western Central Pacific. It has been recorded in tropical waters from Thailand and the Philippines to northern Australia, however, it is absent in the oceanic islands [29]. The species usually inhabits silty soft sand or mud bottoms at depths of 10 m to 80 m. These preferences may account for its restricted distribution and absence from oceanic islands [33]. It mainly feeds on small fishes and crustaceans [34].
The sixbar grouper has a grayish brown head and body with one dark brown bar on the nape and five on the body, as well as pale spots scattered all over. The soft dorsal, caudal and pelvic fins have a dusky gray color, while the pectoral fins are grayish or orange-red and not fleshy [29]. While a maximum length of up to 40 cm has been reported, the standard length attained is usually 21 cm. Little is known about the biological or population status of this fish species. There is also not much reported with respect to fishery. *E. sexfasciatus* is commonly caught by trawls and can often be found in local markets [34]. In Indonesia it also occurs in inshore fisheries.

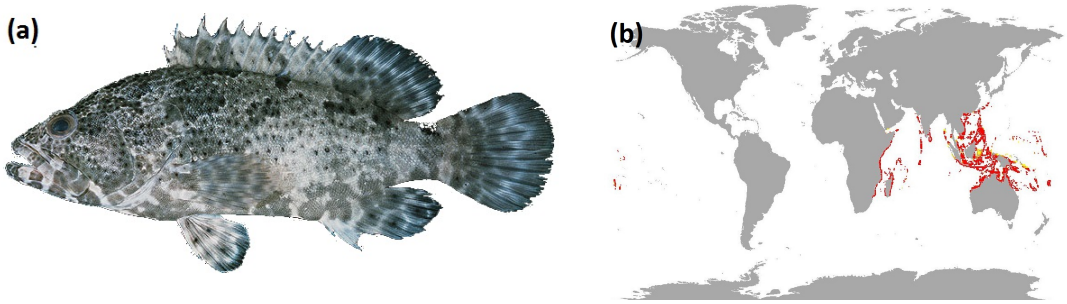


Figure 5: ***Epinephelus fuscoguttatus*** **(a) and occurrence of the species (b). a:** The body is pale yellowish-brown and covered with large, irregular dark brown blotches, as well as tiny brown spots along the back. [29]. **b:** The red and yellow areas show the probability of occurrence of *Epinephelus fuscoguttatus* within a range of 95% to 100% (modified from [30, 1]).

1.3.3 *Epinephelus fuscoguttatus*

The brown-marbled grouper, scientifically known as *Epinephelus fuscoguttatus* (order Perciformes, family Serranidae), is a benthic marine fish species widely distributed in the tropical and subtropical waters of the Indo-Pacific region as well as the Red Sea. The species inhabits shallow waters, preferably near coral reefs and rocky bottoms at depths of 60 m, with juveniles found mainly in seagrass areas [34]. It has a pale yellowish-brown scaled body covered with large, irregular dark brown blotches. The back and sides of the head are also covered by tiny brown spots.
E. fuscoguttatus is one of the largest fish predators on coral reefs and mainly active at dusk. Its diet consists of smaller fishes, crustaceans and cephalopods. It is reported to be ciguatoxic in some regions [35]. As a protogynous hermaphrodite, *E. fuscoguttatus* begins its life cycle as female and some specimens change sex to male at later ages. The change of sex is reported to be socially mediated, involving fishes of 4-5 years of age [36]. It can be cultured by hatcheries, but is also extensively taken from the wild, whereby adults or large juveniles are marketed directly and small juveniles are grown out to market size in captivity. The destruction of seagrass beds and coral reefs, as well as intensive fishing, qualifies it for near-threatened status on the red list [37].

1.3.4 *Penaeus monodon*

Penaeus monodon, commonly known as the giant tiger prawn (order Decapoda, family Penaeidae), is a marine crustacean inhabiting the coasts of the Indo-Pacific ranging from East Africa and the Arabic peninsula to South Asia, Southeast Asia and Australia. While juveniles can be found in estuaries, lagoons and mangroves, adults inhabit rocky or muddy bottoms, ranging in depths of 0 to 110 m [38].
With a maximum length of 33 cm, *Penaeus monodon* represents the largest species

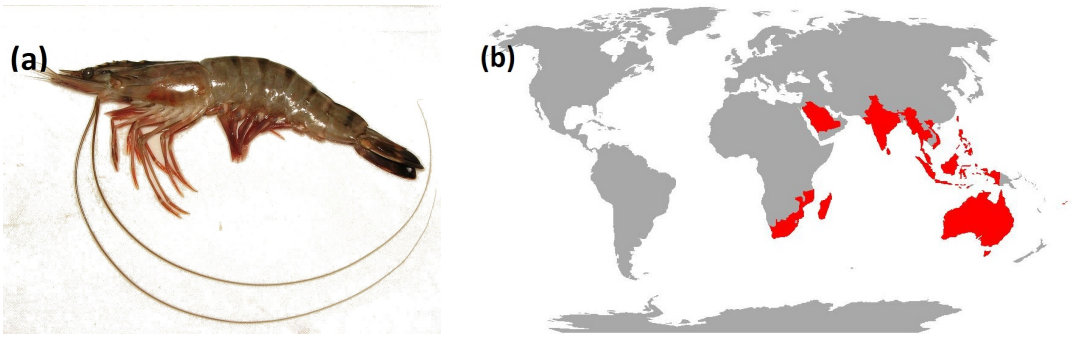


Figure 6: Caught adult specimen of *Penaeus monodon* (a) and main producing countries (b). a: A typical prawn body includes a head, tail, five pleopods, five pereopods, numerous head appendages, a rostrum and carapace. A distinct feature of *Penaeus monodon* are the typical black and white stripes on the back and tail. [38]. **b:** The main producing countries of *Penaeus monodon* from hatcheries and aquaculture combined are highlighted in red on the world map (modified from [39, 1])

of the genus *Penaeidae*, with females typically attaining a larger size than male specimens [39]. The body is typical for a prawn, which includes a head, tail, five pairs of swimming legs (pleopods), five pairs of walking legs (pereopods) and numerous head appendages [38]. The cephalotorax is enclosed by a carapace. A well-developed dorsally and ventrally toothed rostrum follows as an extension of the carapace, which covers the head. In contrast to other species of the *Penaeidae* genus, they have no expod on their fifth pereiopods. The base body color varies from green, brown, red, gray and blue, with typical black and white stripes on the back and tail. In addition, these stripes can alternate between black/yellow and blue/yellow on the abdomen [38].

In the larval stages, *Penaeus monodon* filter feeds on plankton, diatoms and other small organisms, which changes to benthic feeding in the adult stage. The giant tiger prawn represents an important economic factor for fisheries in Southeast Asia, which has seen a rapid increase in the number of aquacultures.

1.4 Objective

The increase in marine pollution in Indonesian waters has had a huge impact on coastal ecosystems. This can especially be seen in the area of Jakarta Bay in North Jakarta. The thirteen rivers flowing through the coastal mega-city receive enormous amounts of untreated wastewaters, which is discharged into Jakarta Bay [40, 41]. The effects of this highly polluted environment on the microbiome of marine host species are yet unknown.

This thesis investigates the fecal microbiome of three important Indonesian food fish species (the migrating *Atule mate*, the less mobile *Epinephelus sexfasciatus* and the *Epinephelus fuscoguttatus*) and samples of *Penaeus monodon* (one of the most widely consumed marine crustaceans worldwide) collected from different water bodies. Samples of a highly polluted marine environment were obtained from Jakarta Bay in the north of Java (*A. mate*, *E. sexfasciatus*, *P. monodon*) and comparative samples representing cleaner water bodies were collected at Pulau Seribu (*E. fuscuguttatus*), a chain of islands located to the north of Jakarta Bay, as well as from the coastal waters of Cilacap (*A. mate*, *E. sexfasciatus*), a city on the southern coast of Central Java. In addition, samples of *P. monodon* were collected from the clean waters of Bali Bay and a traditional aquaculture farm in Bali.

First, a metagenomic amplicon-based approach using 16S rRNA amplicon sequencing was applied to the fecal samples of the three host fish species (*A. mate*, *E. sexfasciatus* and *E. fuscoguttatus*) in order to investigate and compare the fecal bacterial communities and biodiversity between the highly polluted marine environment of Jakarta Bay and the cleaner water bodies around Cilacap (*A. mate*, *E. sexfasciatus*) as well as samples from the free-living and mariculture environment (*E. fuscoguttatus*) at the marine national park Thousand Islands [42].

In a more detailed study, samples of *E. fuscoguttatus* were analyzed using a whole metagenomic sequencing approach in order to examine the bacterial community composition in more detail and determine the eukaryotic components within the fecal samples. Using a whole metagenomic sequencing approach enables a functional analysis of the predominant phyla in the microbiome to explore and compare the functional composition between the free-living and mariculture environment [43]. Furthermore, the bacterial communities of *Penaeus monodon* were investigated and compared between highly polluted (Jakarta Bay) and non-polluted (Bali Bay) environments and with *P. monodon* obtained from a traditional aquaculture in Bali. For this purpose, an amplicon-based approach was applied. For this study, we focus on the bacterial and viral pathogens that are typical for *P. monodon*. The bacterial community composition and potential candidates for pathogenic bacteria were determined by bioinformatical analyses, and the results confirmed and analyzed further by qPCR [44].

2 List of Publications

1. **Hennersdorf P**, Kleinertz S, Theisen S, Abdul-Aziz MA, Mrotzek G, Palm HW, Saluz HP. (2016) Microbial Diversity and Parasitic Load in Tropical Fish of Different Environmental Conditions.
PLoS ONE 11 (3): e0151594. doi: 10.1371/journal.pone.0151594

   Status: published in PLoS ONE

   Summary: The first study analyzes and compares the fecal bacterial communities and parasites of three important Indonesian fish species collected from the highly polluted Jakarta Bay (*E. sexfasciatus*, *A.mate*) with those of the less polluted Indonesian area of Cilacap (*A. mate*, *E. sexfasciatus*) and Thousand Islands (*E. fuscoguttatus*). In addition, *E. fuscoguttatus* specimens from an open water mariculture facility were compared against free-living specimens. The study included the bacterial community composition, both core and shared microbiomes and a parasitological investigation. It showed that the microbial composition of phylogenetically distant fish species, i.e., *A. mate* and *E. sexfasciatus* from Jakarta Bay and Cilacap, were more closely related than the microbial composition of more phylogenetically related species, i.e., *E. fuscoguttatus* and *E. sexfasciatus* from Jakarta Bay, Cilacap and Thousand Islands. The core and shared microbiomes showed the same classes of bacteria, whereas the proportions varied significantly.

   Author contribution: PH 25%, SK 25%, ST 5%, GM 20%, MAAA 5%, HWP 10%, HPS 10%

2. **Hennersdorf P**, Mrotzek G, Abdul-Aziz MA, Saluz HP (2016) Metagenomic analysis between free-living and cultured *Epinephelus fuscoguttatus* under different environmental conditions in Indonesian waters.
Marine Pollution Bulletin 110 (2): 726-734,
doi: 10.1016/j.marpolbul.2016.05.009.

   Status: published in Marine Pollution Bulletin

   Summary: This study focused on the comparison of fecal metagenome of free-living and cultivated *Epinephelus fuscoguttatus*. Using a whole metagenomic sequencing approach enabled the elucidation of prokaryotic and eukaryotic DNA. Feces samples from mariculture revealed a highly stable distribution of several orders of bacteria compared to highly diverse

free-living samples. In addition, a functional annotation of the dominating bacteria was performed, which revealed a number of functions related to DNA metabolic processes and the response to antibiotics.

Author contribution: PH 45%, GM 35%, MAAA 10%, HPS 10%

3. Oetama VSP, **Hennersdorf P**, Abdul-Aziz MA, Mrotzek G, Haryanti H, Saluz HP (2016) Microbiome analysis and detection of pathogenic bacteria of
Penaeus monodon from Jakarta Bay and Bali.
Marine Pollution Bulletin 110 (2): 718–725,
doi: 10.1016/j.marpolbul.2016.03.043.

Status: published in Marine Pollution Bulletin

Summary: This study examines and compares the fecal microbiota of *Penaeus monodon* from highly polluted waters around Jakarta Bay with those of the less polluted waters of Bali. Next-generation sequencing techniques were used to identify the potential bacterial pathogens and common viral diseases of shrimp. Furthermore, qPCR profiling was used as a confirmatory step and further revealed that the two potentially pathogenic species *Vibrio alginolyticus* and *Photobacterium damselae* were present in most of the samples. In addition, viral diseases of shrimp were also discovered among the samples.

Author contribution: VPO 35%, PH 20%, MAAA 5%, GM 25%, HH 5%, HPS 10%

3 Publications

Microbial Diversity and Parasitic Load in Tropical Fish of Environmental Conditions

Hennersdorf P, Kleinertz S, Theisen S, Abdul-Aziz MA, Mrotzek G, Palm HW, Saluz HP. **(2016)**

PLoS ONE 11 (3): e0151594
doi: 10.1371/journal.pone.0151594

PLOS | ONE

RESEARCH ARTICLE

Microbial Diversity and Parasitic Load in Tropical Fish of Different Environmental Conditions

Philipp Hennersdorf[1,2☯], Sonja Kleinertz[3☯], Stefan Theisen[3], Muslihudeen A. Abdul-Aziz[5], Grit Mrotzek[1], Harry W. Palm[3,4], Hans Peter Saluz[1,2]*

1 Cell and Molecular Biology, Leibniz-Institute for Natural Product Research and Infection Biology, Jena, Germany, **2** Friedrich-Schiller-University, Jena, Germany, **3** Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany, **4** Faculty of Veterinary Medicine, Udayana University, Kampus Bukit Jimbaran, Indonesia, **5** Australian Centre for Ancient DNA, University of Adelaide School of Biological Sciences, Adelaide, Australia

☯ These authors contributed equally to this work.
* hanspeter.saluz@leibniz-hki.de

Citation: Hennersdorf P, Kleinertz S, Theisen S, Abdul-Aziz MA, Mrotzek G, Palm HW, et al. (2016) Microbial Diversity and Parasitic Load in Tropical Fish of Different Environmental Conditions. PLoS ONE 11 (3): e0151594. doi:10.1371/journal.pone.0151594

Editor: Pankaj Kumar Arora, Yeungnam University, REPUBLIC OF KOREA

Received: November 20, 2015

Accepted: March 1, 2016

Published: March 28, 2016

Data Availability Statement: We confirm that all relevant data has been included within the paper and Supporting Information files. In the paper, we have mentioned accession numbers that are linked to the storage of RAW data from the microbiome analysis. Raw sequence data are deposited at NCBI's Short Read Archive under accession number SRP059667.

Funding: Financial support was provided through the German Academic Exchange Service (DAAD) (ST) German Federal Ministry for Education and Science within the framework of the joint Indonesian-German research programs SPICE III - MABICO (Science for the Protection of Indonesian Coastal Marine

Abstract

In this study we analysed fecal bacterial communities and parasites of three important Indonesian fish species, *Epinephelus fuscoguttatus*, *Epinephelus sexfasciatus* and *Atule mate*. We then compared the biodiversity of bacterial communities and parasites of these three fish species collected in highly polluted Jakarta Bay with those collected in less polluted Indonesian areas of Cilacap (*E. sexfasciatus*, *A. mate*) and Thousand Islands (*E. fuscoguttatus*). In addition, *E. fuscoguttatus* from net cages in an open water mariculture facility was compared with free living *E. fuscoguttatus* from its surroundings. Both core and shared microbiomes were investigated. Our results reveal that, while the core microbiomes of all three fish species were composed of fairly the same classes of bacteria, the proportions of these bacterial classes strongly varied. The microbial composition of phylogenetically distant fish species, i.e. *A. mate* and *E. sexfasciatus* from Jakarta Bay and Cilacap were more closely related than the microbial composition of more phylogentically closer species, i.e. *E. fuscoguttatus*, *E. sexfasciatus* from Jakarta Bay, Cilacap and Thousand Islands. In addition, we detected a weak negative correlation between the load of selected bacterial pathogens, i.e. *Vibrio* sp. and *Photobacterium* sp. and the number of endoparasites. In the case of *Flavobacterium* sp. the opposite was observed, i.e. a weak positive correlation. Of the three recorded pathogenic bacterial genera, *Vibrio* sp. was commonly found in *E. fuscoguttatus* from mariculture, and lessly in the vicinity of the net cages and rarely in the fishes from the heavily polluted waters from Jakarta Bay. *Flavobacterium* sp. showed higher counts in mariculture fish and *Photobacteria* sp. was the most prominent in fish inside and close to the net cages.

Ecosystems, BMBF Grant Nos. 03F0641B, 03F0641D).

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Indonesia's population growth and rapid economic development has led to an increased production of wastewater, from industry, farming and households [1]. Furthermore, inadequately purified wastewater is regularly disposed into coastal waters, resulting in a negative influence on marine ecosystems and its inhabitants [2]. Other anthropogenic activities such as capture fisheries and aquaculture also affect benthic communities as well as local fish communities and their environment [1, 3–5]. The consequences of these factors on the microbiome of fish and the possible implications on fish health are yet unknown. Within the coral triangle, Indonesian marine biodiversity exceeds that of any other place on earth [6]. This unique diversity includes all kinds of aquatic organisms, including marine fish, their parasites and pathogens. Fish parasites have been recognized as important sentinel organisms that are able to detect changes in environmental conditions [7–9]. Their diversity in tropical Indonesian waters is high, resulting in more than 80 different fish parasite species that have been recorded from groupers (*Epinephelinae*) kept under mariculture conditions [10]. It has been noted that the number of wild fish parasites exceeds that of mariculture fish [11]. This contrasts the observation that viral and bacterial disease outbreaks occur more regularly in mariculture fish, however, without any evidence for e.g. vibrioses or other bacteria caused skin diseases on Indonesian wild fish. It can be assumed that the environmental conditions, parasite infections and viral or bacterial disease outbreaks are linked and influence each other. According to Brown *et al.* [12], diet-induced altered microbiota results in dysbiosis that may result in inflammatory diseases in humans and contribute to an inappropriate inflammatory response. The microbiome of fish has been recently studied, with common core microbiome detected for certain fish species [13, 14]. The microbiome of marine fish revealed a rich biodiversity that predictably reacts to changing intestinal conditions. Xia *et al.* [15] recorded 33 phyla, 66 classes, 130 orders and 278 families in the intestinal microbiome of Asian seabass (*Lates calcarifer*). They also reported *Proteobacteria* (48.8%), *Firmicutes* (15.3%) and *Bacteroidetes* (8.2%) as the three most abundant bacteria taxa. Under starvation, *Bacteroidetes* were found to be dramaticaly enriched, while *Betaproteobacteria* was significant depleted. A comparison of the microbiome of fish from different environmental conditions such as mariculture and free-living has not yet been studied. In addition, while detailed parasitological investigations on important fish species such as e.g. *Lates calcarifer* [16] and *Epinephelus* spp. [9–11, 17] has been done, possible effects of parasite infection on fish microbiome is still unknown. As a result, we have sampled three important perciform Indonesian food fish species, the migrating pelagic yellowtail scad *Atule (A.) mate*, family Carangidae, less mobile sixbar grouper *Epinephelus (E.) sexfasciatus*, family Serranidae, and brown-marbled grouper *Epinephelus (E.) fuscoguttatus*, family Serranidae, from different water bodies and regions of Java. The samples were obtained from Jakarta Bay in the North of Jakarta (*A. mate*, *E. sexfasciatus*), a booming coastal megacity in Indonesia with over nine million inhabitants. The thirteen rivers that flow through this area receive enormous amounts of untreated wastewater from households and industries and discharge these high pollutant loads into Jakarta Bay [18, 19]. Comparative samples, representing cleaner water bodies, were collected at Pulau Seribu (*E. fuscuguttatus*), a chain of islands located to the North of Jakarta Bay, consisting of 110 islands stretching 45 km North into the Java Sea, added to the Thousand Islands Marine National Park in 2002, and from coastal waters off Cilacap (*A. mate*, *E. sexfasciatus*), a city at the South coast of Central Java.

Materials and Methods

Sample collection and examination

A total of 12 brown-marbled groupers *Epinephelus fuscoguttatus* (Forsskål, 1775), six sixbar groupers *Epinephelus sexfasciatus* (Valenciennes, 1828) and six yellowtale scads *Atule mate*

Table 1. Experimental overview.

sample	host species	sampling location	sampling site	Number of parasites	*Vibrio* sp.	*Flavobacterium*	*Photobacterium* sp.
am1	*A. mate*	Jakarta	free-living	85	1.73	0.001	4.50
am2	*A. mate*	Jakarta	free-living	56	0.07	0.000	0.18
am3	*A. mate*	Cilacap	free-living	0	0.15	0.000	0.54
am4	*A. mate*	Cilacap	free-living	3	0.17	0.000	0.62
es1	*E. sexfasciatus*	Jakarta	free-living	58	0.019	0.002	0.18
es2	*E. sexfasciatus*	Jakarta	free-living	56	0.17	0.002	0.37
es3	*E. sexfasciatus*	Jakarta	free-living	70	0.48	0.001	0.44
es4	*E. sexfasciatus*	Cilacap	free-living	127	0.39	0.000	0.39
es5	*E. sexfasciatus*	Cilacap	free-living	51	2.56	0.001	1.48
es6	*E. sexfasciatus*	Cilacap	free-living	114	0.07	0.001	0.16
ef1	*E. fuscoguttatus*	Thousand Islands	free-living	23	48.43	0.001	20.66
ef2	*E. fuscoguttatus*	Thousand Islands	free-living	7	10.91	0.001	72.72
ef3	*E. fuscoguttatus*	Thousand Islands	free-living	9	2.86	0.001	90.87
ef4	*E. fuscoguttatus*	Thousand Islands	free-living	48	0.68	0.007	84.72
ef5	*E. fuscoguttatus*	Thousand Islands	mariculture	3	53.24	0.306	1.47
ef6	*E. fuscoguttatus*	Thousand Islands	mariculture	4	25.71	0.000	53.18
ef7	*E. fuscoguttatus*	Thousand Islands	mariculture	44	23.13	0.001	57.29
ef8	*E. fuscoguttatus*	Thousand Islands	mariculture	14	0.48	0.006	66.29
ef9	*E. fuscoguttatus*	Thousand Islands	mariculture	5	77.54	0.011	3.93
ef10	*E. fuscoguttatus*	Thousand Islands	mariculture	25	23.37	0.000	46.39

Investigated samples are listed by host species, sampling location and sampling site. The name of each sample is also given and corresponding to the figures and text. Additional the number of parasites per sample and the measured bacterial content for three fish pathogenic bacteria are listed.

doi:10.1371/journal.pone.0151594.t001

(Cuvier, 1833) were studied from i) Jakarta Bay fish markets (Pasar Ikan Pelelangan: 6°06′17.7″S 106°46′31.5″E), ii) the 50 km remote Thousand Islands (Pulau Seribu: 5°44′13.3″S 106°36′31.0″E) National Park (North Java) and iii) Penyu Bay fish markets (Tempat Pelelangan Ikan (TPI)—Pelabuhan Perikanan Samudera Cilacap, 7°43′25.0″S 109°01′22.7″E), Cilacap (South Java), Indonesia (Table 1). All samples were collected during the 2012 rainy season.

The fish species *A. mate* and *E. sexfasciatus* were obtained from local fishermen; *E. fuscoguttatus* originated from an open water mariculture facility (Nusa Karamba Aquaculture) respectively were caught in the direct surrounding of the net cages by fish traps or with a fishing rod. All fish purchased from the market were declared as fresh for human consumption. Fish from Thousand Islands were dissected in the local laboratory (Nusa Karamba Aquaculture) right after catching, purchased fishes were separated into plastic bags and transported immediately to the laboratory or kept on ice and then frozen (≈ −20°C) until subsequently dissected at the Faculty of Biology, Jenderal Soedirman University, Purwokerto (UNSOED) and the Faculty of Veterinary Medicine. Total fish length (TL), standard fish length (SL), total weight (TW), slaughter weight (SW) and liver weight (not shown, used for the calculation of the hepatosomatic index were measured to the nearest 0.1 cm and 0.1 g prior to the parasitological examination [20] (Table 2).

Parasitological examination followed Palm & Bray [21]. Skin, fins, eyes, gills, nostrils, mouth- and gill cavity were examined for ectoparasites. Inner organs such as the digestive tract, liver, gall bladder, spleen, kidneys, gonads, heart and swim bladder were separated and transferred into saline solution for microscopically examination under the stereomicroscope (Zeiss Stemi DV4) in order to allow a quantitative parasitological examination of each organ; belly flaps and

Table 2. Fish morphometrical data.

fish species	area	sample collection point	n	TL [cm]	SL [cm]	TW [g]	SW [g]	m	f	juvenile	HSI	K	H' total	H' endo	E total	E endo
Epinephelus fuscoguttatus	Thousand Islands	fish pond	7	24.90 (23.1–26.2)	21.25 (20.1–22.4)	296.98 (249.5–354.2)	260.20 (221.7–302.2)	5	-	2	1.4	1.685	1.123	0.195	0.689	0.282
Epinephelus fuscoguttatus	Thousand Islands	free-living	5	27.60 (18.5–36.4)	24.02 (16.3–32.0)	475.44 (115.7–955.0)	427.16 (101.1–843.7)	-	-	5	1.1	2.032	1.000	0.300	0.5	0.22
Epinephelus sexfasciatus	Cilacap	fish market	3	27.33 (26.7–28.2)	22.60 (22.0–23.4)	330.13 (299.9–369.2)	283.07 (185.9–358.0)	-	3	-	1.39	0.006	1.301	0.981	0.626	0.548
Epinephelus sexfasciatus	Jakarta	fish market	3	24.83 (24.2–25.2)	20.50 (19.7–21.1)	231.07 (208.3–256.0)	218.77 (198.6–238.6)	-	-	3	1.43	0.006	0.933	0.620	0.521	0.386
Atule mate	Cilacap	fish market	3	24.20 (22.2–25.6)	19.50 (18.1–20.7)	150.00 (114.3–180.3)	140.07 (107.7–168.1)	2	-	1	1.4	0.988	0.451	0.451	0.650	0.650
Atule mate	Jakarta	fish market	3	26.20 (25.3–27.4)	21.20 (20.4–22.5)	177.20 (166.7–188.9)	165.27 (155.9–178.5)	3	-	-	0.8	0.919	1.436	1.118	0.578	0.54

total (TL) and standard length (SL) in cm, total (TW) and slaughter weight (SW) in g, hepatosomatic index (HSI) and condition factor (K) from different Indonesian sampling sites during rainy season 2012. Additionally given are the Shannon-Wiener diversity index and Evenness for all parasites in a sample (H' total; E total) and calculated only for the endoparasites (H' endo; E endo), m: male, f: female

doi:10.1371/journal.pone.0151594.t002

musculature (fillets) were examined on a candling table. Isolated parasites were fixed in 4% borax-buffered formalin and preserved in 70% ethanol. Finally, the musculature was sliced into 0.5–1 cm thick filets, pressed between two petri dishes to identify and isolate parasites from the musculature. Nematoda were dehydrated in a graduated ethanol series and transferred to 100% glycerine (Riemann, 1988). Digeneans, monogeneans and cestodes were stained with acetic carmine, dehydrated, cleared with eugenol and mounted in Canada balsam, whereas crustaceans were dehydrated and transferred directly into balsam. The identification of parasites was based on original descriptions [22–39].

During parasitological investigation feces samples were collected. The intestine was carefully cut and feces (without bones or big, solid components) was scraped with a scoop and stored in 99.9% EtOH for subsequent analyses at the Leibniz Institute for Natural Product Research and Infection Biology e.V. Hans-Knöll-Institute (HKI), Jena, Germany.

Parasitological parameters

Parasitological calculations followed Bush *et al.*[40] The present study applies the method by Palm *et al.*[7], Palm & Rueckert [9], Kleinertz & Palm [41] and Kleinertz *et al.* [42] to monitor the parasite community of Indonesian fish. This is based on the assumption that data and parameters based on the prevalence of certain parasites are characteristic for undisturbed environmental conditions with high parasite diversity. Ecological parameters were evaluated to indicate regional differences, such as the diversity indices Shannon-Wiener and Evenness (both for all parasites as H' total resp. E total, and for the endoparasites exclusively as H' endo resp. E endo, see Kleinertz & Palm [41] and Kleinertz *et al.* [42]), fish ecological indices such as the hepatosomatic index, condition factor and parasitological parameters such as ecto- endoparasite ratio and differences in prevalence of metazoan parasite infections [7, 9, 42].

The total diversity (Shannon–Wiener diversity index [43] H') and the Evenness index (E) of Pielou [44] were calculated for each fish species. According to Kleinertz *et al.* [41] the hepatosomatic index was calculated as a descriptor of a possible pollution impact to the fish host, which may affect increasing liver weights (LW) in relation to the total weight (TW) of the host [45].

Isolation of microbial genomic DNA

Feces samples (5 to 10 mg per specimen) were homogenized in lysis buffer (Bio u. Sell) using a Precellys tissue homogenizer with 1.4 mm ceramic beads and a 3× 30 s homogenization time with 30 s pause at 5000 rpm. Samples were incubated for one hour at 37°C. Subsequently, RNaseA was added and incubation was done for one hour at 37°C and subjected to an over-night proteinase K digestion. Whole genomic DNA was extracted using a Phenol/Chloroform/Isoamyl alcohol extraction followed by ethanol precipitation and quantification.

PCR amplification and Sequencing

Universal prokaryotic primers F515/R806 was used to amplify the V4 region of the bacterial/archaeal 16S rRNA gene [46, 47]. Primers were modified to include Illumina Nextera flowcell adapter sequences, additional forward and reverse primer pads to avoid primer-dimer formation, and a 2-bp linker sequence not matching against any 16S rRNA sequence immediately upstream of the gene primer [47]. The reverse primers also incorporated 12-bp error-correcting Golay barcodes [47].

For each individual sample, three 20 μl PCR reactions (and a negative control) were set up containing 10 ng genomic DNA, 1.25 U TaKaRa SpeedSTAR HS DNA polymerase, 0.2 μM of each primer, corresponding Fast Buffer 1, and 200 μM dNTP final concentration. Reactions were performed with an initial denaturation step for 3 min at 95°C followed by a 40-cycle amplification (95°C for 10 s, 62°C for 30 s), and a final elongation step of 2 min at 72°C on an Applied Biosystems 9800 Fast Thermal Cycler. PCR products were visualized using gel electrophoresis and for successful samples blue, replicate reactions were combined and primer multimers, polymerase, and dNTPs were removed using the Agencourt AMPure XP post-PCR cleanup kit (Beckman Coulter).

Cleaned PCR-products were quantified using the Agilent 2100 Bioanalyser, pooled in equimolar concentration, and subjected to 250-bp paired-end amplicon sequencing on an Illumina MiSeq platform at StarSEQ GmbH (Mainz). Raw sequence data are deposited at NCBI's Short Read Archive under accession number SRP059667.

Ethics statement

In this study, experiments were not performed on live vertebrates. Instead, freshly caught dead fish was used and therefore no ethics statement is required. Samples were taken within the INDONESIAN GERMAN JOINT RESEARCH COOPERATION "Science for the protection of Indonesian marine Coastal Ecosystems—A GERMAN INDONESIAN INITIATIVE IN EARTH SYSTEM RESEARCH. With research permit from RISTEK, the INDONESIAN STATE Ministry of research and technology.

Data processing and statistical analyses

Raw sequence base call files (bcl) were converted into FASTQ format and demultiplexed using the CASAVA v1.8.2 (Illumina) software. Clustering of the reads into Operational Taxonomic Units (OTUs) was performed using the uparse pipeline as implemented in usearch 7.0.1090 [48]. Before clustering, a number of preprocessing steps were carried out: Paired reads were

merged using the fasta_mergepairs command with a minimum Phred score cutoff threshold of 5 and a minimum overlap length of 75 bp. Merged reads were trimmed to a length of 250 bp and filtered if the expected number of errors exceeded 0.5 (fastq_filter). Filtered reads were pooled across samples and dereplicated using the derep_fulllength command. The dereplicated reads were sorted by abundance and all singletons were discarded.

The resulting high-quality sequences were grouped into OTUs using the UPARSE-OTU algorithm [48] (cluster_otus) at a 97% sequence similarity cutoff. This step includes chimera filtering based on models built from more abundant reads. An additional reference-based chimera filtering step was performed using the UCHIME algorithm [49] (uchime_ref) and the ChimeraSlayer reference database (http://microbiomeutil.sourceforge.net). The remaining sequences were considered OTU representative sequences or phylotypes, and mapped against the filtered sample reads at an identity threshold of 97% (usearch_global) to create an OTU abundance table.

OTU sequences were assigned to a taxonomic lineage by inferring the lowest common ancestor for the top BLAST matches against the Greengenes database [50]. Only BLAST hits with a query coverage above 75% and a bitscore above a cutoff value of 97% of the bitscore achieved by the best hit, were considered. Community analyses were performed in R [51] using the packages phyloseq [52] and DESeq2 [53], as well as vegan [54] for diversity analysis.

Results

Analysis of fecal bacterial communities

In total, 8,453,888 valid sequence reads binned into 484 Operational taxonomic units (OTU) were retrieved from 24 fecal samples of the three target species. Three samples (am5, am6 and ef12) were excluded from further analyses due to low number of reads (180, 17,022, and 519 reads, respectively). In addition, one sample from *E. fuscoguttatus* (ef11) was filtered out, due to stochastic behavior of the microbial community, caused by sampling problems. The remaining 20 samples yielded on average 413,100 mapped reads ranging from 154,600 to 719,100 reads. Two OTUs with 1,938 and 40 reads, respectively, were of mitochondrial and chloroplast origin and excluded from subsequent analyses. The remaining 482 OTUs were classified into a total of 19 different phyla, in decreasing order of abundance: *Proteobacteria* (85.93%), *Firmicutes* (11.47%), *Fusobacteria* (1.84%), *Spirochaetes* (0.48%), *Actinobacteria* (0.11%), *Bacteroidetes* (0.06%), *Acidobacteria* (0.04%), *Chlamydiae* (0.02%), *Lentisphaerae* (0.02%), *Cyanobacteria* (0.01%), *Verrucomicrobia*, *Planctomycetes*, *Armatimonadetes*, *WPS-2*, *Tenericutes*, *Chloroflexi*, *Nitrospirae*, *TM6*, *Thermi* (<0.01%).

For all three fish species, the predominant bacterial phyla were *Proteobacteria* (average (avg) 85.39%, standard deviation (s.d.) 19.73%) followed by *Firmicutes* (avg 11.88%, s.d. 20.15%) and *Actinobacteria* (avg 0.13%, s.d. 0.26%) (Fig 1a). Considering only bacterial phyla with a relative abundance of more than 0.1%, *E. fuscoguttatus* showed a markedly higher bacterial diversity than both *E. sexfasciatus* and *A. mate* (eight vs. three phyla).

Across the rare phyla, *A. mate* and *E. sexfasciatus* had an increased bacterial diversity with 16 phyla against 11 phyla for *E. fuscoguttatus*, notably the five additional phyla appeared predominantly in *E. fuscoguttatus* samples. Among the remaining 11 phyla, three different host species had a different composition. Three of the *A. mate* samples, two from Cilacap (am3, am4) and one from Jakarta (am2), showed low abundances, whereas one sample from Jakarta (am1) had a higher abundance of *Spirochaetes* (0.44%) and *Fusobacteria* (0.14%). The samples derived from *E. sexfasciatus* shared in general a similar composition of rare bacterial phyla, consisting mainly of *Bacteriodetes* (0.06%), *Fusobacteria* (0.02%) and *Cyanobacteria* (0.02%). At least one sample of *E. fuscoguttatus*, from outside (ef4) the net cages offered a higher abundance for *Chlamydiae* (0.17%), meanwhile the other samples had nearly none abundance for rare phyla (Fig 1b).

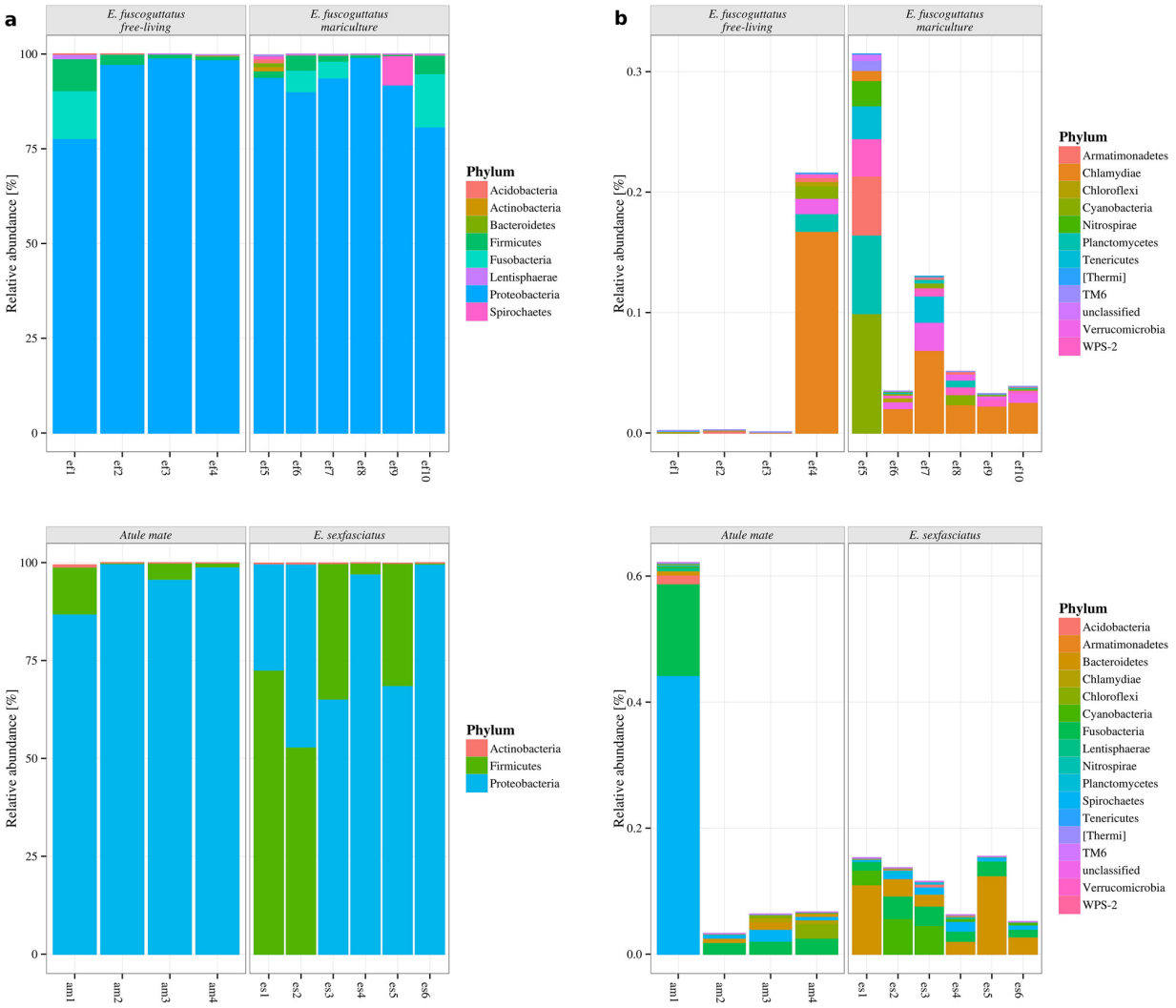


Fig 1. Taxonomic summary of predominant (a) and rare (b) phyla across the three fish species. To determine the predominant phyla only OTUs with an abundance of over 0.01% per sample were selected, resulting in three phyla for *A. mate* and *E. sexfasciatus* and eight phyla for *E. fuscoguttatus*. To determine rare phyla with relative abundance counts of less than 0.01% are included in this plot. ef1-ef10 refers to all samples from *E. fuscoguttatus*, while am1-am4 belongs to *A. mate* and es1-es6 to *E. sexfasciatus*.

doi:10.1371/journal.pone.0151594.g001

In order to assess diversity of the microbial communities, all samples were rarefied to same library size, resulting in 90 OTUs that were sorted out. Following which, three statistical models were applied. The observed OTU richness (Fig 2a) of *A. mate* showed three samples with a related number of OTUs (am1, am3, am4) and one sample from Jakarta (am2) with a reduced number of OTUs, resulting in a median of 178 OTUs. For *E. sexfasciatus*, the median observed OTU richness was 205, distributed to three samples below (es1, es5, es6) and three samples above (es2, es3, es4) the median value. The samples of *E. fuscoguttatus* displayed a higher variation. Free-living samples offered the lowest median observed OTU richness with 136 OTUs, whereas deviation between the samples was very high. In contrast samples derived from mariculture showed a higher observed OTU richness with a median value of 217 OTUs. The non-parametric richness estimator Chao1, providing a statistical estimation of the true species

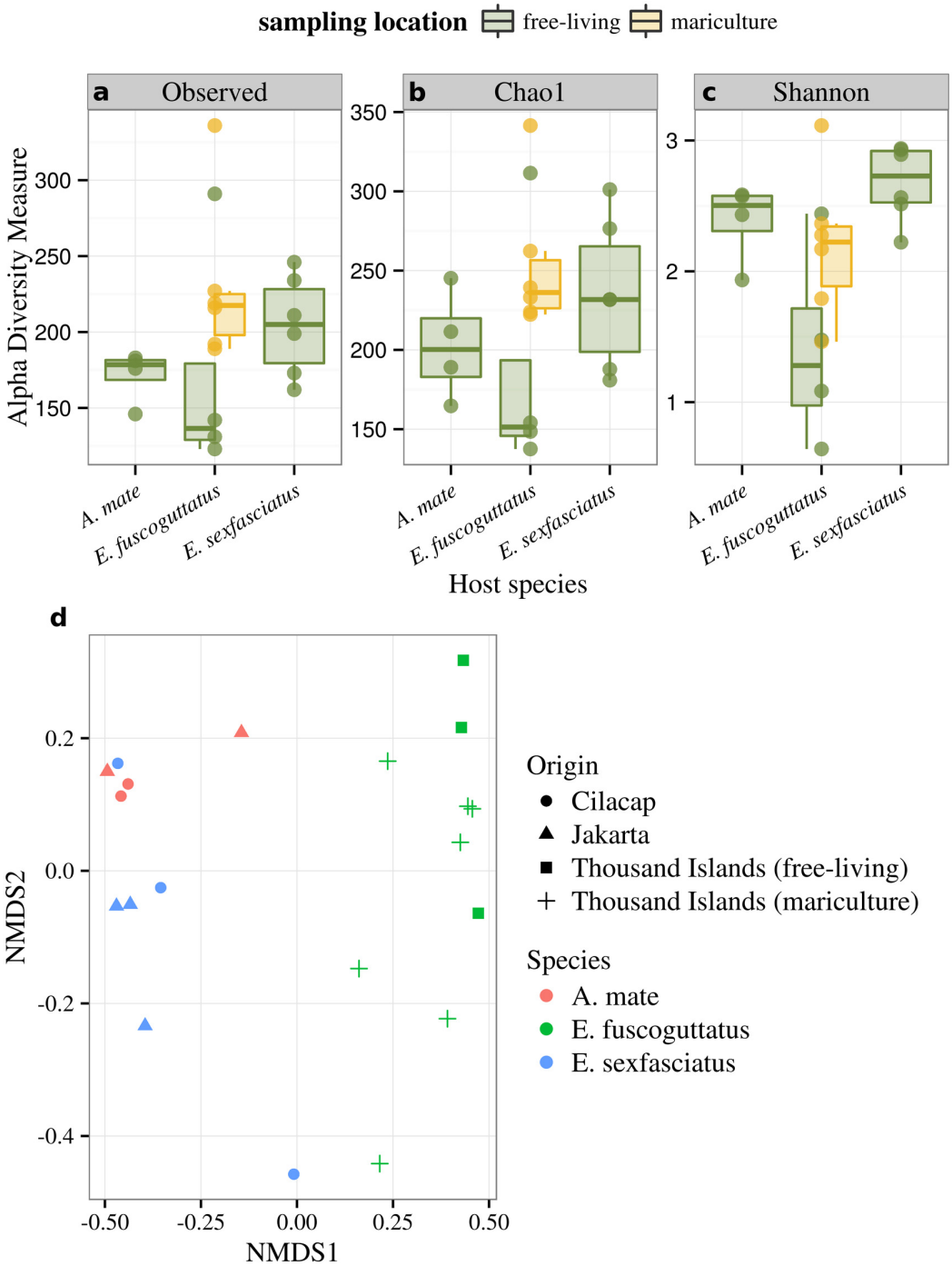


Fig 2. Alpha diversity and beta diversity estimates across fish species, sampling location and visualization of differences in bacterial gut community composition by host species. **a**: observed OTU richness; **b**: Chao1 index that estimates the true species richness of a sample; **c**: Shannon-Wiener diversity index accounting for species abundance and eveness of distribution. Dots represent estimates for individual samples, solid lines constitute the median, boxes the quartiles, and bars the interquartile range. **d**: Beta diversity is estimated with Nonmetric multidimensional scaling (NMDS) of bacterial communities derived from 20 fish specimen coloured by host species. Point shapes indicates differences in the sampling location. Samples derived from mariculture are labelled accordingly. Ordinations are based on between-sample dissimilarities calculated by Bray-Curtis distances.

doi:10.1371/journal.pone.0151594.g002

richness of a community including unobserved species [55], revealed a high difference between the observed and the expected OTU richness in the samples of *A. mate* (200 OTUs), *E. sexfasciatus* (231 OTUs) and *E. fuscoguttatus* from mariculture (236 OTUs). Only free-living samples had a small difference between expected and observed OTUs (151 OTUs) (Fig 2b). This implies that an even greater diversity of bacteria lies undiscovered. The ecological diversity, measured by Shannon-Wiener diversity index, revealed high bacterial diversity in the samples of *A. mate* (2.5) and *E. sexfasciatus* (2.7). In addition, the mariculture samples from *E. fuscoguttatus* displayed higher bacterial diversity (2.2) than the free-living samples (1.3) (Fig 2c). To measure the differences between the bacterial gut community compositions a nonmetric multidimensional scaling method was used to obtain ordinations based on between-sample dissimilarities calculated by Bray-Curtis distances [56]. The ordinations displayed two different clusters, whereas two samples (am1, es5) were outliers. One cluster was formed by *A. mate* and *E. sexfasciatus*. This implied a closer relationship of the microbial communities of *A. mate* and *E. sexfasciatus* than the communities of *E. fuscoguttatus*, which formed the other cluster (Fig 2d). Both fish species were collected from off Cilacap and inside Jakarta Bay, while *E. fuscoguttatus* originated from the Thousand Islands. This relationship stands in contrast to the phylogeny of the host species, whereas *E. fuscoguttatus* and *E. sexfasciatus* belonging to the same genus must be more closely related than to *A. mate* related at order level. Further analysis, using the statistical method adonis, confirmed the significance of these clusters with an p-value below 0.001.

Comparison between the microbiomes of free-living and mariculture *E. fuscoguttatus*

Comparing the fecal bacterial communities of *E. fuscoguttatus* over different sampling sites, inside the net cages and outside on the surrounding reef, revealed only minor differences in the composition of most abundant phyla. Proportions of *Proteobacteria* (avg free-living (free): 93.14% vs. avg mariculture (mari): 91.58%, s.d. free: 10.30% vs. s.d. mari: 6.10%) and *Spirochaetes* (avg free: 0.01% vs. avg mari: 1.52%, s.d. free: 0.02% vs. s.d. mari: 3.06%) differed slightly between free-living samples and samples from mariculture. Also minor differences in the proportion of *Fusobacteria* could be detected (avg free: 3.18% vs. avg mari: 4.06%, s.d. free: 6.30% vs. s.d. mari: 5.46%). The other dominating phyla appeared nearly in same ratios (Fig 1a).

Among the low abundance phyla no difference was detected based on differences in sampling sites. In each case one sample from inside (ef5) and outside the net cages (ef4) displayed a higher abundance of detected phyla, but this could not be assigned to different sampling sites. (Fig 1b). Differences between the bacterial gut community compositions of *E. fuscoguttatus* specimen showed that the samples from inside the net cages formed a subcluster within the cluster of *E. fuscoguttatus* (Fig 2d). Smaller distances between samples from inside pointed out a more conserved community structure in comparison to the samples from outside the net cages. Median observed OTU richness (Fig 2a) revealed a reduced OTU richness for the free-living samples of *E. fuscoguttatus* with 136 OTU. In contrast samples derived from inside the net cages showed a higher observed OTU richness with 217 OTUs, supported by Chao1 richness estimator showing a predicted number of 236 OTUs for samples from inside in contrast to 151 OTUs for the samples from outside (Fig 2b). In addition, Shannon-Wiener diversity index indicated a more diverse bacterial community structure for inside (2.22) than the free-living specimens (1.28) (Fig 2c).

Core and shared microbiomes

The comparison of shared OTUs revealed a different core microbiome for each host species (Fig 3). Core microbiome construction lead to a high number of shared OTUs for each of the

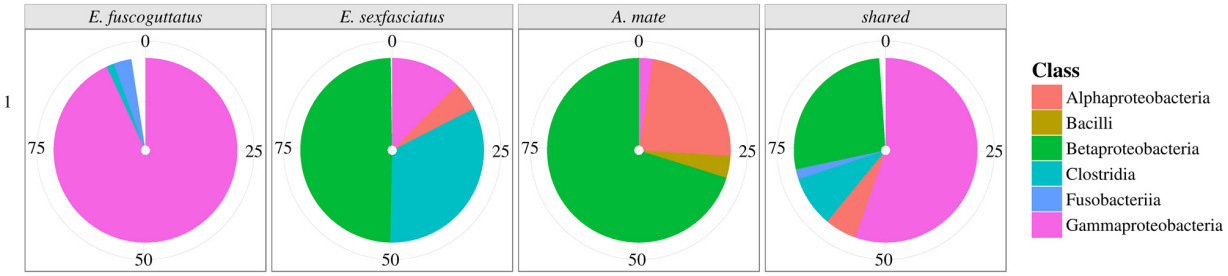


Fig 3. Core microbiome for each host species and shared microbiome on class level. The core microbiome is constructed by counting OTUs that are present in every sample of each host species. The shared microbiome results as a combination of all three core microbiomes, representing all OTUs in all samples of the three host fish species.

doi:10.1371/journal.pone.0151594.g003

three host species (*E. fuscoguttatus*: 106 OTUs, *E. sexfasciatus*: 129 OTUs and *A. mate*: 124 OTUs). Core microbiome of *E. fuscoguttatus*, consisting of 106 OTUs, was dominated by *Gammaproteobacteria* with over 92.59% whilst *Fusobacteria* (3.12%), *Clostridia* (1.35%) and *Betaproteobacteria* (1.00%) constituted the rest of the core microbiome. *E. sexfasciatus* revealed a completely different composition with 129 OTUs belonging to the core microbiome. The core was dominated by *Betaproteobacteria* with a relative abundance of 49.16%. Also a huge portion of *Clostridia* (32.67%) and a lower portion of *Gammaproteobacteria* (12.45%) and *Alphaproteobacteria* (4.97%) were revealed. Core microbiome of *A. mate* consisted of 124 OTUs, and was also dominated by *Betaproteobacteria* (69.84%) and had a large portion of *Alphaproteobacteria* (23.40%). The rest of it was distributed to *Bacilli* (3.82%) and *Gammaproteobacteria* (2.43%). Out of the three different core microbiomes a shared microbiome was constructed, by counting only OTUs present in every sample of the three host species. Thereby the resulted shared microbiome was dominated by *Gammaproteobacteria* (55.08%) and *Betaproteobacteria* (27.07%). *Clostridia* (8.91%), *Alphaproteobacteria* (5.57%), *Fusobacteria* (1.76%), as well as *Bacilli* (0.93%) and *Brevinematae* (0.50%) formed the rest of this shared microbiome.

Parasites

Fish parasitological studies on *E. fuscoguttatus*, *E. sexfasciatus* and *A. mate* from Thousand Islands (Pulau Seribu), Cilacap and Jakarta (Table 1), revealed 28 different parasite species belonging to the following taxa: 10 Digenea, 4 Monogenea, 1 Cestoda, 7 Nematoda, 2 Acanthocephala, 1 Hirudinea and 3 Crustacea (Table 3). In an additional study, using a shotgun sequencing approach on *Epinephelus fuscoguttatus*, all observed parasites were confirmed [57]. Data on prevalence, intensity, mean intensity and mean abundance of the collected parasite species for each fish species are summarized in Table 3. Parasite species richness of up to 12 taxa, calculated and pooled in the fish samples for both sites (Jakarta Bay and Cilacap) was highest in *A. mate* followed by *E. sexfasciatus* with nine taxa. *E. fuscoguttatus* from both sampling locations in the Thousand Islands had only seven taxa, with only five species from the fish in the net cages and seven from the fish caught in the reef beside the net cages.

To analyze parasite composition at each sampling site, the Shannon-Wiener diversity index suggested as an ecological parameter by Palm *et al.* [7] and Palm & Rueckert [9] were calculated (Table 2). Highest total Shannon-Wiener diversity was given in *A. mate* from Jakarta (1.44) followed by *E. sexfasciatus* from Cilacap (1.30) while lowest total diversity was observed in *A. mate* from Cilacap (0.45). Highest endoparasite Shannon-Wiener diversity was seen in *A. mate* in Jakarta (1.12), medium in *E. sexfasciatus* from both samples (0.98 vs. 0.62) and low in *A. mate* from Cilacap (0.45), *E. fuscoguttatus* from the surroundings of the net cages (0.30)

Table 3. Parasitic load.

		Atule mate Jakarta				*Atule mate* Cilacap				*E. fuscoguttatus* free PS				*E. fuscoguttatus* caged PS				*E. sexfasciatus* Cilacap				*E. sexfasciatus* Jakarta			
parasite taxa	**site**	**P[%]**	**MI**	**(I)**	**MA**	**P[%]**	**MI**	**(I)**	**MA**	**P[%]**	**MI**	**(I)**	**MA**	**P[%]**	**MI**	**(I)**	**MA**	**P[%]**	**MI**	**(I)**	**MA**	**P[%]**	**MI**	**(I)**	**MA**
Didymodiclinus **sp. (D)**	**fin, skin**																	100.0	5.7	(5–6)	5.7				
Echinostoma **sp. (D)**	**in., st.**	33.3	1.0	(1)	0.3																				
Genolinea **sp. (D)**	**st.**	66.7	7.5	(3–12)	5.0																				
Prosorhynchus cf. australis **(D)**	**in., pyl.**									20.0	2.0	(2)	0.4												
Prosorhynchus crucibulum **(D)**	**in.**									40.0	1.0	(1)	0.4												
Prosorhynchus **sp. (D)**	**in., pyl., st.**																	100.0	29.3	(9–55)	29.3	66.7	4.5	(2–7)	3.0
Pseudopecoelus **sp. (D)**	**mes. of swb.**									20.0	235.0	(33–437)	94.0	14.3	1.0	(1)	0.1								
Didymozoidae indet. **(D)**	**in., st.**	66.7	44.0	(5–66)	29.3	33.3	1.0	(1)	0.3																
Digenea indet. I (D)	**st.**	66.7	2.5	(2–3)	1.7																				
Digenea indet. II (D)	**in.**	33.3	1.0	(1)	0.3																				
Haliotrema cf. epinepheli **(M)**	**gi., gicv.**																	100.0	14.7	(9–25)	14.7	100.0	36.0	(27–41)	36.0
Pseudorhabdosynochus epinepheli **(M)**	**gi.**									60.0	3.0	(1–6)	1.8	85.7	9.5	(1–22)	8.1								
Pseudorhabdosynochus lantauensis **(M)**	**gi., gicv., op.**									100.0	13.4	(2–34)	13.4	85.7	6.8	(1–19)	5.9								
Mazocraeidea indet. (M)	**gi.**	33.3	1.0	(1)	0.3																				
Callitetrarhynchus gracilis **(C)**	**bcv., mes., gicv., go., li., stw.**	100.0	7.0	(4–9)	7.0																				
Camallanus **sp. I (N)**	**in.**	66.7	1.0	(1)	0.7																				
Camallanus **sp. II (N)**	**pyl.**																	33.3	1.0	(1)	0.3	33.3	1.0	(1)	0.3
Capillaria **sp. (N)**	**gi.**																	33.3	1.0	(1)	0.3	66.7	1.0	(1)	0.7
Hysterothylacium **sp. I (N)**	**in., st.**	66.7	2.0	(1–3)	1.3																				
Hysterothylacium **sp. II (N)**	**bcv. (mes.), go., li., stw.**																	100.0	45.0	(11–101)	45.0	100.0	21.0	(12–35)	21.0
Philometra **sp. (N)**	**bcv., go., li.**																	100.0	1.3	(1–2)	1.3				
Raphidascaris **sp. (N)**	**bcv. in fat; stw.**									40.0	1.0	(1)	0.4	28.6	3.0	(1–5)	0.9								
Rhadinorhynchus lintoni **(A)**	**in.**				66.7	2.5	(2–3)	1.7																	
Serrasentis sagittifer **(A)**	**mes. of go., in., st.**																					33.3	1.0	(1)	0.3
Zeylanicobdella arugamensis **(H)**	**gicv., fin**									60.0	1.6	(3–4)	2.7	85.7	1.7	(1–3)	1.4								
Caligus **sp. (C)**	**mcv.**	100.0	2.7	(1–4)	2.7																				
Lepeophtheirus **sp. (C)**	**gi.**																	33.3	2.0	(2)	0.7				
Bomolochidae indet. (C)	**gi.**	33.3	2.0	(2)	0.7																				
	ectoparasites		3				0				3				3				2				1		
	endoparasites		8				2				4				2				6				5		
	Ec/En ratio		0.4				0.0				0.8				1.5				0.3				0.2		

The prevalence [%], intensity (I), mean intensity (MI) and mean abundance (MA) of ectoparasites and endoparasites from *E. fuscoguttatus*, *E. sexfasciatus* and *A. mate* from different Indonesian waters. Additionally given is the amount of ecto- and endoparasite species as well as the Ec/En ratio. bcv: body cavity, gi: gills, gicv: gill cavity, go: gonads, in: intestine, li: liver, mes: mesenteries, mcv: mouth cavity, pyl: pylorus, st: stomach, stw: stomach wall; A: Acanthocephala, C: Cestoda, Cr: Crustacea, D: Digenea, H: Hirudinea, M: Monogenea, N: Nematoda

doi:10.1371/journal.pone.0151594.t003

respectively inside the net cages (0.20) (Table 2). Ecto-/endoparasite ratios calculated by using the number of ectoparasite species vs. the number of endoparasite species ranged from 0.0 (*A. mate*, Cilacap) up to 1.5 (*E. fuscoguttatus*, Thousand Islands from net cages) (Table 2). Additional ecological parameters, such as the hepatosomatic index and condition factors were calculated (Table 2) and showed highest HSI values in *E. fuscoguttatus* and *A. mate* (cultured at Thousand Islands respectively Cilacap, both = 1.4). All other HSI values ranged between 0.6 and 1.1. The condition factor was highest in both samples from *E. fuscoguttatus* from Thousand Islands (1.69–2.03).

Correlation between pathogenic bacteria and parasites

A Spearman's rank-order correlation between commonly known fish pathogenic bacteria, selected from microbiome analysis and recorded parasite number, in part also reflect the observed biodiversity (Fig 4). The highest parasite numbers were observed for free-living *E. sexfasciatus* from Cilacap (es4, es6), followed by *A. mate* and *E. sexfasciatus* from Jakarta Bay, mariculture and free-living *E. fuscoguttatus* from Thousand Islands. *A. mate* from Cilacap displayed virtually no parasite infection. All fish with a high number of parasites (above 50 individual metazoans) had no potentially pathogenic *Vibrio* sp., *Flavobacterium* sp. or *Photobacterium* sp. This was supported by the results of a Spearman's rank-order correlation test, which revealed a medium negative correlation for *Vibrio* sp. ($\rho = -0.4592765$, $p = 0.04164$) and *Photobacterium* sp. ($\rho = -0.4429808$, $p = 0.05045$). On the other hand, the highest *Vibrio* sp. counts were found in *E. fuscoguttatus* from inside the net cages (ef10, ef5) and, to a much lower degree, in *E. fuscoguttatus* outside the net cages (ef1) from the surrounding reef. An increased detected value for *Flavobacterium* sp. was only recorded from a fish inside the net cages without metazoan parasites (ef5), resulting in a weak positive correlation ($\rho = 0.1329735$, $p = 0.05762$) with a high p-value. *Photobacterium* sp. could only be recorded from *E. fuscoguttatus*, from free-living and mariculture fish from Thousand Islands, without any record from Jakarta Bay and off Cilacap.

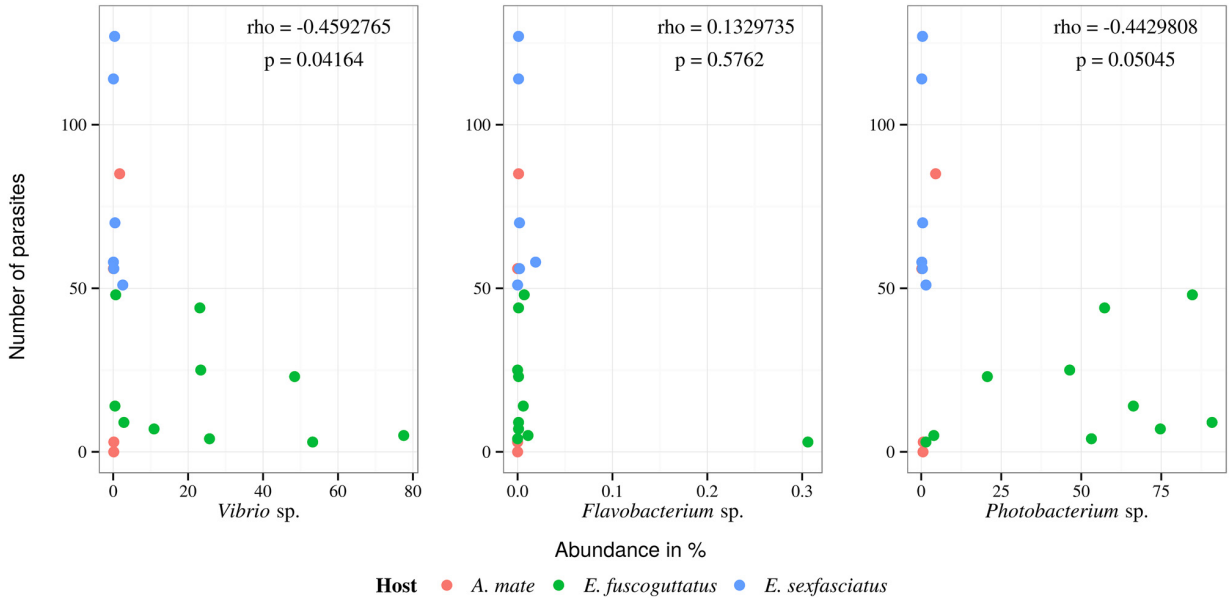


Fig 4. Spearman's rank-order correlation was performed to determine a relationship between the numbers of parasites against the abundance of three known fish pathogenic bacteria. It showed for *Vibrio* sp. ($\rho = -0.4592765$, $p = 0.04164$) and *Photobacterium* sp. ($\rho - 0.04429808$, $p = 0.05045$) a medium negative correlation and a weak positive correlation for *Flavobacterium* sp. ($\rho = 0.1329735$, $p = 0.5762$).

doi:10.1371/journal.pone.0151594.g004

Discussion

Our investigation of microbial composition reveals three phyla that are predominantly present across all three host species: *Proteobacteria*, *Firmicutes* and *Actinobacteria*. Particularly, *Proteobacteria* immensely dominates on phyla level in all samples, composed of *Gammaproteobacteria*, *Betaproteobacteria* and *Alphaproteobacteria* on the class level. In addition, *Firmicutes* and *Actinobacteria* was detected in all samples. In line with other metagenomic studies of fish microbiomes, these three phyla have been recognized as the characterstic components of the fish microbiome [16, 58]. For example, Asian sea bass has, *Proteobacteria* (48.8%), *Firmicutes* (15.3%), *Bacteroidetes* (8.2%) and *Fusobacteria* (7.3%) as the four most abundant bacterial phyla [15]. While members of the predominant bacterial communities at phylum level appears equal over all three host species and similar to earlier studied fish, the proportion of these communities is unique for every host species. This assumption is supported by beta diversity analysis (Fig 2d), which presents a distinct cluster for each of the three host species. Investigations on rare phyla (below 0.1%) results in altered phyla and proportions for every sample [59], with no detectable difference on sampling locations or host species.

Comparing samples derived from *E. fuscoguttatus* under mariculture and free-living conditions, we detected similar compositions of predominant bacterial communities. Further investigation of differences in bacterial community composition revealed that samples from mariculture formed a subcluster within free-living samples, indicating robustness of the bacterial communities within these samples compared to communities in free-living samples. In contrast, alpha diversity measurements exposed higher species richness for the mariculture samples, furthermore mean number of the expected species richness was higher compared to free-living samples. This indicated that the number and distribution of phyla within both conditions were the same, but deeper taxonomical levels revealed a more diverse bacterial community for the mariculture samples in contrast to the free-living ones.

The exploration of the core microbiomes resulted in three cores with similar members under different proportions. While the core microbiomes of *A. mate* and *E. sexfasciatus* consisted of a dominating parts of *Betaproteobacteria*, the core of *E. fuscoguttatus* is highly dominated by *Gammaproteobacteria* with over 90%, also a member in the other core microbiomes, but only with a portion of 12% (*E. sexfasciatus*) and 3% (*A. mate*). In addition, *Clostridia* was proportionally higher in *E. sexfasciatus*, but had negligible presence in *E. fuscoguttatus* and *A. mate*. This was also observed with *Alphaproteobacteria*, present in large proportions in *A. mate* compared to, *E. sexfasciatus* and none on *E. fuscoguttatus*. Three bacteria were only detected in one of the core microbiomes: *Fusobacteria*, *Brevinematae* in *E. fuscoguttatus* and *Bacilli* in *A. mate*. The shared microbiome derived by combining the three core microbiomes from the host species, resulted in the composition of all three core microbiomes, whereas *Gammaproteobacteria* (54.72%) and *Betaproteobacteria* (26.59%) dominated. These three different core microbiomes are supported by calculations from beta diversity analysis, showing three separate clusters, each consisting only of one host species including all environmental conditions (Fig 2d). Furthermore, core microbiomes showed that bacterial communities differ with host species. With a unique bacterial community per host species, the core microbiome is also unique.

Previous fish parasitological studies in Indonesian waters have revealed a rich species diversity, naming nearly 80 different taxa from mariculture groupers alone, belonging to the three genera *Epinephelus*, *Cromileptes* and *Plectropomus*. For cultured epinephelids in total 60 different parasite species were found. The highest parasite diversity was recorded for *E. fuscoguttatus* with 46 parasite species/taxa, 25 of which were ectoparasites and 21 were endoparasites. Another frequently cultivated fish, *Epinephelus coioides*, harbours 36 parasite species/taxa (21 ecto- and 15 endoparasites). While the lowest parasite diversity was found for *Epinephelus*

areolatus (three ectoparasites only) [10]. Independent data from *E. sexfasciatus* and *A. mate* from Indonesian waters so far are unavailable.

In this study, we record seven different parasite species for *E. fuscoguttatus* from Thousand Islands, five from within and seven from outside the net cages, comparatively less than reported from the same location in earlier studies [17]. In general, wild fish has been observed to be infected by fish parasites more than cultured fish [10, 11]. Palm *et al.*[7] used fish parasites to monitor long-term change in finfish grouper mariculture in Indonesia. A total of 210 *Epinephelus fuscoguttatus* were sampled in six consecutive years between 2003/04 and 2008/09 from the same mariculture facility and, using the same methodology, examined for parasites. While fish from inside the net cages in the first dataset had 14–16 different parasite species, this number decreased to eight in the rainy season 2008/09. Palm *et al.*[7] stated that the diminishing parasite richness over time may reflect changing environmental conditions at the site, from the initiation of mariculture activity (beginning of the parasite monitoring) until increased fish production six years later. However, the authors sampled only fish from net cages. In the present study, only five parasite species occurred inside the fish from the cages, reflecting a further decrease in parasite richness in rainy season 2012. Our data demonstrates that parasite richness at the present time is even further reduced. More importantly, fish from outside the net cages had only seven different parasite species. This fact strongly supports the notion that not only the feed within the mariculture facility but also environmental conditions must have changed during the last and present investigation.

Rueckert *et al.*[17] studied distinctly fed groupers, *E. coioides* from an Indonesian finfish mariculture farm for ecto- and endohelminth parasites. Pellet-fed *E. coioides* were infested with 13 parasite species/taxa of which six had a monoxenous (single host) and seven a heteroxenous (multi host) life cycle. A total of 14 parasite species/taxa were found in the fish that were fed with different trash fish species, four of them with a monoxenous and ten with a heteroxenous life cycle. The use of pellet food significantly reduced the transfer of endohelminths and the number of parasites with a heteroxenous life cycle. The risk of parasite transfer can be also reduced by feeding selected trash fish species with a lower parasite burden, using only trash fish musculature or minimizing the abundance of invertebrates (fouling) on the net cages. For *E. fuscoguttatus* Rueckert *et al.*[11] recorded a parasite infracommunity ranging from one to nine (cultured) and three to 14 parasite species (wild) also in Lampung Bay. In the present study, *E. fuscoguttatus* from the net cages had less parasites than those caught in the surrounding reef, however, at a low level.

The highest Shannon-Wiener diversity (total biodiversity) was revealed for *A. mate* from Jakarta Bay (1.4), followed by *E. sexfasciatus* from Cilacap (1.3). With respect to endoparasite diversity, we observe a trend with usually higher endoparasite diversity in the free-living epinephelids vs. the cultured *E. fuscoguttatus*, and for fish from Cilacap vs. fish from Jakarta Bay. This is also reflected by the Ec/En ratio from *E. fuscoguttatus* from cultured (1.5) compared with free-living (0.8) fish. The endoparasite diversity is of importance because under natural environmental conditions, the endoparasite richness inside the gut is regularly high and is used as bioindicator [60, 61]. Under polluted and heavily impacted environmental conditions, endoparasites lack the ability to complete their life cycles, and cannot be found in the studied fish. Consequently, our observation follows the general assumptions that the number of endoparasite species is low inside the mariculture fish as well as from polluted waters such as Jakarta Bay. The only exception here is *A. mate* that had similar high endoparasite diversity in Jakarta Bay and Cilacap. However, this might be caused by the small number of analyzed fish, or the migratory behavior that is known for this pelagic species.

According to our data, the number of fish parasites of wild fish exceeds that of mariculture fish [11, 42]. This coincided with the observation that tropical wild fish show fewer signs of

diseases, though potential pathogens can be regularly found in the environment. In contrast, bacterial disease outbreaks occur under aquaculture conditions where only few parasites occur. While the diet (in our case the use of trash fish) and/or the environment can influence the number of revealed endoparasites in the fish, the parasite infracommunity as well might influence the microbiome, and suppress the impact of pathogenic bacteria and subsequently disease outbreaks. Consequently, we would expect to observe differences in the microbiome of the sampled parasitized or less infected fish.

The results of the microbial communities enable the identification of three potentially pathogenic bacteria, i.e. *Vibrio* sp., *Flavobacterium* sp. and *Photobacterium* sp.. Comparing these results with the recorded parasite numbers using a Spearman's rank-order correlation test, shows a weak negative correlation for *Vibrio* sp. and *Photobacterium* sp. In case of *Flavobacterium* sp. a weak positive correlation could be detected. The highest number of parasites were observed for free-living *E. sexfasciatus* from Cilacap, followed by *A. mate* and *E. sexfasciatus* from Jakarta Bay, and free-living and mariculture *E. fuscoguttatus* from Thousand Islands. All highly parasitized fish (above 50 individual metazoans) had no potentially pathogenic *Vibrio* sp., *Flavobacterium* sp. or *Photobacterium* sp. Instead, highest *Vibrio* sp. counts were only found in *E. fuscoguttatus* from inside the net cages and, to a much lower degree, in *E. fuscoguttatus* outside the net cages from surrounding reef. *Flavobacterium* sp. was only recorded from a fish inside the net cages without metazoan parasites, and *Photobacterium* sp. was recorded only from *E. fuscoguttatus*, from free-living and mariculture fish from Thousand Islands, without any record from the other two sampled fish species from Jakarta Bay and off Cilacap. This coincides with our assumption that there is a positive influence of the metazoan parasite infection on fish health and the occurrence of potential pathogenic bacteria inside the fish. However, this requires verification in future studies with a larger sample size.

Conclusions

Notably the core microbiomes of both phylogenetically related and distant related fish species, *Epinephelus fuscoguttatus*, *Epinephelus sexfasciatus* and *Atule mate*, contained approximately the same classes of bacteria independent on the degree of pollution. However, the proportions of these bacterial classes strongly varied. The microbial biodiversity of two phylogenetically distant fish species, *A. mate* and *E. sexfasciatus* from Jakarta Bay and Cilacap were more closely related than those of the two phylogenetically adjacent species, *E. fuscoguttatus* and, *E. sexfasciatus* from Jakarta Bay, Cilacap and Thousand Islands. In addition, we detected weak negative correlation between the load of selected bacterial pathogens, *Vibrio* sp., *Photobacterium* sp. and the number of endoparasites. In the case of *Flavobacterium* sp. we found the opposite weak positive correlation. Of the three pathogenic bacterial genera, *Vibrio* sp. were found predominantly in *E. fuscoguttatus* from mariculture, and fewer in the vicinity of the net cages and rarely in fish from the heavily polluted waters from Jakarta Bay. *Flavobacterium* sp. showed highest counts inside mariculture fish and *Photobacteria* sp. was most prominent inside and close to the net cages. Due to our sample size, further study is required to make general statements concerning these findings, which are highly relevant for future finfish mariculture activities and management practices.

Supporting Information

S1 Fig. Number of raw sequence reads, merged paired reads, post-QA/QC sequence reads, and number of taxonomically classified reads. The light grey and dark grey dashed lines represent the average number of raw sequence reads and taxonomically classified sequence reads across all samples.
(TIFF)

S2 Fig. Rarefaction curves.
(TIFF)

Author Contributions

Conceived and designed the experiments: PH GM HPS HWP ST SK. Performed the experiments: GM ST SK. Analyzed the data: PH ST SK. Contributed reagents/materials/analysis tools: PH ST SK. Wrote the paper: PH MAA HPS HWP ST SK.

References

1. Dsikowitzky L, Nordhaus I, Jennerjahn TC, Khrycheva P, Sivatharshan Y, Yuwono E, et al. Anthropogenic organic contaminants in water, sediments and benthic organisms of the mangrove-fringed Segara Anakan Lagoon, Java, Indonesia. Marine Pollution Bulletin. 2011; 62: 851–862. doi: 10.1016/j.marpolbul.2011.02.023 PMID: 21414637
2. Rinawati, Koike T, Koike H, Kurumisawa R, Ito M, Sakurai S, et al. Distribution, source identification, and historical trends of organic micropollutants in coastal sediment in Jakarta Bay, Indonesia. Journal of Hazardous Materials. 2012; 217–218: 208–216. doi: 10.1016/j.jhazmat.2012.03.023 PMID: 22480710
3. Cooper TF, Gilmour JP, Fabricius KE. Bioindicators of changes in water quality on coral reefs: review and recommendations for monitoring programmes. Coral Reefs. 2009; 28: 589–606. doi: 10.1007/s00338-009-0512-x
4. Eng CT, Paw JN, Guarin FY. The environmental impact of aquaculture and the effects of pollution on coastal aquaculture development in Southeast Asia. Marine Pollution Bulletin. 1989; 20: 335–343. doi: 10.1016/0025-326X(89)90157-4
5. Nordhaus I, Hadipudjana FA, Janssen R, Pamungkas J. Spatio-temporal variation of macrobenthic communities in the mangrove-fringed Seagara Anakan lagoon, Indonesia, affected by anthropogenic activities. Regional Environmental Change. 2009; 9: 291–313. doi: 10.1007/s10113-009-0097-5
6. Carpenter KE, Springer GV. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environmental Biology of Fishes. 2005; 72: 467–480. doi: 10.1007/s10641-004-3154-4
7. Palm HW. Fish parasites as biological indicators in a changing world: can we monitor environmental impact and climate change? In: Mehlhorn H, Progress in Parasitology. Springer Berlin Heidelberg 2011; 2: pp. 223–250.
8. Palm HW, Kleinertz S, Ruckert S. Parasite diversity as an indicator of environmental change?—an example from tropical grouper (*Epinephelus fuscoguttatus*) mariculture in Indonesia. Parasitology. 2011; 138: 1–11. doi: 10.1017/S0031182011000011
9. Palm HW, Ruckert S. A new approach to visualize fish and ecosystem health by using parasites. Parasitology Research. 2009; 105: 539–553. doi: 10.1007/s00436-009-1423-z
10. Palm HW, Yulianto I, Theisen S, Rueckert S, Kleinertz S. *Epinephelus fuscoguttatus* mariculture in Indonesia: implications from fish parasite infections. Regional Studies in Marine Science. 2015; (in press) doi: 10.1016/j.rsma.2015.07.003
11. Rueckert S, Klimpel S, Palm HW. Parasites of cultured and wild brown-marbled grouper *Epinephelus fuscoguttatus* (Forsskål, 1775) in Lampung Bay, Indonesia. Aquaculture Research. 2010; 41: 1158–1169. doi: 10.1111/j.1365-2109.2009.02403.x
12. Brown K, DeCoffe D, Molcan E, Gibson DL. Diet-induced sysbiosis of the intestinal microbiota and the effects on immunity and sisease. Nutrients. 2012; 4: 1095–1119. doi: 10.3390/nu4081095 PMID: 23016134
13. Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, et al. Evidence for a core gut microbiota in the zebrafish. The ISME journal. 2011; 5: 1595–608. doi: 10.1038/ismej.2011.38 PMID: 21472014
14. Givens CE, Ransom B, Bano N, Hollibaugh JT. Comparison of the gut microbiome of 12 bony fish and 3 shark species. Mar Ecol Prog Ser. 2015; 518: 209–223. doi: 10.3354/meps11034
15. Xia J, Lin G, Fu GH, Wan ZY, Lee M, Wang L, et al. The intestinal microbiome of fish under starvation. BMC Genomics. 2014; 15(1): 266. doi: 10.1186/1471-2164-15-266 PMID: 24708260
16. Rueckert S, Palm HW, Klimpel S. Parasite fauna of seabass (*Lates calcarifer*) under mariculture conditions in Lampung Bay, Indonesia. J Appl Ichthyol. 2008; 25: 321–327. doi: 10.1111/j.1439-0426.2008.01064.x

17. Rueckert S, Klimpel S, Al-Quraishy S, Mehlhorn H, Palm HW. Transmission of fish into grouper mariculture (Serranidae: *Epinephelus coioides* (Hamilton, 1822)) in Lampung Bay, Indonesia. Parasitology Research. 2009; 104: 523–532. doi: 10.1007/s00436-008-1226-7

18. van der Meij SET, Moolenbeek RG, Hoeksema BW. Decline of the Jakarta Bay molluscan fauna linked to human impact. Marine Pollution Bulletin. 2009; 59(4–7): 101–107. doi: 10.1016/j.marpolbul.2009.02.021 PMID: 19342065

19. van der Meij SET, Suharsono, Hoeksema BW. Long-term changes in coral assemblages under natural and anthropogenic stress in Jakarta Bay (1920–2005). Marine Pollution Bulletin. 2010; 60(9): 1442–1454. doi: 10.1016/j.marpolbul.2010.05.011 PMID: 20615515

20. Ruckert S, Hagen W, Yuniar AT, Palm HW. Metazoan parasites of Segara Anakan Lagoon, Indonesia, and their potential use as biological indicators. Regional Environmental Change. 2009; 9: 315–328. doi: 10.1007/s10113-008-0076-2

21. Palm HW, Bray RA. Marine Fish Parasitology in Hawaii. Westarp; 2014.

22. Anderson RC. Nematode parasites of vertebrates, their development and transmission. CABI Publishing; 2002.

23. Boxshall GA, Halsey SH. An introduction to copepod diversity Volume 1 & 2. London: The Ray Society; 2004.

24. Bu SSH, Leong TS, Wong SY, Woo YSN, Foo RWT. Three diplectanid monogeneans from marine finfish (*Epinephelus* sp.) in the Far East. Journal of Helminthology. 1999; 73: 301–312.

25. Bray RA, Gibson DI, Jones A. Keys to the Trematoda Volume 3. London: The Natural History Museum; 2008.

26. Bray RA, Palm HW. Bucephalids (Digenea: *Bucephalidae*) from marine fishes off the south-western coast of Java, Indonesia, including the description of two new species of *Rhipidocotyle* and comments of the marine fish digenean fauna of Indonesia. Zootaxa. 2009; 2223: 1–24. doi: 10.11646/zootaxa.2223.1.1

27. De Silva PHDH. Zeylanicobdella arugamensis gen. nov. and sp. nov. from Arugam Kalapu, Eastern Province, Ceylon. Spolia Zeylanica. 1963; 30: 47–53.

28. Dewi K, Palm HW. Two new species of philometrid nematodes (Nematoda: *Philometridae*) in *Epinephelus coioides* (Hamilton, 1822) from the South Bali Sea, Indonesia. Zootaxa. 2013; 3609: 49–59. doi: 10.11646/zootaxa.3609.1.3 PMID: 24699571

29. Gibson DI, Jones A, Bray RA. Keys to the Trematoda Volume 1. London: The Natural History Museum; 2002.

30. Golvan YJ. Systématique des Acanthocéphales (Acanthocephala Rudolphi 1801). Paris: Editions du Mus éum; 1969.

31. Jones A, Bray RA, Gibson DI. Keys to the Trematoda Volume 2. London. The Natural History Museum; 2005.

32. Moravec F. Nematodes of freshwater fishes of the neotropical region. Praha: Academia; 1989.

33. Palm HW. The Trypanorhyncha Diesing, 1863. Bogor: PKSPL-IPB Press; 2004.

34. Pillai NK. The fauna of India. Copepod parasites of marine fishes. Kalkutta: Zoological Society of India; 1985.

35. Whittington ID. The Capsalidae (Monogenea: Monopisthocotylea): a review of diversity, classification and phylogeny with a note about species complexes. Folia Parasitologica. 2004; 51: 109–122. doi: 10.14411/fp.2004.016 PMID: 15357390

36. Yamaguti S. Parasitic worms mainly from Celebes. Part 1. New digenetic trematodes of fishes. Acta Medicinae Okayama. 1952; 8: 146–198.

37. Yamaguti S. New parasitic copepods from Japanese fishes. Publications of the Seto Marine Biological Laboratory. 1960; 8: 141–155.

38. Yamaguti S. Systema Helminthum, Volume IV: Monogenea and Aspidocotylea. London: Interscience Publishers; 1963.

39. Yamaguti S. Systema Helminthum. Volume V: Acanthocephala. New York, London: Interscience Publishers John Wiley & Sons; 1963.

40. Bush O, Lafferty AD, Lotz JM, Shostak AW. Parasitology meets ecology on his own terms: Margolis et al revisited. Journal of Parasitology. 1997; 83: 575–583. doi: 10.2307/3284227 PMID: 9267395

41. Kleinertz S, Palm HW. Fish parasites of *Epinephelus coioides* (Serranidae) as potential environmental indicators in Indonesian coastal ecosystems. Journal of Helminthology. 2015; 89(1) 86–99.

42. Kleinertz S, Damriyasa M, Hagen W, Theisen S, Palm HW. An environmental assessment of the parasite fauna of the reef- associated grouper *Epinephelus areolatus* from Indonesian waters. Journal of Helminthology. 2014; 88: 50–63. doi: 10.1017/S0022149X12000715 PMID: 23095724

43. Shannon CE, Weaver W. The Mathematical Theory of Communication. Urbana: University of Illinois Press; 1949.

44. Magurran AE. Ecological diversity and its measurement. London: Croom Helm; 1988.

45. Munkittrik KR, Van der Kraak GJ, McMaster ME, Portt DCB, Van den Heuval MR, Servos MR. Survey of receiving-water environmental impacts associated with discharges from pulp mills. II. Gonad size, liver size, hepatic EROD activity and plasma sex steroid levels in white sucker. Environmental Toxicology and Chemistry. 1994; 13: 1089–1101. doi: 10.1002/etc.5620130710

46. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences. 2011; 108: 4516–4522. doi: 10.1073/pnas.1000080107

47. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012; 6: 1621–4. doi: 10.1038/ismej.2012.8 PMID: 22402401

48. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods. 2013; 10: 996–8. doi: 10.1038/nmeth.2604 PMID: 23955772

49. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011; 27: 2194–2200. doi: 10.1093/bioinformatics/btr381 PMID: 21700674

50. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology. 2006; 72: 5069–72. doi: 10.1128/AEM.03006-05 PMID: 16820507

51. R Core Team. R: A Language and Environment for Statistical Computing.; 2014. Available from: http://www.r-project.org/.

52. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013; 8: e61217.14.

53. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11: R106. doi: 10.1186/gb-2010-11-10-r106 PMID: 20979621

54. Hudson PJ, Dobson AP, Lafferty KD, Oksanen J, Guillaume BF, Kindt R, et al. vegan: Community Ecology Package.; 2016. Available from: https://CRAN.R-project.org/package=vegan

55. Chao A. Nonparametric Estimation of the Number of Classes in a Population. Scandinavian Journal of Statistics. 1984; 11(4): 265–270.

56. Bray JR, Curtis JT. An ordination of upland forest communities of southern Wisconsin. Ecological Monographs. 1957; 27: 325–349. doi: 10.2307/1942268

57. Hennersdorf P, Mrotzek G, Muslihudeen AA, Saluz HP. Metagenomic analysis between free-living and cultered *Epinephelus fuscoguttatus* under different environmental conditions in indonesian waters. Marine Pollution Bulletin; 2015 (in press).

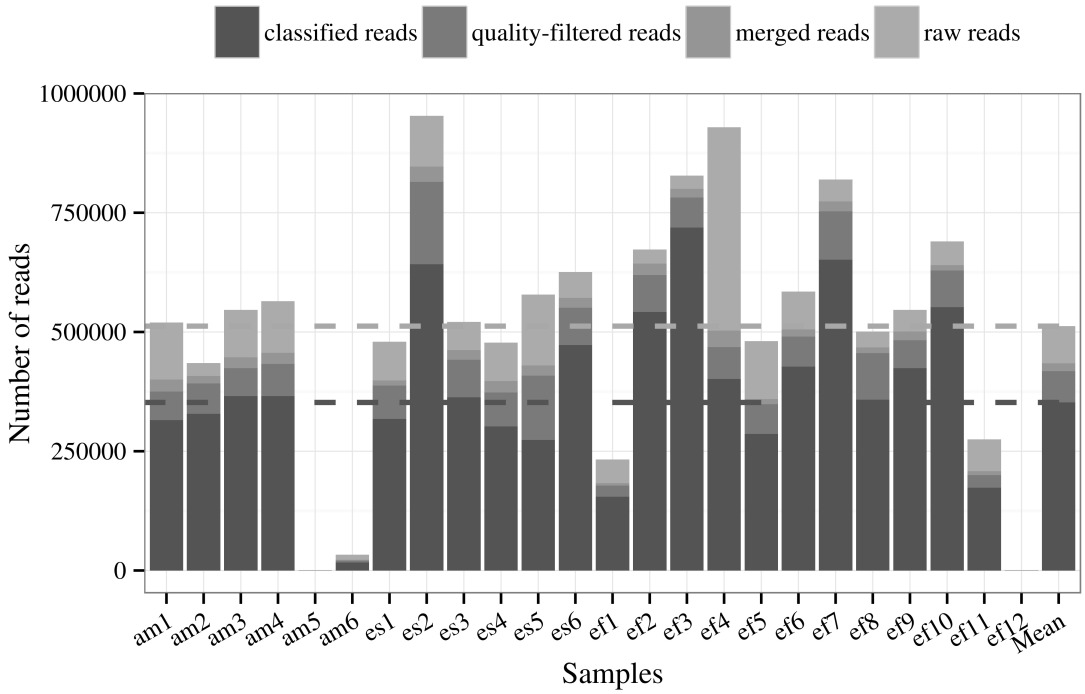
58. Sevellec M, Pavey SA, Boutin S, Filteau M, Derome N, Bernatchez L. Microbiome investigation in the ecological speciation context of lake whitefish (*Coregonus clupeaformis*) using next-generation sequencing. Journal of Evolutionary Biology. 2014; 1029–46. doi: 10.1111/jeb.12374 PMID: 24773032

59. Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight R, et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Molecular Ecology. 2012; 21: 3363–3378. doi: 10.1111/j.1365-294X.2012.05552.x PMID: 22486918

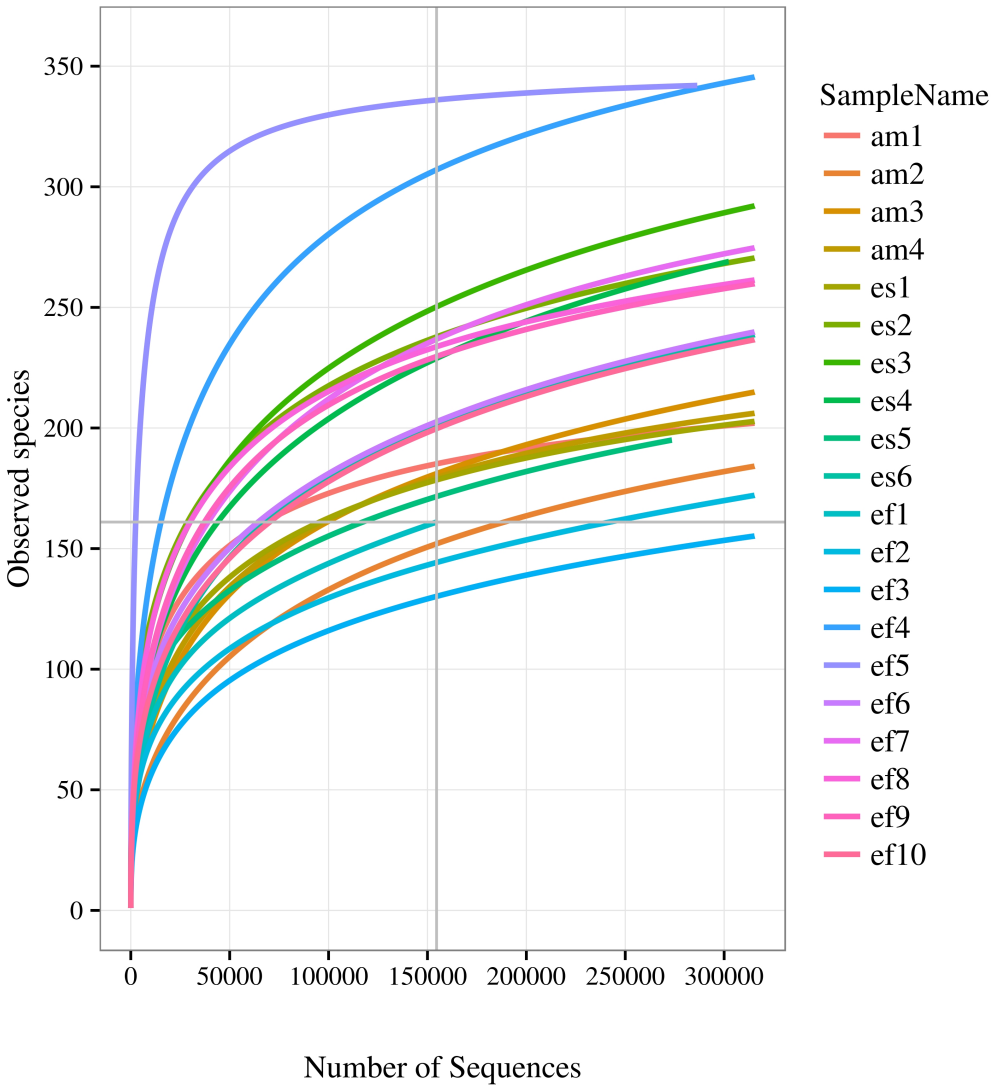
60. Hechinger RF, Lafferty KD, Huspeni TC, Andrew JB, Armand MK. Can parasites be indicators of free-living diversity? Relationships between species richness and the abundance of larval trematodes and of local benthos and fishes. Oecologia. 2007; 151: 82–92. doi: 10.1007/s00442-006-0568-z PMID: 17024376

61. Hudson PJ, Dobson AP, Lafferty KD. Is a healthy ecosystem one that is rich in parasites? Trends in Ecology and Evolution. 2006; 21: 381–385.

Supplement



SFig 1: Number of raw sequence reads, merged paired reads, post-QA/QC sequence reads, and number of taxonomically classified reads. The light grey and dark grey dashed lines represent the average number of raw sequence reads and taxonomically classified sequence reads across all samples.



SFig 2: Rarefaction curves.

Metagenomic analysis between free-living and cultured *Epinephelus fuscoguttatus* under different environmental conditions in Indonesian waters

Hennersdorf P, Mrotzek G, Abdul-Aziz MA, Saluz HP **(2016)**

Marine Pollution Bulletin 110 (2), 726-734
doi: 10.1016/j.marpolbul.2016.05.009.

Marine Pollution Bulletin 110 (2016) 726–734

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Metagenomic analysis between free-living and cultured *Epinephelus fuscoguttatus* under different environmental conditions in Indonesian waters

Philipp Hennersdorf [a,b], Grit Mrotzek [a], Muslihudeen A. Abdul-Aziz [c], Hans Peter Saluz [a,b,*]

[a] Leibniz-Institute for Natural Product Research and Infection Biology, Beutenbergstr. 11a, D-07745 Jena, Germany
[b] Friedrich Schiller University of Jena, Fürstengraben 1, D-07743 Jena, Germany
[c] Australian Centre for Ancient DNA, University of Adelaide, Adelaide, South Australia, Australia

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form 29 April 2016
Accepted 5 May 2016
Available online 20 May 2016

Keywords:
Fish
Epinephelus fuscoguttatus
Jakarta
Microbiome
Metagenome
Functional annotation

ABSTRACT

In this study, we analyzed and compared feces of free-living and cultivated fish species, *Epinephelus fuscoguttatus* under different environmental conditions in Indonesian waters. Metagenome analysis was performed using Illumina MiSeq sequencing of the whole metagenomic DNA isolated from fish feces samples. The analysis covered both prokaryotic and eukaryotic DNA. Feces samples from mariculture fish revealed a highly stable distribution of several orders of bacteria when compared to samples from free-living fish, which were highly diverse and dominated by *Vibrionales*, *Pseudomonales*, *Rhizobiales* and non-classifiable *Alphaproteobacteria*. The eukaryotic content of the samples was dominated by residues of the host and nine additional fish species that formed a portion of the diet. Investigations on functional annotations for predominant bacterial taxa, using Gene Ontology enrichment, revealed a number of functions related to DNA metabolic processes, especially DNA repair, as well as antibiotic response in the free-living fish species.

1. Introduction

Indonesia with its myriad of islands is located within the coral triangle, an area with marine biodiversity that exceeds that of all other places on earth (Carpenter and Springer, 2005; Hutomo and Moosa, 2005). Fishery has always been an important economic factor for the country, based on the natural richness of highly valued fish species and ideal aquaculture conditions (Pet-Soede et al., 1999). However, population growth and rapid economic development has led to the depletion of fishing stock due to over exploitation. In addition, the growth in aquaculture related activities has resulted in an increase in wastewater production. This wastewater is poorly treated and discharged into the environment (Dsikowitzky et al., 2011) resulting in an immensely negative influence on marine ecosystems of Indonesian coastal waters and its inhabitants (Rinawati et al., 2012). Principally, in the region of Jakarta Bay, Northern Jakarta, a booming coastal megacity with over nine million inhabitants, that has been affected by increasing pollution load caused by untreated wastewater originating from households and industry (van der Meij et al., 2009, 2010; Siregar and Koropitan, 2013). Likewise, capture fisheries on Jakarta Bay have an influence on the local fish communities and their environment (Dsikowitzky et al., 2011; Cooper et al., 2009; Eng et al., 1989; Nordhaus et al., 2009). The consequences of these factors on the microbiome and especially the metagenome of fish are still unclear.

The microbiome of various fish species has been recently investigated by a number of studies. Recent findings have detected a core microbiome that appears to be common for certain fish species (Roeselers et al., 2011; Givens et al., 2015). Furthermore, Xia et al. (Xia et al., 2014) revealed the influence of starvation on the microbiome of fish by using a combined analysis on the microbiome and the functions of the intestinal tract. They recorded 33 phyla, 66 classes, 130 orders and 278 families of the intestinal microbiome using 16S amplicon sequencing and revealed that *Proteobacteria* (48.8%), *Firmicutes* (15.3%) and *Bacteroidetes* (8.2%) appeared as the predominant taxa. Specifically, *Proteobacteria* and *Firmicutes* have been widely reported as the characteristically predominant taxa for fish intestinal and gut microbiomes (Sullam et al., 2012; Sevellec et al., 2014; Xing et al., 2013). In contrast, the elucidation of functional annotations for the metagenomic samples of fish gut has only recently begun, resulting in a dearth of studies at this point in time (Xing et al., 2013). Furthermore, the microbiome of mariculture and free-living fish, and particularly a comparison between fish from these two different environments are yet to be studied. The brown-marbled grouper *Epinephelus* (*E.*) *fuscoguttatus* (order Perciformes, family Serrinadae) is a benthic marine fish species, widely distributed over the Indo-Pacific region and represents an important Indonesian fish species. *E. fuscoguttatus*, is a protogynous hermaphrodite and begins its lifecycle as female and changes its sex to male at later ages (Sugama et al., 2012). It is cultured in hatcheries, as well as extensively sourced from the wild. Adults or large juveniles are marketed directly and small juveniles are grown out to market size. In contrast to the difficult conditions for culturing, the species

* Corresponding author at: Leibniz-Institute for Natural Product Research and Infection Biology, Beutenbergstr. 11a, D-07745 Jena, Germany.

http://dx.doi.org/10.1016/j.marpolbul.2016.05.009
0025-326X/

represents an important commercial fish species in Indonesia. However, the destruction of seagrass beds, vital for juveniles, and coral reefs, as well as intensive fishing has led to a near threatened status on the red list for this species (Cornish). In this study we collected 12 samples of *E. fuscoguttatus*, living under free-living and mariculture conditions, from Pulau Seribu, a chain of islands located to the North of Jakarta Bay. This area, stretching 45 km in the North into the Java Sea, consists of 110 islands and is part of the Thousand Island Marine National Park. We extracted total DNA from fish feces and sequenced it with Illumina MiSeq technology, using a whole metagenomic shotgun sequencing approach. We aimed at the identification and comparison of the taxonomical composition of the microbiome, particularly the predominant taxa and biodiversity, elucidating the functional components of these bacterial taxa, as well as the eukaryotic components within the fecal samples, providing information on food components and potential residues of parasitic sequences. (See Table 1.)

2. Results

2.1. Clustering of bacterial taxa for different environmental conditions

In total 12 samples, distributed over two environmental conditions, five free-living ($n = 5$) and seven mariculture ($n = 7$) samples were used to analyze bacterial composition. To test for the presence of two distinct environmental conditions, the samples were clustered, using a hierarchical clustering method with average linkage method (Supplementary Fig. S1). The clustering shows two distinct clusters representing the two different environmental conditions and one outlier (free2). While the first cluster consists only of samples from mariculture, the other consisted of free-living samples, except the outlier free2, and two mariculture samples (mari2, mari3). Using the statistical test Adonis, provided by R package vegan (Oksanen et al., 2016), we confirmed the composition of these clusters with a p-value of 0.034. As a result the samples free2, mari2 and mari3 which were statistically insignificant were excluded from further analysis.

2.2. Elucidation of bacterial communities under different environmental conditions

The annotated bacterial sequences of the remaining nine taxa revealed 16 different phyla, in decreasing order of abundance: *Proteobacteria* (71.00%), *Firmicutes* (23.71%), *Spirochaetes* (3.76%), *Actinobacteria* (1.13%), *Bacteroidetes* (0.40%), *Tenericutes* (0.07%), *Fusobacteria*, *Cyanobacteria*, *Deinococcus-Thermus*, *Verrucomicrobia*, *Acidobacteria*, *Lentisphaerae*, *Chlorobi*, *Aquificae*, *Chlamydiae* and *Synergistetes* (<0.01%).

Whereas, all samples were dominated by four bacterial phyla: *Proteobacteria* (71.00%), *Firmicutes* (23.71%), *Spirochaetes* (3.76%) and *Actinobacteria* (1.13%). The distribution of these predominant bacterial phyla exposed major differences between taxonomic level (Fig. 1a). While free-living samples revealed an asymmetrical structure of bacterial communities, dominated by one or two bacterial taxa per sample, the bacterial communities within the mariculture samples were more symmetrically distributed. On the free-living side, samples free3 and free5 were dominated by *Vibrionales* (free3: 88.77%) and *Bacillales* (free5: 75.89%), whereas free1 had two bacteria dominating: *Rhizobiales* (29.25%) and one unclassified bacteria (50.24%) at the order level. This unclassified annotation belonged to one taxa, annotated to *Proteobacteria* at phylum level. A more detailed annotation could not be achieved. However, the mariculture samples showed a more balanced distribution of bacteria. Most of the bacterial taxa appeared in a similar proportions. For each mariculture sample *Actinomycetales* (mean: 2.18%, sd: 0.41%), *Burkholderiales* (mean: 10.23%, sd: 0.59%), *Pseudomonadales* (mean: 24.10%, sd: 1.56%) *Enterobacteriales* (mean: 15.52%, sd: 6.36%) and *Rhizobiales* (mean: 5.23%, sd: 0.43%) occurred in a similar distribution. Further analysis, calculating a log2 fold change to determine significant differences between the samples, revealed a significant difference for *Vibrionales*, *Rhizobiales* and *Bacillales* over-represented in free-living samples and a significant difference for *Enterobacteriales* in the mariculture samples (Supplementary Fig. S2). While each free-living sample is dominated by one of the three bacteria *Vibrionales*, *Rhizobiales* and *Bacillales*, they have very low abundance for *Enterobacteriales* (mean: 1.72%, sd: 1.45%). In contrast the mariculture samples show a high stable occurrence for *Enterobacteriales* (mean: 15.52%, sd: 6.36%). Thereby the log2 fold change analysis support the findings of a more balanced distribution of bacteria in the mariculture samples.

On analyzing rare phyla, each sample revealed occurrences for *Bacteriodetes* (mean free-living: 0.19% vs. mean mariculture: 0.58%) and *Tenericutes* (mean free-living: 0.05% vs. mean mariculture: 0.10%) (Fig. 1b) with no significant differences between the two environmental conditions. In addition, sample free1 also showed abundance for

Table 1
Top 10 enriched GO terms per environmental condition. The enriched GO term were ordered by p-value and for every environmental condition 10 most enriched were used for functional analysis. In addition the distance to root of the ontology and number of matched Pfam identifier are listed.

GO term	Description	# annotated Pfam identifier	z-Score	p-Value	Level
Free-living					
GO:0006298	Mismatch repair	39	8.36	3.4e−14	8
GO:0006281	DNA repair	39	5.63	7.5e−08	6
GO:0006974	Response to DNA damage stimulus	39	5.29	3.3e−07	5
GO:0006835	Dicarboxylic acid transport	16	5.02	1.7e−06	9
GO:0031532	Actin cytoskeleton reorganization	8	4.67	4.9e−06	6
GO:0046677	Response to antibiotic	27	4.45	1.2e−05	4
GO:0044237	Cellular metabolic process	403	3.95	0,00003	2
GO:0008152	Metabolic process	429	3.87	4.2e−05	1
GO:0009636	Response to toxic substance	27	4.07	4.9e−05	3
GO:0042493	Response to drug	7	3.89	0.00009	3
Mariculture					
GO:0006259	DNA metabolic process	3	1.56	0.35	5
GO:0090304	Nucleic acid metabolic process	3	1.09	0.76	4
GO:0050896	Response to stimulus	3	1.02	0.85	1
GO:0044237	Cellular metabolic process	8	884	0.12	2
GO:0046483	Heterocycle metabolic process	4	833	0.12	3
GO:0034641	Cellular nitrogen compound metabolic process	4	784	0.13	3
GO:0050789	Regulation of biological process	3	719	0.14	2
GO:0006725	Cellular aromatic compound metabolic process	4	735	0.14	3
GO:0006139	Nucleobase-containing compound metabolic process	3	643	0.15	3
GO:1901360	Organic cyclic compound metabolic process	4	655	0.16	3

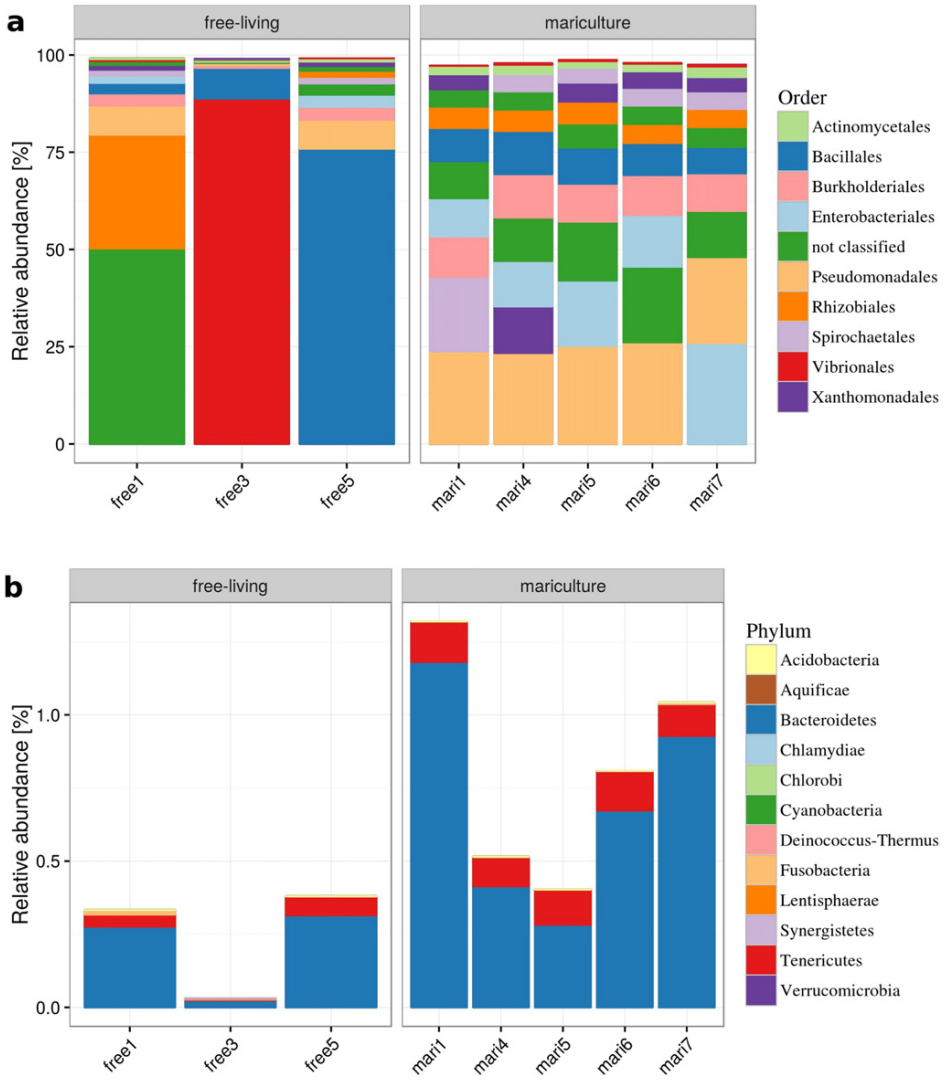


Fig. 1. Visualization of the bacterial communities divided per environmental condition. (a) Corresponds to nine bacterial taxa on order level appearing predominatly within host species *E. fuscoguttatus*, whereas (b) regards to 12 detected rare phyla of the bacterial community. Both graphs are divided between the two environmental conditions: free-living and mariculture.

Fusobacteria (free1: 0.01%). Other bacterial phyla, in particular *Acidobacteria*, *Aquificae*, *Chlamydiae*, *Chlorobi*, *Deinococcus-Thermus*, *Lentisphaerae*, *Synergistetes* and *Verrucomicrobia*, could be detected, but at very low abundance (<0.01%).

In order to obtain alpha diversity measurements, three statistical methods were applied. First, observed taxonomic richness (Fig. 2a), a simple counting of the number of appearing taxa, was performed, which indicated higher richness in free-living samples (median: 175 taxa) in contrast to mariculture samples (median: 105 taxa). Due to small sample size, free1 was considered as median representation, while free3 represents an outlier above the median value with high number of observed taxa (451 taxa) and free5 an outlier below the median (108 taxa). The mariculture samples displayed one outlier with 143 taxa (mari4), whereas the other samples were clustered at the median value. The Chao1 index (Chao, 1984), an nonparametric richness estimator, providing a statistical estimation of true species richness, including unobserved species within a community, revealed a high number of unobserved species (median: 240.34 taxa) for the free-living samples (Fig. 2b). On the other hand, mariculture samples revealed a minor increase in expected taxa (median: 125.75 taxa), in contrast to the observed species richness (median: 105 taxa). To measure ecological diversity, Shannon-Wiener diversity index (Shannon and Weaver) was applied, indicating a decreased diversity for mariculture samples (median: 0.94) in contrast to samples, derived under free-living conditions (1.52) (Fig. 2c). Results of the alpha diversity measurements were supported by rarefaction curves (Supplementary Fig. S3). An analysis of beta diversity (Fig. 2d), a measurement of the difference between the bacterial compositions under different environmental conditions, was done by using non-metrical dimensional scaling method in conjunction with Bray-Curtis distances (Bray and Curtis, 1957). The analysis revealed one cluster constructed entirely from the mariculture samples, whereby the free-living samples spread around this cluster. This result was supported by initial hierarchical clustering. Further analysis on the beta diversity, confirmed the results by using statistical test Adonis, resulting in a *p*-value of 0.011.

2.3. Accessing the eukaryotic content

Taking into account the eukaryotic taxa we could detect 18 eukaryotic phyla in decreasing order of abundance: *Chordata* (99.9479%), *Streptophyta* (0.0425%), *Mollusca* (0.0018%), *Platyhelminthes* (0.0015%), *Arthropoda* (0.0015%), *Ascomycota* (0.0011%), *Cnidaria* (0.0005%), *Acanthocephala* (0.0004%), *Nematoda* (0.0003%), *Annelida* (0.0002%), *Echinodermata* (0.0002%), *Basidiomycota* (0.0002%),

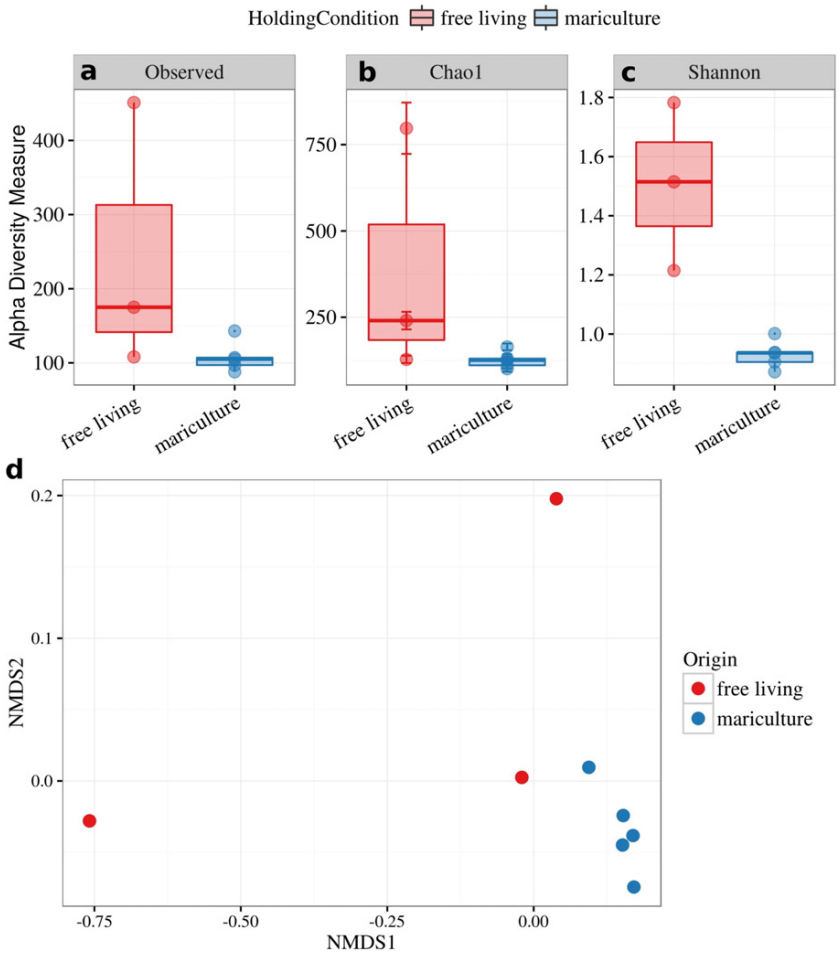


Fig. 2. Alpha diversity estimates across the two environmental conditions free-living and mariculture. (a) Observed refers to the observed taxonomical richness; (b) Chao1 represents the Chao1 index that estimates the true species of a sample; (c) Shannon represents the Shannon-Wiener diversity index accounting for species abundance and evenness of distribution. Dots represent estimates for individual samples. A solid line represents the median, boxes the quartiles and bars the interquartile range. (d) Beta diversity is estimated with nonmetric multidimensional scaling (NMDS) of bacterial communities colored by environmental conditions. Ordinations are based on between-sample dissimilarities calculated by Bray-Curtis distances.

Nemertea (0.0001%), *Bryozoa*, *Brachiopoda*, *Apicomplexa*, *Hemichordata*, *Chlorophyta*(<0.0001%). While the samples are highly dominated by the phylum *Chordata* (99.9479%) due to the fact that we had not filtered out host sequences before sequencing, a large proportion of this annotated reads belongs to *E. fuscoguttatus*. On the order level, ten taxa could be detected, however the most annotated sequences belonged to *Perciformes* (74.38%), an element of the linage of the host species *E. fuscoguttatus* (Fig. 3a). The additional nine taxa were distributed to other members of the class *Actinopteri* in decreasing order of abundance: *Cichliformes* (3.56%), *Cypriniformes* (2.08%), *Tetraodontiformes* (0.60%), *Cyprinodontiformes* (0.43%), *Beloniformes* (0.36%), *Spariformes* (0.30%), *Pleuronectiformes* (0.26%), *Salmoniformes* (0.17%) and *Centrarchiformes* (0.11%). The remaining annotated sequences (17.75%) revealed a taxonomical hit on class level (*Actinopteri*) but no further information on order level was available. In order to obtain a nuanced view on the phyla excluding *Chordata*, all annotated sequences belonging to this taxa were excluded for subsequent analysis of rare phyla present in the samples (Fig. 3b). The resulting composition of rare phyla could be divided into three distinct groups. One group was formed by parasitic taxa, with a majority of the samples displaying abundance for phyla belonging to known *endo*-parasites such as *Platyhelminthes* (0.0010%), *Acanthocephala* (0.0004%), *Nematoda* (0.0003%), *Annelida* (0.0002%), and *Nemertea* (0.0001%), as well as *Apicomplexa* (<0.0001%), a large phylum of parasitic protists. The largest detected group amongst these parasites, was *Platyhelminthes*, which was distributed over each sample. Another group of rare phyla was formed by taxa that were assigned as nutriment or residues from eaten nutriments. Phyla such as *Mollusca* (0.0018%), *Arthropoda* (0.0015%), *Cnidaria* (0.0005%), *Echinodermata* (0.0002%), *Hemichordata*, *Bryozoa* and *Brachiopoda* (<0.0001%) belonged to this group. The rest of the rare phyla were assigned to two fungal taxa, *Ascomycota* (0.0011%) and *Basidiomycota* (0.0002%), as well as *Streptophyta* (0.0425%) and *Chlorophyta* (<0.0001%), two plant taxon.

2.4. Functional annotation of the bacterial composition

To obtain a functional annotation, taxonomical annotated sequences belonging to predominantly appearing taxa were extracted from the dataset and pooled for the different environmental conditions. While, only sequences belonging to *Proteobacteria* on phylum level and deeper annotated sequences of this phylum could be successfully mapped to functional annotations. Firstly, the sequences were annotated with Pfam identifiers (Finn et al., 2014), providing information about protein coding parts within the sequences. Next, these Pfam identifiers were used to perform a Gene Ontology enrichment analysis, a computational method to identify classes of genes or proteins showing statistically significant differences between the two different environmental conditions. Therefore annotated Pfam identifiers were assigned to Gene

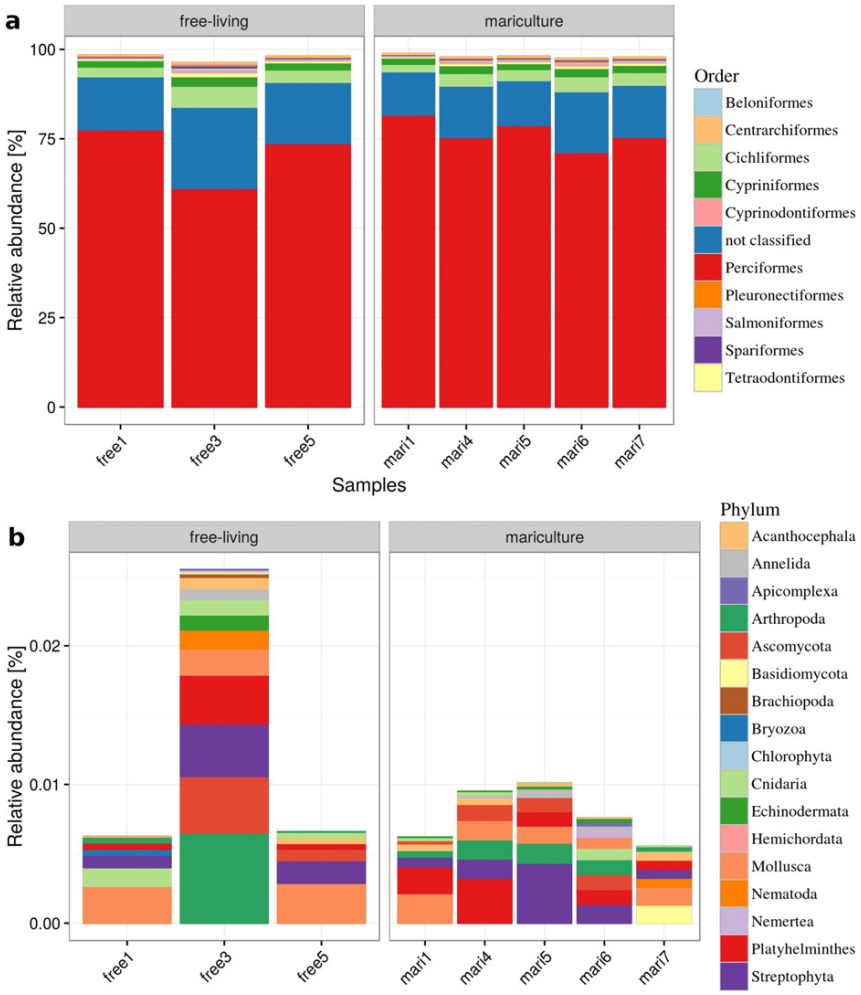


Fig. 3. Relative abundance of the metagenomic eukaryotic content detected inside the feces of the host species *E. fuscoguttatus*. (a) Corresponds to ten eukaryotic taxa at order level detected inside the predominantly appearing phylum *Chordata*; (b) shows the detected 17 rare phyla after removing the phylum *Chordata*. Both barplots are divided by the two environmental conditions: free-living and mariculture.

Ontology terms in combination with a fisher's exact test to identify significant sets of over-represented functions. Due to the scope of this analysis only the biological process ontology, describing series of molecular events or molecular functions (The Gene Ontology Consortium, 2000) were used. Instead of having only annotations of proteins and protein subdomains the assignment of Gene ontology terms provided a more targeted view on the functional content inside the samples.

Both groups of samples show high abundance as well as a significant appearance for the GO term cellular metabolic process (*GO:0044237*, *p*-value free-living: 0,00003, *p*-value mariculture: 0.012) a general term in the biological process ontology, directing to a number of GO terms representing more detailed functions (Fig. 4a). A closer look on the free-living samples revealed an increased abundance for DNA Repair (*GO:0006281*), that further point at DNA mismatch repair (*GO:0006298*) the most abundant term inside the free-living samples. Whereby both terms appeared significant enriched inside the free-living samples (*GO:0006281*, *p*-value: $7.5xe^{-8}$, *GO:0006298*, *p*-value: $3.4xe^{-14}$). Considering the taxonomical annotation revealed a linkage of these functions, described by these GO terms, with *Vibrionales*, *Photobacterium* and *Enterbacteriales*. In addition, the more general term DNA damage stimulus (*GO:0006974*), linked to DNA Repair (*GO:0006281*) displayed increased abundance as well as a significant enrichment (*p*-value: $3.3xe^{-7}$). An increased number of annotated Pfam identifiers were detected for the function response to chemical stimulus (*GO:0042221*, *p*-value: 0.00021) separating into three more detailed and significant enriched functions. Whereas response to drug (*GO:0042493*, *p*-value: 0.00009) showed a distinct subset of annotated Pfam identifiers, the functions response to toxic substance (*GO:0009636*, *p*-value: $4.9xe^{-5}$) as well as response to antibiotic (*GO:0046677*, *p*-value: $1.2xe^{-5}$) were linked together by a shared number of annotated Pfam identifier. Thereby response to antibiotic (*GO:0046677*) represented a subset of response to toxic substance (*GO:0009636*). Concerning these three functional annotations all annotated Pfam identifiers were exclusively linked to *Vibrionales* in the free-living samples and mostly annotated to membrane and transport related protein domains. Actin cytoskeleton reorganization (*GO:0031532*) represented another significant enriched GO term (*p*-value: $4.9xe^{-6}$) displaying high abundance of annotated Pfam identifier related to tRNA protein domains. Analysis of functional content of the mariculture samples (Fig. 4b) revealed four enriched functions (*GO:0034641*, *GO:0046483*, *GO:0006725*, *GO:1901360*) concerning metabolic processes each linked to the GO term nucleobase-containing compound metabolic process (*GO:0006139*, *p*-value: 0.15). Whereby this term pointed at nucleic acid metabolic process (*GO:0090304*) and DNA metabolic process (*GO:0006259*). Furthermore two abundant and enriched GO terms, describing a response to stimulus (*GO:0050896*, *p*-value: 0.085), as well as regulation of biological processes (*GO:0050789*, *p*-value: 0.14) could be detected. Functional annotations to the other two remaining

Fig. 4. Graph visualization of the biological process ontology used by the 10 most enriched GO terms of functional annotation. The 10 most enriched terms were ordered by *p*-value. The coloring of the graph nodes represents the abundance of Pfam identifiers mapped to a specific GO term, whereby all nodes with solid black line show an enrichment.

ontologies of the Gene Ontology were not considered due to low informative content.

3. Discussion

A number of studies concerning the gut microbiome of marine organisms such as fish (Xia et al., 2014; Xing et al., 2013) shrimp (Liu et al., 2011; Zhang et al., 2014) and oyster (King et al.) have indicated the influence of the microbiome on the metabolism (Xia et al., 2014) and immunity (Guarner and Malagelada, 2003; Balcazar et al., 2006) of the host species. In this study we applied a next generation sequencing approach, based on the Illumina MiSeq platform, to elucidate the bacterial composition of the microbiome, the corresponding functional annotation for predominant bacteria and the eukaryotic content of *E. fuscoguttatus* under free-living and mariculture conditions. In contrast to previous studies, where a combination of 16S rRNA analysis for the taxonomic analysis and a total DNA metagenome sequencing approach to access functional annotations was done, here we use a whole metagenomic shotgun sequencing approach, to obtain both the microbiome and the functional annotations and additionally the eukaryotic content (Muzquiz, 2006). The bacterial communities were dominated by four phyla that were similarly distributed over all samples. Particularly *Proteobacteria* (71.00%) and *Firmicutes* (23.71%), which were universally represented in the samples, have been reported to be a common part of the fish gut microbiome, as they have been described in similar distributions in a number of other studies (Xia et al., 2014; Sullam et al., 2012; Sevellec et al., 2014). In addition, *Spirochaetes* (3.76%) and *Actinobacteria* (1.13%) have also been reported as dominant components of the fish bacterial gut community (Xing et al., 2013). On a deeper level of taxonomy, the distributions of bacterial communities between the two environmental conditions were more dissimilar. While the mariculture samples showed a more stable distribution of microbial taxa over the samples, represented by comparatively even distributions of the microbiome, the free-living samples had a more diverse microbiome per sample. Each of the free-living samples revealed a unique bacterial community not similar to any of the other samples. Particularly free3, which was highly dominated by *Vibrionales* (88.77%) and free5, which was highly dominated by *Bacillales*

(75.89%) indicating the influence of environmental conditions on the bacterial communities of the host species. Xia et al.'s (Xia et al., 2014) investigation of the influence of starvation on the intestinal microbiome of fish, revealed shifts in the distribution of bacterial communities under different feeding conditions which supports our opinion that controlled feeding and environmental conditions appear to stabilize the microbiome of the host species. Our investigation of the eukaryotic reads exposed no significant differences in feed between the two environmental conditions. However, a connection appears to exist between the environmental conditions and the stability of the bacterial communities. The comparable distribution of bacteria within the mariculture samples indicated, that the more controlled environment of the mariculture leads to a more stable bacterial community, compared to free-living samples. In contrast, alpha diversity measurements pointed out that the number of observed and estimated taxa in the mariculture samples are lower in contrast to the free-living samples. The usage of a whole metagenomic sequencing approach enabled us to determine the eukaryotic content of the samples. It revealed the dominance of *Chordata* (99.95%), a taxa linked to the host, *E. fuscoguttatus*, and other eukaryotic orders of fish representing feed. Residues of host cells in the fecal samples constituted nearly 74.47% of the taxonomical annotated sequences, while some of these annotations may also belong to other fishes of the same order that were eaten by the host species. Further study of the other constituents of the taxa reveal nine additional order of fish taxa, living mainly in salt water or brackish water around the island of Java. Exceptions are *Cichliformes* and *Salmoniformes*. While all members of *Cichliformes* are exclusively fresh water fishes, *Salmoniformes* wander between fresh and salt water during their life cycle, but both fish species are not endemic to Indonesian waters. These two taxa are the results of an erroneous annotation due to high sequence similarity with a wrong hit sequence in the database. Following the filtering out of the annotated *Chordata* taxa, 13 rare phyla could be detected. Each phylum fits into one of three groups: parasites, fungi and food residues. Over all sample footprints of annotated parasitic sequences were found to be distributed over seven parasitic taxa. While, *Platyhelminthes*, *Arthropoda* and *Acanthocephala* appear in nearly all samples, the other four parasite phyla, namely *Nemertoda*, *Nemertea*, *Annelida* and *Apicomplexa*, are only detected in single samples. Previous parasitological studies of aquaculture and free-living fish in Indonesian waters had revealed the presence of ecto- and endoparasites as endemic (Sharpton, 2014; Rueckert et al., 2010). Therefore the discovery of parasitic sequences in the metagenome of the host species is not surprising. Additionally, a variety of eukaryotic marine phyla were found in low quantities. These were *Brachiopoda*, *Cnidaria*, *Echinodermata*, *Mollusca*, and *Hemichordata*. All of which could be categorized as food directly eaten by the host species or as undigested feces residues from an eaten fish (Sullam et al., 2012). Moreover, two fungal phyla were detected as rare phyla: *Ascomycota* and *Basidiomycota*, both pooled in the subkingdom *Dikarya* (Hibbet and et al., 2007). As these fungi are known to be terrestrial in combination with the low abundance values of these annotations, it is more likely false classification from the database. Therefore, these two phyla have to be excluded from the eukaryotic content. Finally, two plant taxa, *Chlorophyta* and *Streptophyta*, were also detected as rare phyla of eukaryotic taxa, whereby both phyla include algae these taxa could also be explained by food residues.

Investigation of functional components of the predominantly appearing bacterial taxa was focused on the biological process ontology of the Gene Ontology (The Gene Ontology Consortium, 2000), which describes a series of events accomplished by a number of organized assemblies of molecular functions. Reviewing the other two ontologies (The Gene Ontology Consortium, 2000) resulted in a low informational content for the cellular component and molecular function ontology. The results are completely linked to *Proteobacteria*, annotated from phylum level down to genus level of taxonomy. They reveal enrichment for DNA repair functions. Especially functions related to the reparation of mismatches, in combination with the discovered enrichment for a response to DNA damage can be annotated inside the free-living samples. Whereby samples from mariculture also point out enriched terms directing to the function of DNA metabolic processing, but show no enrichment for the DNA repair function. This could be a result of the annotation process used in this study. Whereas this function is assigned to *Vibrionales*, *Pseudomondales* and *Enterobacteriales* in the free-living samples, the same function can be annotated to *Proteobacteria* only on the mariculture side. Nevertheless the appearance and increased abundance of this specific function could point to a higher contamination level for the free-living samples in contrast to the samples of the mariculture, living in a controlled environment. In addition, another interesting result of the functional annotation is the linkage between *Vibrionales* and a response to antibiotics in the free-living samples. As described in the results section the corresponding Pfam identifier (Finn et al., 2014) belonging to this enriched function are assigned to membrane and transport related protein domains. The exclusive appearance of *Vibrionales*, as a response to antibiotics is expected in samples from mariculture and not in the free-living ones. The appearance of this functional classification could be a result of higher contamination for free-living *E. fuscoguttatus*.

4. Conclusion

The intestinal and gut microbiomes of fish have been investigated by various studies, reporting characteristic predominant taxa and also the influence of starvation on the bacterial communities. In this study we demonstrate that the environmental conditions also have an influence on the distribution of fish bacterial communities. We showed that bacterial communities of mariculture samples have overall a more similar distribution than bacterial communities of free-living samples, which are dominated by overall different bacteria. As a result of using a whole metagenomic sequencing approach, we were able to exclude influences of feed. By performing a taxonomical annotation of the eukaryotic components of the dataset, we disclosed the dominance of the phylum *Chordata*, comprising the taxa of the host, *Epinephelus fuscoguttatus*, and other fish related taxonomic orders. In addition, we detected a variety of parasitic footprints present in each of the samples, belonging to the phyla *Platyhelminthes*, *Arthropoda* and *Acanthocephala*. Furthermore, investigations on the functional annotation of the predominant taxa revealed functions mainly related to DNA repair in free-living samples, as well as more general to DNA metabolic processes in the mariculture samples. Moreover, we were able to identify functions involved in antibiotic response in the free-living samples.

5. Material and methods

5.1. Sample collection

Brown-marbled groupers *E. fuscoguttatus* (Forsskål, 1775) used in this study were collected from the Thousand Islands (Pulau Seribu) Marine National Park, 45 km North of Jakarta Bay during rainy season 2012 (November and March). Cultured samples ($n = 7$) originated from an open water mariculture facility (Nusa Karamba Aquaculture), while free-living samples ($n = 6$) were obtained nearby the mariculture facility, using net cages, fish traps and fishing rods of local fishermen. Immediately following sampling, the samples were dissected in a local laboratory (Nusa Karamba Aquaculture), where feces was extracted and stored in 99.9% EtOH for subsequent DNA extraction.

5.2. Isolation of genomic DNA and sequencing

Tissue obtained from the fish samples (5 to 10 mg per sample) were homogenized in lysis buffer (Bio & Sell, Germany) using Precellys tissue homogenizer tubes with 1.4 mm ceramic beads at 3×30 s homogenization time. Samples were incubated for one hour at 37 °C in lysis buffer.

Following which, samples were incubated with RNase A for one hour at 37° C and subjected to an overnight proteinase K digest. Whole genomic DNA was extracted using phenol-chloroform-isoamylalcohol extraction followed by ethanol precipitation (Sambrook et al., 1989). Following extraction, IMPLEN NanoPhotometer TM was used to quantify DNA concentration and purity. DNA samples were then subjected to 250-bp paired-end sequencing on an Illumina MiSeq platform at StarSEQ GmbH (Mainz, Germany). Raw sequence data has been deposited in NCBI Short Read Archive under accession number SRP059771. Sample metadata is provided as Supplementary Table 1.

5.3. DNA analysis

The resulting 40,823,868 (10 Gbp) paired-end raw sequences were obtained in FASTQ format. The sequences were trimmed and filtered based on their quality (minimum length of 200 bp and a quality threshold of 18) using Trimmomatic (Bolger et al., 2014). To improve the accuracy of subsequent annotation tasks the paired-end sequences were merged to one sequences, based on overlapping regions, using FLASh (Magoc and Salzberg, 2011) with a minimum overlap threshold of 10 bp. Subsequently, merged sequences were pooled with unmergable sequences and annotated with blastn (Altschul et al., 1990) against the nt database (Sayers et al., 2009) to annotate sequences originating from eukaryotic organisms, as well as bacterial organisms (Supplementary Fig. 4). To assign a taxonomical annotation to the hits for every query, sequences were filtered to a number of blast hits with a query coverage above 50% and a bit score above a cutoff value of 85% of the bit score achieved by the best hit. These sequences were used to perform an additional database assignment against the NCBI taxonomy database (Altschul et al., 1990; National Center for Biotechnology Information, 2014). Due to the possibility of multiple different assignments for one query sequence a lowest common ancestor algorithm was used to determine a taxonomical annotation for the query sequence. In total 11,757,803 (28.80%) sequences obtained a taxonomical annotation, spread over 3,217,992 annotated sequences of eukaryotic origin and 8,539,811 bacterial sequences. A large proportion of the bacterial reads (5,762,047 sequences) could only be assigned to the taxa bacteria on superkingdom level and was excluded from analysis, due to the low informative content. In addition, one sample was excluded from analysis, due to low number of reads (free4). Finally, after filtering, the 3,217,992 eukaryotic and 2,774,233 bacterial annotated reads (14.68%) were converted into a Biological Observation matrix file (McDonald et al., 2012) for subsequent analysis, which was performed in R (R Core Team, 2014) using the packages phyloseq (McMurdie and Holmes, 2013), DESeq2 (Anders and Huber, 2010) and vegan (Oksanen et al., 2016).

In order to obtain a functional annotation of the predominant bacteria, the corresponding sequences were extracted from the quality filtered and merged sequence datasets and subsequently translated into open reading frames, by using biopython (Cock et al., 2009). Subsequently, the resulted open reading frames were clustered by using cd-hit (Weizhong, 2009; Li and Godzik, 2006) with a sequence identity of 90%. The representative sequences for the different clusters were used for an annotation against the Pfam protein family database (Finn et al., 2014), using the software package hmmsearch (Finn et al., 2011), based on a hidden markov motive search algorithm. A subsequent functional analysis of the functional annotated sequences was performed in R (R Core Team, 2014), using the additional library packages phyloseq (McMurdie and Holmes, 2013) and dcGOR (Fang, 2014). Thereby Pfam annotated sequences belonging to Bacteriophage or viral origins were filtered out. The remaining sequences were further annotated by Gene Ontology terms (The Gene Ontology Consortium, 2000) referring to the biological process ontology. Subsequently a GO enrichment analysis on the annotated results was performed and further analyzed.

Acknowledgements

We wish to thank Harry W. Palm and his group at the University of Rostock for providing us with the samples used in this study. We are grateful to SPICE III for financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.marpolbul.2016.05.009.

References

Altschul, S.F., Gish, W., M. W, Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.
Balcazar, J.L., Blas, I.D., Ruiz-Zarzuela, I., Cunningham, D., Vendrell, D., Muzquiz, J.L., 2006. The role of probiotics in aquaculture. Vet. Microbiol. 114, 173–186.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics (btu170).
Bray, J.R., Curtis, J.T., 1957. An ordination of upland forest communities of southern Wisconsin. Ecol. Monogr. 27, 325–349.
Carpenter, K.E., Springer, G., 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ. Biol. Fish 72, 467–480.
Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Stat. 11 (4), 265–270 (URL http://www.jstor.org/stable/4615964).
Cock, et al., 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25 (11), 1422–1423.
Cooper, T.F., Gilmour, J.P., Fabricius, K.E., 2009. Bioindicators of changes in water quality on coral reefs: review and recommendations for monitoring programmes. Coral Reefs 28, 589–606.
Cornish, A., 2004. *Epinephelus fuscoguttatus*, The IUCN Red List of Threatened Species e.T44673A10934320 http://dx.doi.org/10.2305/IUCN.UK.2004.RLTS.T44673A10934320.en .
Dsikowitzky, L., et al., 2011. Anthropogenic organic contaminants in water, sediments and benthic organisms of the mangrove-fringed Segara Anakan Lagoon, Java, Indonesia. Mar. Pollut. Bull. 62, 851–862.
Eng, C.T., Paw, J.N., G. F. Y, 1989. The environmental impact of aquaculture and the effects of pollution on coastal aquaculture development in Southeast Asia. Mar. Pollut. Bull. 20, 335–343.
Fang, H., 2014. dcGOR: an R package for analyzing ontologies and protein domain annotations. PLoS Comput. Biol. 10 (10), e1003929. http://dx.doi.org/10.1371/journal.pcbi.1003929 (URL http://dx.doi.org/10.1371/journal.pcbi.1003929).
Finn, R.D., Clements, J., Eddy, S.R., 2011. Hmmer web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.
Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. The Pfam protein families database. Nucleic Acids Res. 42, D222–D230.
Givens, C.E., Ransom, B., Bano, N., Hollibaugh, J.T., 2015. Comparison of the gut microbiome of 12 bony fish and 3 shark species. Mar. Ecol. Prog. Ser. 518, 209–223.
Guarner, F., Malagelada, J.-R., 2003. Gut flora in health and disease. Lancet 360, 512–519.
Hibbet, D.S., et al., 2007. A higher-level phylogenetic classification of the fungi. Mycol. Res. 111, 509–547.
Hutomo, M., Moosa, M., 2005. Indonesian marine and coastal biodiversity: present status. Indian J. Marine Sci. 34, 88–97.
King, G.M., Judd, C., Kuske, C.R., Smith, C., 2012. Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. PLoS One 7.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
Liu, H., Wang, L., Liu, M., Wang, B., Jiang, K., Ma, S., et al., 2011. The intestinal microbial diversity in Chinese shrimp (*Fenneropenaeus chinensis*) as determined by PCR-DGGE and clone library analyses. Aquaculture 317. http://dx.doi.org/10.1016/j.aquaculture.2011.04.008 (32–6).
Magoc, T., Salzberg, S., 2011. Flash: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963.
McDonald, D., Clemente, J.C., Kuczynski, J., Rideout, J.R., Stombaugh, J., Wendel, D., Wilke, A., Huse, S., Hufnagle, J., Meyer, F., Knight, R., Caporaso, J.G., 2012. The Biological Observation Matrix (biom) format or: how I learned to stop worrying and love the ome-ome. Giga Science 1, 7. http://dx.doi.org/10.1186/2047-217×-1-7.
McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.14.
Muzquiz, J., 2006. The role of probiotics in aquaculture. Vet. Microbiol. 114, 173–186.
National Center for Biotechnology Information, 2014. NCBI genomes ftp site (April URL ftp://ftp.ncbi.nlm.nih.gov/genomes/).
Nordhaus, I., Hadipudjana, F.A., Janssen, R., Pamungkas, J., 2009. Spatio-temporal variation of macrobenthic communities in the mangrove-fringed Seagara Anakan Lagoon, Indonesia, affected by anthropogenic activities. Reg. Environ. Chang. 9, 291–313.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2016. Vegan: Community Ecology Package, R Package Version 2.3-4 (URL https://CRAN.R-project.org/package=vegan).

Pet-Soede, C., MacHiels, M., Stam, M., Van Densen, W., 1999. Trends in an Indonesian coastal fishery based on catch and effort statistics and implications for the perception of the state of the stocks by fisheries officials. Fish. Res. 42, 41–56. http://dx.doi.org/10.1016/S0165–7836(99)00034-X.

R Core Team, 2014. R: A Language and Environment for Statistical Computing (Jan. URL http://www.r-project.org/).

Rinawati, et al., 2012. Distribution, source identification, and historical trends of organic micropollutants in coastal sediment in Jakarta Bay, Indonesia. J. Hazard. Mater. 217–218, 208–216.

Roeselers, G., et al., 2011. Evidence for a core gut microbiota in the zebrafish. ISME J. 5, 1595–1608.

Rueckert, S., Klimpel, S., Palm, H.W., 2010. Parasites of cultured and wild brown-marbled grouper *Epinephelus fuscoguttatus* (Forsskål, 1775) in Lampung Bay, Indonesia. Aquac. Res. 41, 1158–1169.

Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor laboratory press, New York.

Sayers, et al., 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 37, D5–D15.

Sevellec, M., Pavey, S.A., Boutin, S., Filteau, M., Derome, N., Bernatchez, L., 2014. Microbiome investigation in the ecological speciation context of lake whitefish (*Coregonus clupeaformis*) using next-generation sequencing. J. Evol. Biol. http://dx.doi.org/10.1111/jeb.12374 (1029–46).

Shannon, C.E., Weaver, W., 1841. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Sharpton, T., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5, 209. http://dx.doi.org/10.3389/fpls.2014.00209.

Siregar, V., Koropitan, A., 2013. Primary productivity of jakarta bay in a changing environment: anthropogenic and climate change impacts. Biotropia 20, 89–103.

Sugama, K., Rimmer, M.A., Isma, S., Koesharyani, I., Suwirya, K., Giri, N.A., Alava, V.R., 2012. Hatchery Management of Tiger Grouper (*Epinephelus fuscoguttatus*): A Best-Practice Manual. Australian Centre for International Agricultural Research.

Sullam, K.E., et al., 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Mol. Ecol. 21, 3363–3378.

The Gene Ontology Consortium, 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

van der Meij, S.E.T., Moolenbeek, R.G., Hoeksema, B.W., 2009. Decline of the Jakarta Bay molluscan fauna linked to human impact. Mar. Pollut. Bull. 59 (4–7), 101–107. http://dx.doi.org/10.1016/j.marpolbul.2009.02.021.

van der Meij, S.E.T., Suharsono, Hoeksema, B.W., 2010. Long-term changes in coral assemblages under natural and anthropogenic stress in Jakarta Bay (1920–2005). Mar. Pollut. Bull. 60 (9), 1442–1454. http://dx.doi.org/10.1016/j.marpolbul.2010.05.011.

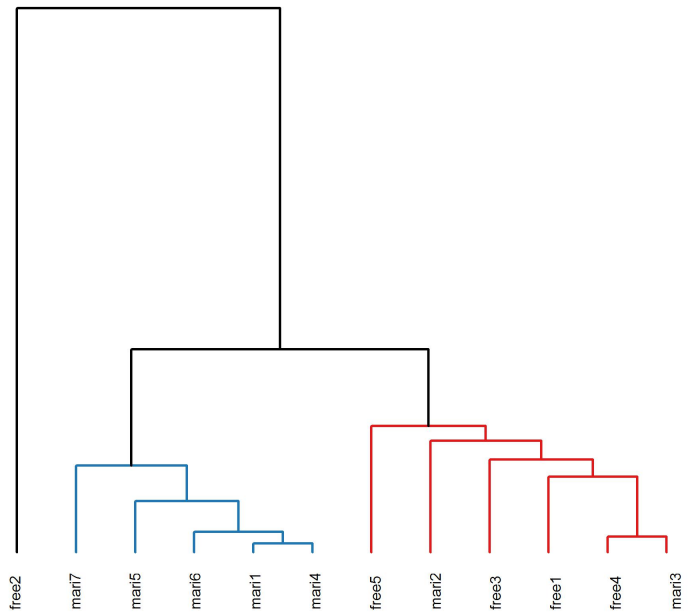
Weizhong, L., 2009. Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinf. 10, 359. http://dx.doi.org/10.1186/1471–2105–10-359.

Xia, J., et al., 2014. The intestinal microbiome of fish under starvation. BMC Genomics 15 (1), 266. http://dx.doi.org/10.1186/1471–2164–15-266.

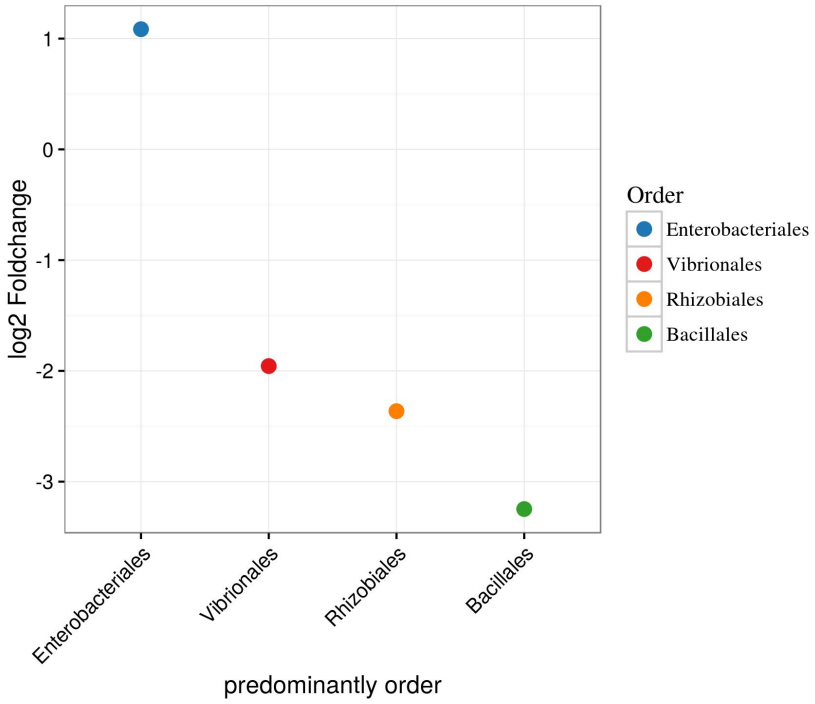
Xing, M., Hou, Z., Yuan, J., Liu, Y., Qu, Y., Liu, B., 2013. Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*). FEMS Microbiol. Ecol. 86, 432–443.

Zhang, M., Sun, Y., Chen, K., Yu, N., Zhou, Z., C. L, et al., 2014. Characterization of the intestinal microbiota in Pacific white shrimp, *Litopenaeus vannamei*, fed diets with different lipid sources. Aquaculture 434, 449–455.

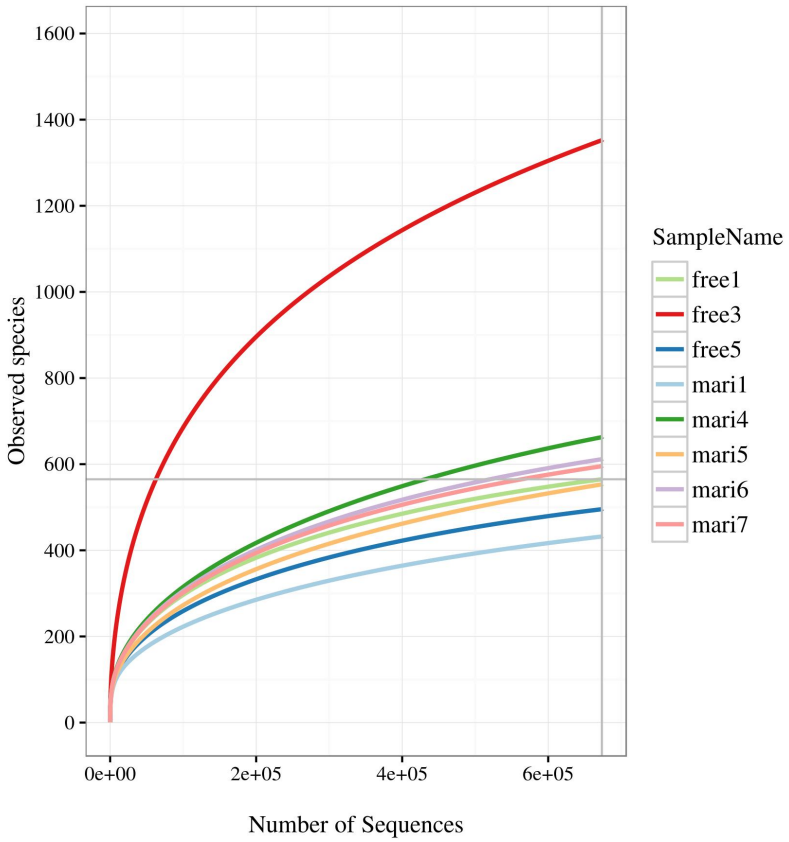
Supplement



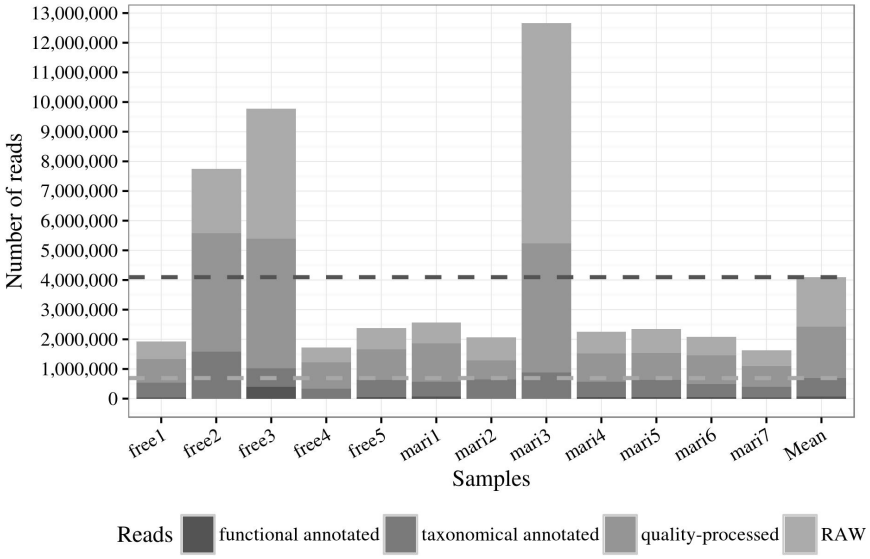
Supplementary Figure 1: **Initial hierarchical clustering of the samples, to determine samples not related to the corresponding environmental conditions.** The clustering was performed by using an euclidean distance measurement in combination with an average agglomeration method.



Supplementary Figure 2: **Log2 fold change to determine significant differences between the two environmental conditions.** All taxa showing a fold change with alpha value below 0.5 were assumed as significant different.



Supplementary Figure 3: **Rarefaction curves were constructed by counting the number of species and increasing the number of sequence reads.** A flatter curve to the right indicate a reasonable number of individual samples was taken. More intensive sequencing will only achieve few new species.



Supplementary Figure 4: **Number of raw sequence reads, post-QA/QC sequence reads, number of taxonomically annotated reads and number of functional annotated reads.** The light grey and dark grey dashed lines represent the average number of raw sequence reads and taxonomically annotated sequence reads across all samples.

Sample	Host species	Origin	longitude, latitude	Environment	season	habitat	Date	total Length (in cm)	total Weight (in g)	sex / gender	Number of parasites
free1	Epinephelus fuscoguttatus	Pulau Seribu	5.738372 S 106.608917 E	free-living	rainy season	free-living	March 2012	26.5	391.2	juvenile	23
free2	Epinephelus fuscoguttatus	Pulau Seribu	5.738372 S 106.608917 E	free-living	rainy season	free-living	March 2012	22.5	187.5	juvenile	7
free3	Epinephelus fuscoguttatus	Pulau Seribu	5.738372 S 106.608917 E	free-living	rainy season	free-living	March 2012	18.5	115.7	juvenile	9
free4	Epinephelus fuscoguttatus	Pulau Seribu	5.738372 S 106.608917 E	free-living	rainy season	free-living	March 2012	34.1	727.8	juvenile	48
free5	Epinephelus fuscoguttatus	Pulau Seribu	5.738372 S 106.608917 E	free-living	rainy season	free-living	March 2012	36.4	955.0	juvenile	474
mari1	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	25.8	305.4	male	20
mari2	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	24.2	271.6	juvenile	3
mari3	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	26.2	354.2	male	4
mari4	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	23.1	249.5	male	44
mari5	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	25.3	316.1	juvenile	14
mari6	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	25.7	299.0	male	5
mari7	Epinephelus fuscoguttatus	Pulau Seribu	5.736825 S 106.609228 E	mariculture	rainy season	net cage	March 2012	24.9	291.5	male	25

Supplementary Table 1: **Metadata**

SampleName	RAW	merged	quality controlled	taxonomical annotated	functional annotated
free1	1.401.439	1870846	1289136	531757	45488
free2	6.155.862	8514557	5580070	1589109	/
free3	8.743.868	5550285	4990929	1030024	403117
free4	1.388.403	1768584	1222168	333854	/
free5	1.736.051	2415265	1612967	637906	55133
mari1	1.991.046	2738325	1780660	570845	53924
mari2	1.422.876	2005226	1294682	65168	/
mari3	11.769.965	5.813.667	5240131	885528	/
mari4	1.682.298	2.268.302	1476652	565899	162204
mari5	1.713.223	2.388.324	1488598	633121	48962
mari6	1.585.176	2.157.281	1401802	491562	47896
mari7	1.233.661	1.609.940	1068888	401950	51876
total	81.647.736	39.100.602	28446683	8323193	37484

Supplementary Table 2: **Number of Reads.**

Microbiome analysis and detection of ogenic bacteria of *Penaeus monodon* from Jakarta Bay and Bali

Oetama VSP, **Hennersdorf P**, Abdul-Aziz MA, Mrotzek G, Haryanti H, Saluz HP **(2016)**

Marine Pollution Bulletin 110 (2), 718–725
doi: 10.1016/j.marpolbul.2016.03.043.

Marine Pollution Bulletin 110 (2016) 718–725

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Microbiome analysis and detection of pathogenic bacteria of *Penaeus monodon* from Jakarta Bay and Bali

Vincensius S.P. Oetama [a,b], Philipp Hennersdorf [a,b], Muslihudeen A. Abdul-Aziz [b,c], Grit Mrotzek [a], Haryanti Haryanti [d], Hans Peter Saluz [a,b,*]

[a] Leibniz-Institute for Natural Product Research and Infection Biology, Beutenbergstr. 11a, D-07745 Jena, Germany
[b] Friedrich Schiller University of Jena, Fürstengraben 1, D-07743 Jena, Germany
[c] Australian Centre for Ancient DNA, University of Adelaide, Adelaide, South Australia, Australia
[d] Gondol Research Institute for Mariculture GRIM, ds Penyabangan, Br. Gondol, PO. Box 140, Singaraja 81101, Bali, Indonesia

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 9 March 2016
Accepted 20 March 2016
Available online 15 April 2016

Keywords:
Penaeus monodon
Fecal microbiome
Jakarta
Bali
Free-living
Aquaculture

ABSTRACT

Penaeus monodon, the Asian black tiger shrimp is one of the most widely consumed marine crustaceans worldwide. In this study, we examine and compare the fecal microbiota of *P. monodon* from highly polluted waters around Jakarta Bay, with those of less polluted waters of Bali. Using next generation sequencing techniques, we identified potential bacterial pathogens and common viral diseases of shrimp. *Proteobacteria* (96.08%) was found to be the most predominant phylum, followed by *Bacteriodetes* (2.32%), *Fusobacteria* (0.96%), and *Firmicutes* (0.53%). On the order level, *Vibrionales* (66.20%) and *Pseudoaltermonadales* (24.81%) were detected as predominant taxa. qPCR profiling was used as a confirmatory step and further revealed *Vibrio alginolyticus* and *Photobacterium damselae* as two potential pathogenic species present in most of the samples. In addition, viral diseases for shrimp were discovered among the samples, WSSV in Jakarta free-living samples, YHV in Bali free-living samples and IHHNV in both.

1. Introduction

The black tiger shrimp, *Penaeus monodon* is one of the most widely traded marine crustaceans over the last few decades (Liu et al., 2011; FAO: Food Agric Organ United Nations, 2014). Due to its high economic value, the number of *P. monodon* aquaculture farms has rapidly increased. But several problems have emerged which have resulted in decreasing shrimp production (Acre and Moss, 2003; Venkateswara-Rao, 1998; Lightner, 2003). Low quality control of feed and an ineffective disease management for shrimp represents a major problem. Previous studies have reported the importance of bacterial communities in intestinal tract to maintain the metabolism and immunity of their host, as well as to evade viral and bacterial diseases (Zhang et al., 2014; Chaiyapechara et al., 2012). Due to its aquatic environment, *P. monodon* is highly exposed to potential pathogens (Rungrassamee et al.). In order to support the innate immunity of the host, fecal microbiota has evolved an established mechanism for host defense (Xia et al., 2014). Previous studies have shown, that the composition of bacterial communities is influenced by host phylogeny and diet (Zhang et al., 2014; Xia et al., 2014). Furthermore, ecological and environmental interaction also has an influence on alteration of bacterial diversity (Sullam et al., 2012; Asplund, 2013). A comprehensive understanding of bacterial community composition and alteration factors is essential for the enhancement of aquaculture quality. Currently, studies of fecal microbiota in shrimp have been done (Liu et al., 2011; Zhang et al., 2014; Luis-Villasenor et al., 2012), particularly in *P. monodon* (Chaiyapechara et al., 2012; Rungrassamee et al.). Comparative analyses between wild and domesticated *P. monodon* will provide a solid base for the development of probiotics. As previously reported (Partida-Arangure et al., 2013; Balcazar et al., 2006; Luis-Villasenor et al., 2012), probiotics have the ability to reshape bacterial communities and act as a host defense mechanism or in the improvement of surrounding conditions. Rungrassamee et al. showed that *Proteobacteria*, *Bacteriodetes*, *Firmicutes*, *Fusobacteria*, and *Actinobacteria* represent the predominant phyla in free-living *P. monodon*, as well as in aquaculture ones (Rungrassamee et al.). The bacterial community composition differs in each phylum under different conditions. Moreover, predominant genus that have been discovered are potential pathogens for shrimp and human, for instance, *Vibrio* sp. and *Photobacterium* sp. (Rungrassamee et al.; Vaseeharan et al., 2007). Furthermore, several common viral diseases in shrimp also play an important role for aquaculture improvement. They could cause a great loss, either by reducing the production or collapsing the whole aquaculture (Xia et al., 2014; Wegner et al., 2013; Guarner and Malagelada, 2003; O.L. et al., 2007). In this study, we use a 16S rRNA approach on an Illumina MiSeq

* Corresponding author at: Leibniz-Institute for Natural Product Research and Infection Biology, Beutenbergstr. 11a, D-07745 Jena, Germany.

http://dx.doi.org/10.1016/j.marpolbul.2016.03.043
0025-326X/

sequencing platform to elucidate the fecal microbiome of 24 *P. monodon* samples from three different locations, representing three different environments, polluted water around Jakarta Bay, non-polluted water of Bali Bay and cultured *P. monodon* from a traditional aquaculture farm in Bali. We focus on this three way comparison in order to determine differences in the bacterial community composition between these three environments. In addition, we also reveal the presence of several potential bacterial and viral pathogens that could have an impact on shrimp and human health using qPCR (quantitative Polymerase Chain Reaction).

2. Results

2.1. Bacterial community analysis

The Illumina MiSeq sequencing platform yielded in total 6,705,305 reads over all 22 samples. We had to exclude two free-living samples from Jakarta (Jk1: 2096 reads, Jk7: 159 reads) due to low number of reads. On average 304,800 mapped reads were obtained, ranging from 166,400 reads up to 581,400 reads (Supplementary Fig. 1), which were assigned to 935 Operational Taxonomic Units (OTUs) in order to assess the bacterial communities of the fecal samples. Selecting only phyla with an abundance level at above 0.1%, the bacterial communities revealed four bacterial phyla appearing predominantly: *Proteobacteria* (96.08%), *Bacteriodetes* (2.32%), *Fusobacteria* (0.96%), and *Firmicutes* (0.53%). These four phyla spread over twelve bacteria on the order level, which also appeared predominantly in all samples (Fig. 1a). On average 66.20% of *Vibrionales* accounted for the bacterial community composition, followed by 24.81% of *Alteromonadales*. These two dominant orders were relatively similar distributed in each sample, except in Jk3, Jk4, and Jk6 which showed higher numbers of *Vibrionales* and absence of *Alteromonadales*. Ba7 pointed out a different pattern, showing *Alteromonadales* (68.42%) as predominant order, followed by *Vibrionales* (28.83%). Aquaculture samples exhibited a significant number of bacteria at the order level beyond *Vibrionales* and *Alteromonadales*. For instance, Aq4 showed one unclassified bacteria, which we were unable to annotate on order level, but belonging to *Alphaproteobacteria* (21.64%), whereas *Sphingobacteriales* (15.01%) appeared in high numbers in Aq9. Bali free-living samples, particularly Ba4, showed a large portion of *Fusobacteriales* (20.21%). Rare phyla, showing abundance below 0.1% were also assessed. *Actinobacteria* (0.0623%), followed by *Cyanobacteria* (0.0070%), *Verrucomicrobia* (0.0039%), *Deinococcus-Thermus* (0.0013%), *Planctomycetes* (0.0006%), *Spirochaetes* (0.0005%), *Chloroflexi* (0.0004%), *Tenericutes* (0.0002%), *Acidobacteria* (0.0001%) and *Ignavibacteria* (<0.0001%) were found among all samples (Fig. 1b). Samples derived from Bali aquaculture showed *Actinobacteria* as the most prevalent rare phyla, specifically in Aq4 (0.30%) and Aq7 (0.07%). A similar trend was observed in Jakarta free-living samples, particularly in Jk2 (0.25%) whereas *Cyanobacteria* tended to be the most prevalent rare phylum in Bali free-living samples, especially in Ba6 (0.21%).

2.2. Diversity of the microbiome

Three different approaches were employed to investigate bacterial diversity between the samples. First, observed OTU richness, number of unique OTUs, revealed a high median number of observed OTUs for

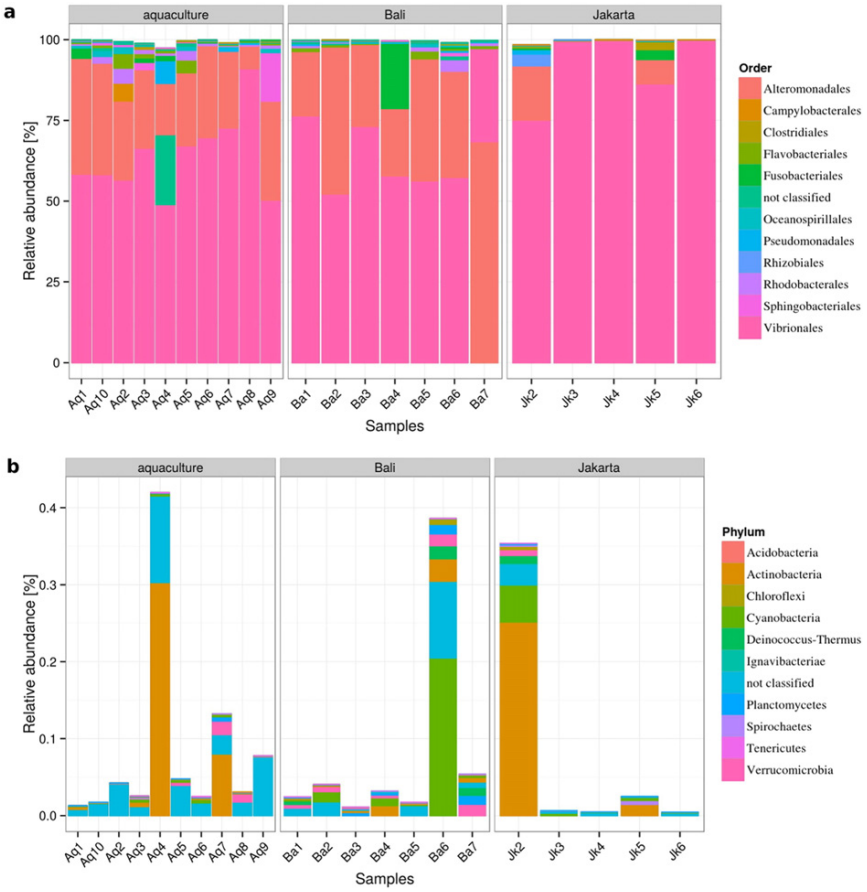


Fig. 1. Bacterial community composition of shrimp samples under Bali traditional aquaculture, Bali free-living, and Jakarta free-living condition. (a) corresponds to the predominantly appearing bacterial taxa on order level of taxonomy. Only bacterial taxa with an abundance over 0.1% were selected. (b) shows the rare phyla appearing in each sample under the different conditions.

Bali free-living samples (295 OTUs) and aquaculture (269 OTUs). However, most of the samples were similar in the number of observed OTUs, ranging from 200 OTUs up to 293 OTUs in Bali aquaculture samples and 224 to 316 OTUs for the Bali free-living samples (Fig. 2a). Two samples displayed a larger number of OTUs (Aq7: 357, Ba6: 484). The median for samples derived from Jakarta (122 OTUs) were more than two times lower with a range from 80 OTUs up to 169 OTUs, displaying one sample with an increased number of OTUs (Jk2: 322 OTUs). The Chao1 richness estimator (Chao, 1984), a statistical model for calculating the richness of a community including the undetected species, resulted in slightly higher median values as the observed richness (Fig. 2b). The samples from aquaculture (median: 303.23 OTUs) and free-living derived from Bali (median: 329.29 OTUs) showed a higher increase of expected OTUs, than the samples from Jakarta (median: 140.47 OTUs). In addition, the Shannon–Wiener diversity index (Shannon and Weaver, 1949) revealed the highest bacterial diversity for the Bali aquaculture samples (median: 2.58) and Bali free-living samples (median: 2.57) (Fig. 2c). In contrast, samples, derived under free-living conditions in Jakarta, showed low bacterial diversity (median: 0.93). The distribution of individual samples from Jakarta showed a higher variation in diversity than the individual samples of the other conditions, with Shannon–Wiener diversity from 0.54 up to 2.28. In order to measure differences between the different sampling locations and environmental conditions, a nonmetric multidimensional scaling method, using Bray–Curtis distances (Bray and Curtis, 1957) was employed to obtain an ordination matrix, based on between-sample dissimilarities. The ordination plot displayed two distinct clusters, whereas one cluster was constructed by three samples from Jakarta free-living (Jk3, Jk4, Jk6) and the other cluster consisted of samples from Bali free-living and Bali aquaculture. Nevertheless, four samples did not belong to any of the two clusters (Ba7, Aq8, Jk5 and Jk2). To confirm these findings, an Adonis analysis, provided by the statistical package vegan (Oksanen et al., 2016), was used. It confirmed the found clusters with a p-value below 0.001. The result of this beta diversity tests imply that the sample location has an huge influence on the bacterial community composition of the samples, due to the two distinct clusters depended mostly on the sampling location and not on the different environmental conditions. (See Fig. 3.)

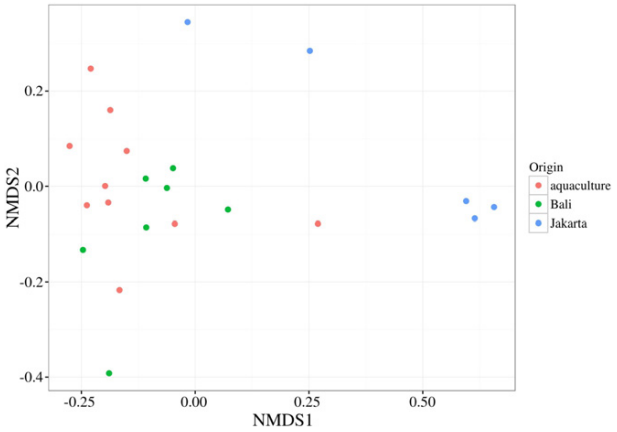


Fig. 3. Non-metric multidimensional scaling (NMDS) ordination plot of bacterial communities based on Bray Curtis distances. The ordination displays the similarity of gut microbiota and cluster as different condition applied. Bali aquaculture and free-living are clustered together apart from Jakarta free-living. Red represents Bali Aquaculture sample, blue represents Bali free-living, and green represents Jakarta free-living.

2.3. Detection of bacterial pathogen

The bacterial community analysis revealed three predominant bacterial orders containing potentially pathogenic species for shrimp and humans. Quantitative PCR profiling was applied to detect a selection of pathogens, by using species specific primers for the most dominant order, *Vibrionales* (Table 1). A total of six samples were used as a representative for Jakarta free-living, and four samples for Bali free-living. On the other hand, due to the sample availability and quality, nine samples were used for Bali aquaculture. A confirmation of 16S and 23S housekeeping genes exhibited positive results in each sample. In addition, the presence of phyla *Bacteriodetes* and *Firmicutes* was also confirmed. Furthermore, a genus specific primer was used to determine the number of contaminated samples for *Pseudomonas* sp., *Pseudoalteromonas*

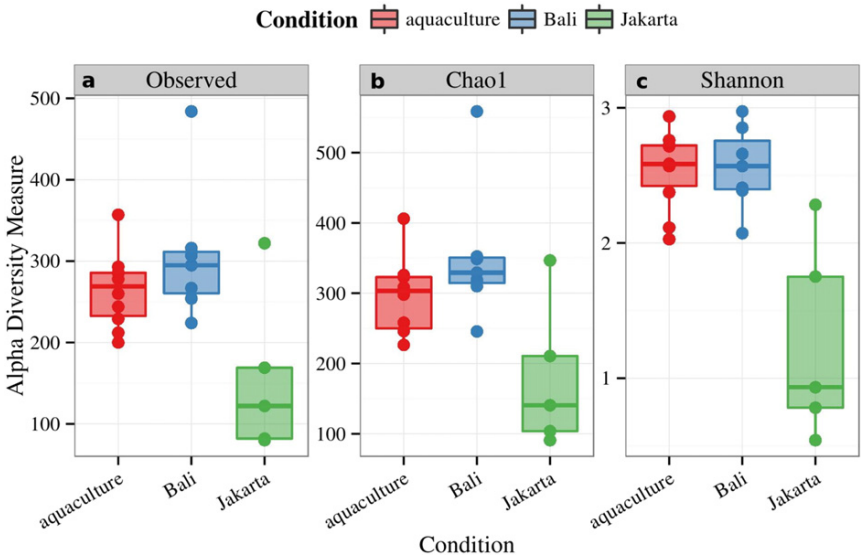


Fig. 2. Diversity analyses for aquaculture, Bali free-living, and Jakarta free-living are based on a. observed OTU richness, b. Chao1 richness estimator, and c. Shannon indices. Color represents the holding condition of the sample, red represents the aquaculture, blue represents Bali free-living, and green represents Jakarta free-living. Dots represent the measurement for every single sample. Boxes represent the quartiles and bars represent the interquartiles. Solid lines represent the median from all samples.

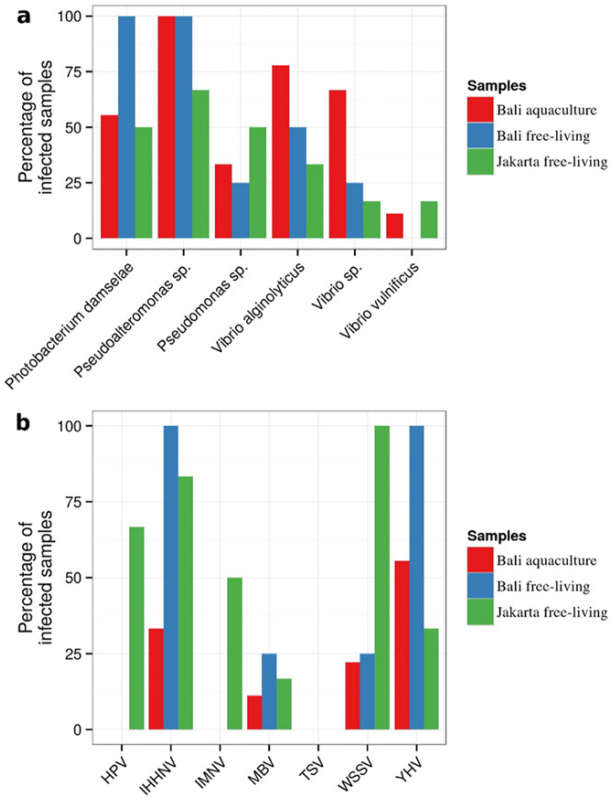


Fig. 4. Results of qPCR profiling for potential bacterial and viral pathogens for each of the three environmental conditions. (a) shows percentages of six different bacterial pathogens. (b) corresponds to seven viral pathogens. All positive samples were divided by total amount of samples, yielded the relative percentage in six different bacterial and seven different viral pathogens. Bar represents the relative percentage of infection level in Y-axis. Names of bacterial and viral pathogens are displayed in the X-axis.

sp. and *Vibrio* sp. In total 50% of the Jakarta free-living samples showed presence for *Pseudomonas* sp., whereas in Bali aquaculture (33.3%) and Bali free-living (25.0%) fewer samples were contaminated. On the other hand, *Pseudoalteromonas* sp., belonging to order *Altermonadales*, was detected in each of the samples from Bali free-living (100%) and Bali aquaculture (100%), in contrast to Jakarta free-living samples, that showed only 67% contamination with *Pseudoalteromonas* sp. The detection for *Vibrio* sp. revealed the highest contamination in Bali aquaculture samples (67.67%), whereas the free-living samples from Bali (25.0%) and Jakarta (16.67%) were less contaminated. Furthermore, to assess potentially pathogenic species inside the *Vibrio* sp., species specific primers for *V. cholerae*, *V. parahaemolyticus*, *V. vulnificus*, and *V. alginolyticus* were utilized, as well as primers for *Photobacterium damselae*, which belongs to *Vibrionales* on order level. *P. damselae* could be detected in each of the samples from Bali free-living (100%), followed by Bali aquaculture (55.56%) and Jakarta free-living (50.0%). For the detection of *V. cholerae*, *V. parahaemolyticus* and *V. vulnificus*, two different primers referring to the *vvh* and *toxR* gene, were employed. *V. cholerae* and *V. parahaemolyticus* could not be detected among any of the samples. *V. vulnificus* was detected in one sample from Jakarta free-living (Jk6) and one sample from Bali aquaculture (Aq6), but not in any of the Bali free-living samples. However, detection of *V. alginolyticus* revealed to be at the highest level of contamination in Bali aquaculture samples (77.78%), followed by Bali free-living (50.00%) and a relatively lower contamination for Jakarta free-living samples (33.33%). Due to the high contamination of *Vibrio* sp. and *Vibrio alginolyticus* in Bali aquaculture and *P. damselae* in Bali free-living samples, the DNA copy number was quantified in order to apprehend the level of infection yielded from qPCR. Standard curves were generated, using a 10 fold dilution of PCR product in the range of 1 ng/μl up to 1 fg/μl (Supplementary Fig. 2). A quantification of DNA copy numbers was obtained, using standard curves in combination with a formula (Table 1). Highest number of DNA copies for *Vibrio* sp. for a specific 16S gene was detected in Bali aquaculture (Aq9: 7.09×10^4 copies), followed Bali free-living (Ba3: 2.02×10^3 copies). Among the Bali aquaculture samples, four samples showed quantifiable number of DNA copies in a decreasing order of abundance: Aq4 (1.83×10^3 copies), Aq8 (1.43×10^3 copies), Aq5 (1.41×10^3 copies) and Aq3 (2.93×10^2 copies). For Jakarta free-living samples the level of infection could only be determined for one sample (Jk6: 8.8×10^2 copies). Concerning the level of infection with *V. alginolyticus*, using *toxR* gene, revealed four samples with quantifiable results, showing the highest number of DNA copies in Bali free-living sample Ba3 (6.48×10^3 copies), followed by Jakarta free-living Jk4 (2.9×10^3 copies) and Bali aquaculture (Aq9: 8.3×10^2 copies, Aq4: 1.42×10^2 copies). Selected DNA copy numbers for *ppb* gene in *P. damselae* showed results in nine samples. The highest level of infection was determined for Bali free-living (Ba3: 4.43×10^3 copies, Ba4: 3.96×10^3 copies and Ba1: 1.5×10^2 copies), and Jakarta free-living (Jk4: 3.1×10^3 copies and Jk6: 1.36×10^3 copies). The samples from Bali aquaculture displayed a low level of infection, in decreasing order of abundance: Aq4 (6.86×10^2 copies), Aq9 (4.99×10^2 copies), Aq5 (1.72×10^2 copies) and Aq2 (1.37×10^2 copies). Samples not mentioned showed no infection or could not be determined. Subsequently, a statistical analysis, based on Kolmogorov–Smirnov, Levene, and Kruskal–Wallis tests was applied. Due to the standard deviation and normality distribution among the samples; there was no significant difference and no correlation between the three different sampling locations, Jakarta free-living, Bali free-living, and Bali aquaculture with the level of infection of *Vibrio* sp. and *V. alginolyticus* and *P. damselae*.

Table 1
qPCR was applied to quantify the DNA copy numbers of sample and simultaneously the detection level from three different specific primers. Respectively, *Vibrio* sp. was determined with specific primers for 16S rRNA gene, *Vibrio alginolyticus* with *toxR* gene, and *Photobacterium damselae* with *ppb* gene. There were several samples that could not be determined because no Ct was obtained from the raw data. There were also several samples with DNA copy numbers below 10^2 that can be considered as insignificant data.

Sample	DNA copy numbers		
	Vibrio sp.(16S)	*Vibrio alginolyticus* (*toxR*)	*Photobacterium damselae* (*ppb*)
Jk2	Nd	6*	3*
Jk3	Nd	Nd	3*
Jk4	Nd	2.9×10^3	3.1×10^3
Jk5	Nd	6*	6*
Jk6	8.8×10^2	60*	1.36×10^3
Ba1	Nd	6*	1.5×10^2
Ba2	Nd	6*	6*
Ba3	2.02×10^3	6.48×10^3	4.43×10^3
Ba4	Nd	33*	3.96×10^3
Aq1	1.08×10^2	6*	6*
Aq2	Nd	42*	1.7×10^2
Aq3	2.93×10^2	18*	30*
Aq4	1.83×10^3	1.42×10^2	6.86×10^2
Aq5	1.41×10^3	24*	1.72×10^2
Aq6	Nd	18*	30*
Aq7	Nd	12*	3*
Aq8	1.43×10^3	36*	60*
Aq9	7.09×10^4	8.3×10^2	4.99×10^2

2.4. Detection of viral diseases

Moreover, we examined seven viral diseases, commonly found in *P. monodon*, using a qPCR approach (Fig. 4b). All Jakarta free-living samples (100%) indicated contamination with White Spot Syndrome Virus (WSSV), whereas samples from Bali free-living (25.0%) and Bali aquaculture (22.22%) revealed lower levels of contamination. The presence of Infectious Hypodermal and

Hematopoietic Necrosis Virus (IHHNV) was confirmed in each of the Bali free-living samples (100%), followed by a high contamination in Jakarta free-living samples (83.33%), and a less presence in Bali aquaculture samples (33.33%). In addition, Yellow Head Virus (YHV) infection was observed in all of the Bali free-living samples (100%), whereas Bali aquaculture (55.56%) and Jakarta free-living samples (33.33%) were less contaminated. The results suggested a low infection for Monodon Baculovirus (MBV), where 25% Bali free-living samples were infected, followed by Jakarta free-living (16.67%) and Bali aquaculture (33.33%). The presence of Hepatopancreatic Parto Virus (HPV) and Infectious Myonecrosis Virus (IMNV) could only be determined in the samples from Jakarta free-living (HPV: 66.67%, IMNV: 50.0%), while the Bali aquaculture and Bali free-living samples showed no infection. The Taura Syndrome Virus (TSV) was not detectable in any of the samples.

3. Discussion

The fecal microbiota in marine organisms, such as fish (Sanchez et al., 2012), oyster (King et al., 2012), and shrimp (Liu et al., 2011; Zhang et al., 2014; Chaiyapechara et al., 2012; Rungrassamee et al.) has been previously reported. As recent findings indicate the positive influence of the microbiome to the host's metabolism and immunity (Balcazar et al., 2006; Guarner and Malagelada, 2003), this may lead to a reduction of the use of antibiotics in aquaculture. In this study, comparison of the bacterial communities in free-living and aquacultured *P. monodon* from two regions in Indonesia were conducted. A next generation sequencing technique — Illumina MiSeq was employed to reveal the fecal microbiome. This study is the first approach using this platform in order to determine bacterial communities of *P. monodon*. Formerly, Roche 454 pyrosequencing was used in parallel with DGGE (Denaturing Gradient Gel Electrophoresis) as a confirmatory step (Rungrassamee et al.). Results revealed *Proteobacteria*, followed by *Bacteriodetes*, *Fusobacteria*, and *Firmicutes* as dominating members of the fecal bacterial community composition. Previous studies support these findings by reporting these bacteria as common members of the microbiome appearing predominantly in many marine organisms (Harris, 1993), especially in shrimp (Chaiyapechara et al., 2012; Rungrassamee et al.; Partida-Arangure et al., 2013). Furthermore, on the order level of taxonomy *Vibrionales* and *Alteromonadales*, belonging to phylum *Proteobacteria*, provide the predominant appearing taxa under Bali free-living and Bali aquaculture condition. In addition, bacterial community composition between these two conditions is very similar, which could be verified by beta diversity analysis. Due to similarities between free-living conditions and conditions of traditional aquaculture, these findings are not uncommon. There were no food additives and probiotics used in the traditional aquaculture (Tang-Nelson and Lightner, 2001), which leads to limited surroundings compared with free-living habitants. Hence, identical bacterial profiles between free-living and aquaculture is observed as such in different model organisms, *Danio rerio* (Roeselers et al., 2011), *Drosophila melanogaster* (Cox and Gilmore, 2007), and *Hydra* (Fraune and Bosch, 2007). In Jakarta free-living samples, *Alteromonadales* could only be detected in two samples, which could be a result of the difference in environmental conditions. In addition, *Sphingobacteriales* and *Fusobacteriales* showed a significant appearance in a few samples, independent from conditions. As previously reported these two orders are commonly found in marine samples (Balcazar et al., 2006; Binns, 2013) and has been observed in previous studies (Chaiyapechara et al., 2012; Rungrassamee et al.; Vaseeharan et al., 2007). Elucidation of rare phyla shows bacterial phyla that are commonly found in shrimp, such as, *Actinobacteria*, *Cyanobacteria*, and *Planctomycetes* (Rungrassamee et al.; Binns, 2013), at only in small portions. Statistical analysis of the elucidated bacterial communities also shows similarities between Bali free-living and Bali aquaculture conditions, and dissimilarities to the Jakarta free-living condition. In reference to the number of unique observed OTUs and Chao1 index, Bali free-living and aquaculture samples show a higher bacterial diversity than Jakarta free-living shrimp samples. The sampling location emphasizes the population dynamics as has been previously reported (Yatsunenko et al., 2012). The bacterial diversity measured by Shannon–Wiener diversity index has also constantly shown that Bali free-living and aquaculture samples have a higher index than Jakarta free-living samples. In addition, a nonmetric multidimensional scaling method points out different clusters between the two different sampling locations, Jakarta Bay and Bali Bay. Whereas Bali free-living and Bali aquaculture samples form a distinct cluster, the Jakarta free-living samples, except two samples (Jk5, Jk2) cluster also together in a distinct cluster. It proves that the geographic factor is associated with the alteration of fecal microbiota. Furthermore, we observed a distinction between the Bali free-living and aquaculture samples. It implies that the free-living samples tend to have a higher trend of bacterial diversity due to the interaction with high variety of bacterial diversity in its ecological surrounding. The environmental turnover under aquaculture condition is lower than under free-living condition, hence the bacterial fluctuation is consequently lower (Sullam et al., 2012; Rönnbäck, 2012). Taking into account these findings reveal that selective pressure of the environment and diet are factors which show influence on the alteration of fecal microbiota. Due to the huge influence of the microbiota on the immunity and metabolism of the shrimp these factors play an important role (Lightner, 2003; Poulos and Lightner, 2006). The profiling of several pathogenic bacteria, belonging to the predominantly order confirmed 16S housekeeping gene and specific genes in phyla *Bacteriodetes* and *Firmicutes*. Genus level primer for *Pseudoalteromonas* sp. reveals a contamination in each sample of Bali free-living and aquaculture. *Pseudoaltermonas* sp. as representative of *Alteromonadales* order, is a common bacterium related to marine organisms, particularly in shrimps (Dheilly et al., 2010). In a previous study, the concern was the antibiotic activity, which was recognized as feasible probiotic (Dheilly et al., 2010). As opposed to *Pseudoalteromonas* sp., *Pseudomonas* sp., showing appearance in some samples, especially from Jakarta free-living, is well-known bacterium, which causes spoilage in fish and shrimp and poses threats to human consumption (Reynisson et al., 2008). Moreover, the contamination of *Vibrio* sp. was highest in Bali aquaculture samples in complete contrast to the other two conditions. It has been reported earlier that *Vibrio* sp. is the cause of disease outbreaks in shrimp aquaculture that led to devastating losses (Venkateswara-Rao, 1998; Iwamoto et al., 2010). However, the existence of *Pseudoalteromonas* sp. in samples derived from aquaculture gives an indication that an immune mechanism protects the host from pathogen invasion (Balcazar et al., 2006; Guarner and Malagelada, 2003; Roeselers et al., 2011). To address the concern about the highly infected aquaculture samples with *Vibrio* sp., species specific primers were applied. The presence of *Vibrio* sp. in aquacultured *P. monodon* could lead to Vibriosis, one of the most prevalent fish diseases. In addition, more than 12 species are known as human pathogens and some of them are pathogenic to marine organisms (Blazer, 1988). As reported earlier, the presence of *V. cholerae* and *V. parahaemolyticus* impose a high threat to humans that could lead to disease as a consequence of toxin intake (Ansede-Bermejo et al., 2012; Ward and Bej, 2006). We could not detect these species in any of the samples, although different set of primers were applied to avoid false positive results. Nonetheless, we found the existence of *V. vulnificus* and *V. alginolyticus* in the samples. *V. vulnificus* has been implicated in primary septicemia in liver disease and wound infection (Bisharat et al., 2005). However, the number of samples infected with *V. vulnificus* was found to be insignificant. *V. alginolyticus*, a human pathogen (Reilly et al., 2011) that can also infect fish and shrimp (Liu et al., 2004; Martins et al., 2010), was significantly found in aquaculture samples. Aside from *Vibrio* sp., *P. damselae* subsp. *piscicida* which is generally found in fish was also detected. It has been reported as a causative agent of Photobacteriosis in fish and shrimp (Rajan et al., 2003), especially in *P. monodon* (Vaseeharan

et al., 2007) and causes great loss in aquaculture. In this study, the dispersion of *P. damselae* was found in all samples irrespective location and conditions. In order to quantify the infections, standard curves of *Vibrio* sp., *V. alginolyticus*, and *P. damselae* were used to obtain a correlation between sampling location and environmental conditions (Lievens et al., 2006), although statistical evaluation revealed no significant correlation between these two factors. However, this study represents the first detection of bacterial pathogens in a comparison of free-living and aquaculture conditions. In addition to bacterial pathogens, viral pathogens represent a major problem for free-living shrimp populations, as well as those in aquaculture conditions. Numerous cases, especially in the United States and Asia (Acre and Moss, 2003) have been documented. Within the scope of this study we revealed contamination by seven different viral pathogens, which have an adverse influence on shrimp health. White Spot Syndrome Virus (WSSV), a double strand DNA virus, belonging to *Nimaviridae*, has been reported as a major cause of economic loss in shrimp aquaculture (Acre and Moss, 2003; Lightner, 2003). Another double strand virus in this study is Monodon Baculovirus (MBV), which is classified as *Baculoviridae*. Single strand DNA viruses, belonging to *Parvoviridae*, such as Hypodermal and Hematopoietic Necrosis Virus (IHHNV) and Hepatopancreatic Parto Virus (HPV) were also identified in this study. Moreover, single strand RNA virus, Yellow Head Virus (YHV), belonging to *Roniviridae* or double strand RNA virus, Infectious Myonecrosis Virus (IMNV), as a member of *Totiviridae* (Acre and Moss, 2003; Lightner, 2003) were detected in several samples. Most free-living *P. monodon* derived from Jakarta are suspected to be infected with WSSV, IHHNV, HPV, and IMNV, whereas free-living *P. monodon* from Bali are mostly infected by IHHNV and YHV. Additionally, the aquaculture samples show only a low level of infection by these viral pathogens, except for YHV. There is a higher frequency for ecological interactions under free-living conditions, this may explain the higher presence of viral pathogen under free-living compared to those under the aquaculture. Many vectors take a role in viral pathogen spreading, as described in case of WSSV (Lightner, 2003). Good aquaculture management leads to less threaten bacterial or viral pathogens, such as the utilization of probiotic that could also reduce the level of viral infection (Partida-Arangure et al., 2013; Balcazar et al., 2006). However, the correlation between viral pathogens and fecal microbiome as an immune response still remains to be revealed. It is important for aquaculture that the issues regarding the analysis of correlations and outcome be further explored.

4. Conclusion

In this study we examine and compare the microbiome of *P. monodon* living in the highly polluted waters of Jakarta Bay, non-polluted waters of Bali Bay and a traditional aquaculture farm in Bali, using a 16S rRNA approach on an Illumina MiSeq platform. We reveal differences in the bacterial community composition between the samples, showing *Vibrionales* and *Altermonadales* as dominating phyla, and the highly-polluted Jakarta Bay samples, dominated by *Vibrionales*. Investigations concerning the bacterial diversity support these findings. In addition, our findings suggest that bacterial diversity in free-living samples is remarkably higher than that in aquaculture samples. A subsequent qPCR profiling on potential bacterial pathogens pointed towards contamination with *V. alginolyticus*, which was detected mainly in the aquaculture samples with greater DNA copy number compared to samples under other conditions. Moreover, potentially pathogenic bacterial species *P. damselae* was found to be prevalent in all studied samples. Furthermore, we detected seven viral pathogens common in shrimp using qPCR profiling. WSSV, HPV, IHHNV, and IMNV were observed in Jakarta free-living samples, whereas YHV was mostly discovered in Bali free-living samples. The aquaculture samples were also slightly infected by viral pathogens.

5. Material and methods

5.1. Sample collection

A total of fourteen *P. monodon* specimen under free-living conditions were collected from Bali (8.101 S 115.087 E, n = 7) and Jakarta Bay (6.028 S 106.840 E, n = 7). Determination of water conditions while sampling showed a salinity of 34 ppt, at a water temperature between 28 °C–32 °C and an pH of 8.4 to 8.5 for Bali Bay, as well as a salinity of 34 to 37 ppt, at 27 °C–34 °C and a pH of 8.5 to 8.9. In addition, ten specimens were also collected from a traditional aquaculture farm in Pejarakan, Singaraja, Bali (n = 10) with similar water conditions to Bali Bay: salinity 24–35 ppt, at 28 °C–32 °C and pH of 8.3 to 8.6. For the traditional aquaculture farm nitrate of 0.1–0.15 mg/l and phosphate of 0.1 mg/l could be measured. The shrimp were fed only with microalgae and phytoplankton that hampered naturally on bottom of the ponds. There were no added food additives and probiotics. Fecal samples were collected while, the collected shrimps both free-living or aquaculture were separated from one another in a tank with 20 l of filtered sea water (the water condition is in average in terms of salinity = 34 ppt, temperature = 30 °C, pH = 8,5) and aerated with oxygen. There was no food supply prior to excretion for 18 to 24 h. After the excretion, the stool was taken and stored in 99.9% EtOH for subsequent DNA extraction.

5.2. Extraction of genomic DNA and 16S rRNA barcoding

DNA was isolated from the fecal samples using QIAamp DNA Stool Mini Kit (Qiagen, Germany) following the manufacturer's instructions, except for the final step where RNase–DNase free water (5 PRIME, Inc., USA) was used for DNA solubilization. Universal prokaryotic primers F515/R806 were used to amplify V4 region of the bacterial 16S rRNA gene (Caporaso et al., 2011, 2012). Subsequently, primers were modified to include Illumina Nextera flowcell adapter sequences, forward and reverse primer pads for the avoidance of primer-dimer formation, and a 2-bp linker sequence not matching against any 16S rRNA sequence immediately upstream of the gene primer (Caporaso et al., 2011). In addition, the reverse primers were incorporated with 12-bp error-correcting Golay barcodes (Caporaso et al., 2011) (Supplementary Table 1). For each individual sample PCR was performed three times, including 10x Fast Buffer 1, 0.25 mM dNTP, 1.25 U TaKaRa SpeedSTAR HS DNA polymerase, 0.2 μM of each primer 20 ng/μl DNA template as a final concentration, and water prior to the reaction volume of 20 μl. Reactions were performed with an initial denaturation step at 95 °C for 5 min, followed by two steps cycled 35 times at 95 °C for 15 s, 64 °C for 45 s, and a final elongation step at 72 °C for 3 min.

5.3. Library preparation and sequencing

PCR products were visualized in a 2% agarose gel (Bio&SELL GmbH, Nürnberg) in bionic buffer 0.5× (Sigma Aldrich, USA) and purified using magnetic beads Agencourt AMPure XP (Beckman Coulter, Inc., USA). Quality and quantity were measured using 2100 Agilent Bioanalyzer (Agilent Technologies, Inc., USA). For the library preparation cleaned PCR-products were pooled in equimolar concentration and subjected to a 250 bp paired-end sequencing on an Illumina MiSeq platform at StarSeq GmbH (Mainz, Germany). Raw sequence data is deposited at NCBI's Short Read Archive under accession number SRP059721.

5.4. Bioinformatic and microbiome analysis

A total of 11,148,953 paired-end sequence reads 250 bp long were obtained in FASTQ format from an Illumina MiSeq platform. An initial quality analysis was done with FastQC (Andrews, 2014) and Trimmomatic (Bolger et al., 2014) by a QC cutoff threshold of

18. Clustering of the reads to Operational Taxonomic Units (OTUs) was performed using the UPARSE pipeline (Edgar, 2013) provided by USEARCH 7.0.1090. Prior to clustering, the paired-end reads were merged using fastq_mergepairs function in the pipeline with minimum overlap of 75 bp and minimum Phred score cutoff threshold of 5. The remaining 7,784,684 sequence reads were trimmed to 250 bp length and additional filtering was performed in the case that the number of errors exceeds 0.5 (fastq_filter). Subsequently, the filtered reads were pooled across samples and dereplication step was done using derep_fulllength function. Eventually, the dereplicated reads were sorted by decreasing abundance. For clustering, the UPARSE-OTU algorithm (cluster_otus) was performed using 7,683,347 high quality sequences as input within a sequence similarity threshold of 97%. The resulting 1382 OTUs were filtered by a reference-based chimera filtering step using the UCHIME algorithm (uchime_ref) (Edgar et al., 2011) and the ChimeraSlayer reference database (http://microbiomeutil.sourceforge.net). In order to create OTU abundance table, the filtered OTUs which were considered as representative sequences, mapped against the filtered reads using sequence identity threshold of 97% (usearch_global). Taxonomic lineages were assigned to the OTU sequences, applying a least common ancestor of the top Blast (Altschul et al., 1990) matches against the NCBI 16SMicrobial database (Sayers et al., 2009). For the assignment, query coverage above 75% and bitscore inside a range of 97% to the best Blast hit were used. The subsequent bacterial community analysis was performed in R (R Core Team, 2014) using the packages phyloseq (McMurdie and Holmes, 2013) and vegan (Oksanen et al., 2016).

5.5. Bacterial pathogen detection using qPCR

A subsequent qPCR was performed for each sample to confirm and determine the potential pathogens, whereas each of the assays was performed twice on a StepOne platform (Life Technologies, USA). qPCR was carried out with two different master mixes, SensiFAST SYBR1 Lo-ROX Kit (Bioline, USA) for a SYBR Green assay and TaqMan Fast Universal PCR Master Mix (Life Technologies, USA) for TaqMan assay. SYBR green assay was performed in 20 μl total volume containing SYBR 2× Lo-ROX master mix, 0.2 μM primers, 20 ng/μl DNA templates as a final concentration, and water. PCR itself was set up at 95 °C for 10 min as a denaturation step, followed by 40 amplification cycles at 95 °C for 15 s and the annealing temperature for 60 s. The primers, master mix, and annealing temperature are listed in Supplementary Table 3. Different optimal annealing temperatures were applied relying on the melting temperature of the primers. Melting temperature curves were also obtained from the PCR reaction. Furthermore, TaqMan assay was done in 20 μl reaction consist of TaqMan Fast Universal PCR Master Mix (2×), 0.6 μM primers, 0.3 μM Probe, 20 ng DNA template as final concentration, and water. Denaturation temperature was set at 95 °C for 5 min, 45 amplification cycles at 95 °C for 15 s and 60 °C for 60 s. Finally, gel electrophoresis was performed for size control, followed by Sanger sequencing (LGC genomics, Berlin, Germany) of the qPCR products for confirmation of the pathogenic sequence by Blast (Altschul et al., 1990).

5.6. Viral pathogen detection using qPCR

The conventional PCR and qPCR were applied for the detection of viral pathogens in shrimp. The conventional PCR was performed and included 10× Fast Buffer 1, 0.25 mM dNTP, 1.25 U TaKaRa SpeedSTAR HS DNA polymerase, 0.2 μM of each primer, 20 ng/μl DNA template as a final concentration, and water prior to the reaction volume of 20 μl. Reactions were performed with an initial denaturation step at 95 °C for 5 min, followed by two steps cycled 40 times at 95 °C for 15 s, 60 °C for 45 s, and a final elongation step at 72 °C for 3 min. On the other hand, StepOne platform (Life Technologies, USA) with TaqMan Fast Universal PCR Master Mix (Life Technologies, USA) for TaqMan assay was occupied for performing the qPCR. TaqMan assay was previously done by Mrotzek et al. (2010). The used primers were listed in Supplementary Table 4. Our detection in WSSV, IHHNV, and MBV were performed as described with no modification. Finally gel electrophoresis was performed for size control.

5.7. Quantification of detected pathogen (Vibrio sp., V. alginolyticus, and P. damselae)

Quantification was done in three genes as representative for the detected potential pathogens, respectively 16S rRNA gene for *Vibrio* sp.; *toxR* specific for *V. alginolyticus*; *ppb* specific for *P. damselae*. A tenfold series of dilution of purified PCR product ranging from 1 ng/μl to 1 fg/μl were calculated to generate a standard curve. Threshold cycle (Ct) versus log DNA concentration was plotted with regression line to establish the standard curve. The PCR efficiency (E) could be accounted from the slope, obtained from the standard curve (Pfaffl, 2004). A calculation to obtain the total copy of three interesting genes was calculated based on the formula (Notes, 2003) described in (Gaibani et al., 2013). Triplicate qPCR for 16S rRNA gene was conducted using TaqMan assay, whereas *toxR* and *ppb* used SYBR Green assay. A melting curve was obtained in order to determine specific peaks in specific melting temperatures in order to validate the result.

Acknowledgments

We thank the Gondol Research Institute for Mariculture GRIM, Bali, Indonesia for providing us with Bali free-living *P. monodon* samples as well as the farmers of traditional shrimp aquaculture in Pedjarakan, Singaraja for providing us with Bali aquaculture *P. monodon* samples and their cooperation. Finally, we would like to acknowledge Harry W. Palm from Aquaculture and Sea-Ranching, Faculty of Agricultural and Environmental Sciences, University of Rostock for providing us with the Jakarta free-living *P. monodon* samples.

Appendix A. Supplementary data

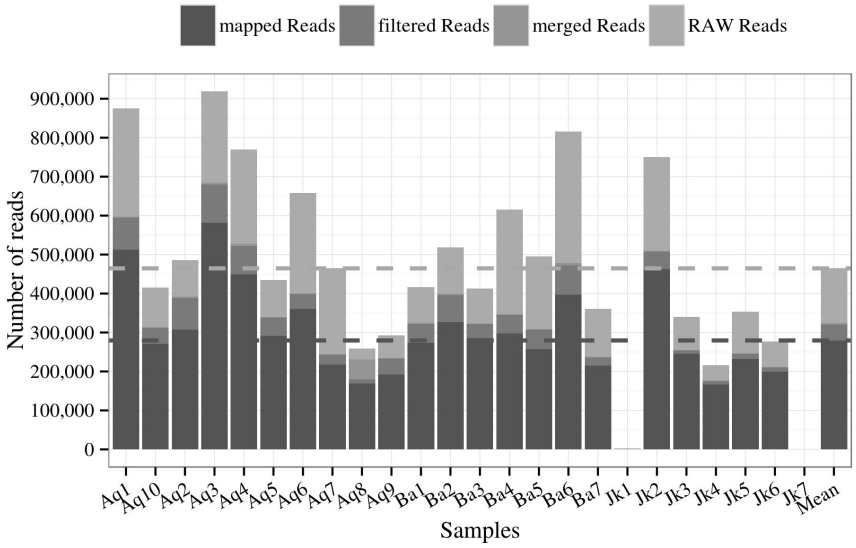
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.marpolbul.2016.03.043.

References

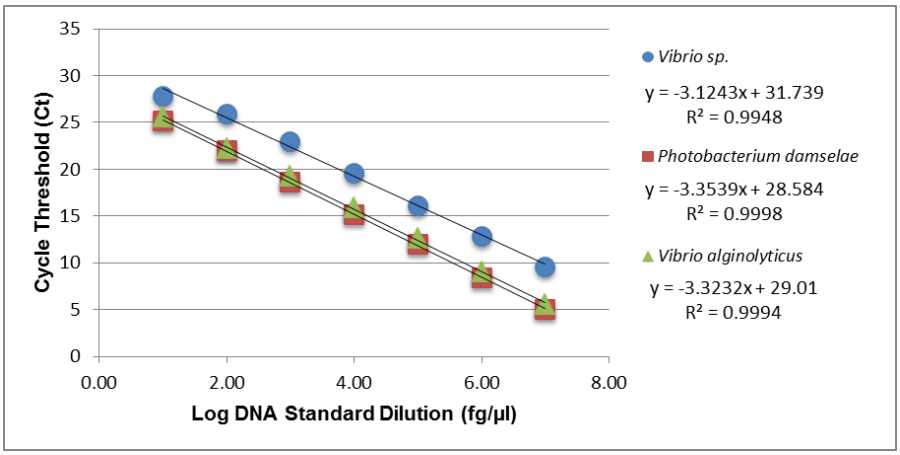
Acre, A.S., Moss, S.M., 2003. Disease prevention strategies for penaeid shrimp culture. Anonime Rosenberry 1–12.
Altschul, S.F., Gish, W., M. W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Andrews, S., 2014. FastQC a quality control tool for high throughput sequence data. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
Ansede-Bermejo, Miranda, A., Rodriguez-Alvarez, M.X., 2012. Ecological determinants of the occurrence and dynamics of *Vibrio parahaemolyticus* in offshore areas. ISME 6 (5), 994–1006.
Asplund, M., 2013. Ecological Aspects of Marine Vibrio Bacteria-exploring Relationships to Other Organisms and A Changing Environment (dissertation) University of Gothenburg.
Balcazar, J.L., Blas, I.D., Ruiz-Zarzuela, I., Cunningham, D., Vendrell, D., Muzquiz, J.L., 2006. The role of probiotics in aquaculture. Vet. Microbiol. 114, 173–186.
Binns, N., 2013. Probiotics, Prebiotics and the Gut Microbiota.
Bisharat, N., Cohen, D.I., Harding, R.M., Falush, D., Crook, D.W., Peto, T., et al., 2005. Hybrid *Vibrio vulnificus*. Emerg. Infect. Dis. 11, 30–35.
Blazer, V.S., 1988. Bacterial fish pathogens. Environ. Biol. Fish 77–79.
Bolger, A., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics.
Bray, J.R., Curtis, J.T., 1957. An ordination of upland forest communities of southern wisconsin. Ecol. Monogr. 27, 325–349.
Caporaso, J.G., et al., 2011. Global patterns of 16s RRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. 108, 4516–4522.
Caporaso, J.G., et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624.
Chaiyapechara, S., Rungrassamee, W., Suriyachay, I., Kuncharin, Y., Klanchui, A., K.N., et al., 2012. Bacterial community associated with the intestinal tract of *P. monodon* in commercial farms. Microb. Ecol. 63, 938–953.

Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Stat. 11 (4), 265–270 (URL http://www.jstor.org/stable/4615964).
Cox, C.R., Gilmore, M.S., 2007. Native microbial colonization of *Drosophila melanogaster* and its use as a model of *Enterococcus faecalis* pathogenesis. Infect. Immun. 75, 1565–1576.
Dheilly, A., Soum-Soutera, E., Klein, G.L., Bazire, A., Compere, C., Haras, D., et al., 2010. Antibiofilm activity of the marine bacterium *Pseudoalteromonas* sp. strain 3j6. Appl. Environ. Microbiol. 76 (201), 3452–3461.
Edgar, R.C., 2013. Uparse: highly accurate otu sequences from microbial amplicon reads. Nat. Methods 10, 996–998.
Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. Uchime improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.
FAO: Food Agric Organ United Nations, 2014. *Penaeus monodon*, Website.
Fraune, S., Bosch, T.C.G., 2007. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan hydra. Proc. Natl. Acad. Sci. U. S. A. 104, 13146–13151.
Gaibani, P., Mariconti, M., Bua, G., Bonora, S., Sassera, D., Landini, M.P., et al., 2013. Development of a broad-range 23s rDNA real-time PCR assay for the detection and quantification of pathogenic bacteria in human whole blood and plasma specimens. Biomed Res. Int. 2013, 13–17.
Guarner, F., Malagelada, J.R., 2003. Gut flora in health and disease. Lancet 360, 512–519.
Harris, J., 1993. The presence, nature, and role of gut microflora in aquatic invertebrates: a synthesis. Microb. Ecol. 25, 195–231.
Iwamoto, M., Ayers, T., Mahon, B.E., Swerdlow, D.L., 2010. Epidemiology of seafood-associated infections in the united states. Clin. Microbiol. Rev. 23, 399–411.
King, G.M., Judd, C., Kuske, C.R., Smith, C., 2012. Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal louisiana, usa. PLoS One 7.
Lievens, B., Brouwer, M., Vanachter, A.C.R.C., Cammue, B.P.A., Thomma, B.P.H.J., 2006. Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. Plant Sci. 171, 155–165.
Lightner, D.V., 2003. The penaeid shrimp viral pandemics due to IHHNV, WSSV, TSV and YHV: history in the Americas and current status. Tech. rep., Tucson, AZ 85721 USA.
Liu, C.H., Cheng, W., Hsu, J.P., Chen, J.C., 2004. *Vibrio alginolyticus* infection in the white shrimp litopenaeus vannamei confirmed by polymerase chain reaction and 16s rDNA sequencing. Dis. Aquat. Org. 61, 169–174.
Liu, H., Wang, L., Liu, M., Wang, B., Jiang, K., Ma, S., et al., 2011. The intestinal microbial diversity in chinese shrimp (*Fenneropenaeus chinensis*) as determined by PCR-DGGE and clone library analyses. Aquaculture 317, 32–36. http://dx.doi.org/10.1016/j.aquaculture.2011.04.008.
Luis-Villasenor, I.E., Campa-Cordova, A.I., Ascencio-Valle, F.J., 2012. Probiotics in larvae and juvenile whiteleg shrimp *Litopenaeus vannamei*. InTech http://dx.doi.org/10.5772/50123.
Martins, M.L., Mourino, J.L.P., Fezer, G.F., Buglione Neto, C.C., Garcia, P., Silva, B.C., et al., 2010. Isolation and experimental infection with *Vibrio alginolyticus* in the sea horse, *Hippocampus reidi* Ginsburg, 1933 (Osteichthyes: Syngnathidae) in Brazil. Braz. J. Biol. 70, 205–209.
McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.14.
Mrotzek, G., Haryanti, Koesharyani, I., Tretyakov, A.N., Sugama, K., Saluz, H.P., 2010. Fast short-fragment PCR for rapid and sensitive detection of shrimp viruses. J. Virol. Methods 168, 262–266.
Notes, I., 2003. Creating standard curves with genomic dna or plasmid dna templates for use in quantitative PCR. ReVision 2003, pp. 1–8.
O.L., La Fauce, K.A., Layton, R., 2007. TaqMan real-time PCR for detection of hepatopancreatic parvovirus from australia. J. Virol. Methods 140, 10–16.
Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Henry, H., Stevens, H., W. H., 2016. Vegan: Community Ecology Package.
Partida-Arangure, B.O., Luna-Gonzalez, A., Fierro-Coronado, J.A., F.-M. C., Gonzalez-Ocampo, H., 2013. Effect of inulin and probiotic bacteria on growth, survival, immune response, and prevalence of white spot syndrome virus (WSSV) in litopenaeus vannamei cultured under laboratory conditions. Afr. J. Biotechnol. 12, 3366–3375.
Pfaffl, M., 2004. Quantification strategies in real-time PCR. A-Z Quant PCR 2004, pp. 87–112.
Poulos, B.T., Lightner, D.V., 2006. Detection of infectious myonecrosis virus (IMNV) of penaeid shrimp by reverse-transcriptase polymerase chain reaction (RT-PCR). Dis. Aquat. Org. 73, 69–72. http://dx.doi.org/10.1186/1471-2164-15-266.
R Core Team, jan 2014. R: a language and environment for statistical computing. URL http://www.r-project.org.
Rajan, P.R., Lin, J.H.Y., Ho, M.S., Yang, H.L., 2003. Simple and rapid detection of *Photobacterium damselae* ssp. piscicida by a PCR technique and plating method. J. Appl. Microbiol. 95, 1375–1380.
Reilly, G.D., Reilly, C.A., Smith, E.G., Baker-Austin, C., 2011. *Vibrio alginolyticus*-associated wound infection acquired in British waters, guernsey, July 2011. Eurosurveillance 16, 3.
Reynisson, E., Lauzon, H.L., Magnusson, H., Hreggvidsson, G.O., Marteinsson, V.T., 2008. Rapid quantitative monitoring method for the fish spoilage bacteria *Pseudomonas*. J. Environ. Monit. 10, 1357–1362.
Roeselers, G., et al., 2011. Evidence for a core gut microbiota in the zebrafish. ISME J. 5, 1595–1608.
Rönnbäck, P., 2002. Environmentally Sustainable Shrimp Aquaculture. Swedish Society for Nature Conservation.
W. Rungrassamee, A. Klanchui, S. Maibunkaew, S. Chaiyapechara, P. Jiravanichpaisal, N. Karoonuthaisiri, Characterization of intestinal bacteria in wild and domesticated adult black tiger shrimp (*Penaeus monodon*)., PLoS One 9.
Sanchez, L.M., Wong, W.R., Riener, R.M., Schulze, C.J., Linington, R.G., 2012. Examining the fish microbiome: vertebrate-derived bacteria as an environmental niche for the discovery of unique marine natural products. PLoS One 7 (5).
Sayers, et al., 2009. Database resources of the national center for biotechnology information. Nucleic Acids Res. 37, D5–15.
Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.
Sullam, K.E., et al., 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Mol. Ecol. 21, 3363–3378.
Tang-Nelson, K., Lightner, D.V., 2001. Development of real-time PCR assays for detection of white spot syndrome virus, yellow head virus, taura syndrome virus, and infectious hypodermal and hematopoietic necrosis virus in penaeid shrimp. Tech. Rep., Tucson, AZ 85721.
Vaseeharan, B., Sundararaj, S., Murugan, T., Chen, J.C., 2007. *Photobacterium damselae* ssp. *damselae* associated with diseased black tiger shrimp *Penaeus monodon* fabricius in India. Lett. Appl. Microbiol. 45, 82–86.
Venkateswara-Rao, 1998. Vibriosis in shrimp. Aquaculture 1–9.
Ward, L.N., Bej, A.K., 2006. Detection of *Vibrio parahaemolyticus* in shellfish by use of multiplexed real-time PCR with TaqMan fluorescent probes detection of *Vibrio parahaemolyticus* in shellfish by use of multiplexed real-time PCR with TaqMan fluorescent probes. Appl. Environ. Microbiol. 72, 2031–2042.
Wegner, K.M., Volkenborn, N., Peter, H., Eiler, A., 2013. Disturbance induced decoupling between host genetics and composition of the associated microbiome. BMC Microbiol. 13 (1), 252.
Xia, J., et al., 2014. The intestinal microbiome of fish under starvation. BMC Genomics 15 (1), 266. http://dx.doi.org/10.1186/1471-2164-15-266.
Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., et al., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227.
Zhang, M., Sun, Y., Chen, K., Yu, N., Zhou, Z., Chen, L., Du, Z., Li, E., 2014. Characterization of the intestinal microbiota in pacific white shrimp, litopenaeus vannamei, fed diets with different lipid sources. Aquaculture 434, 449–455.

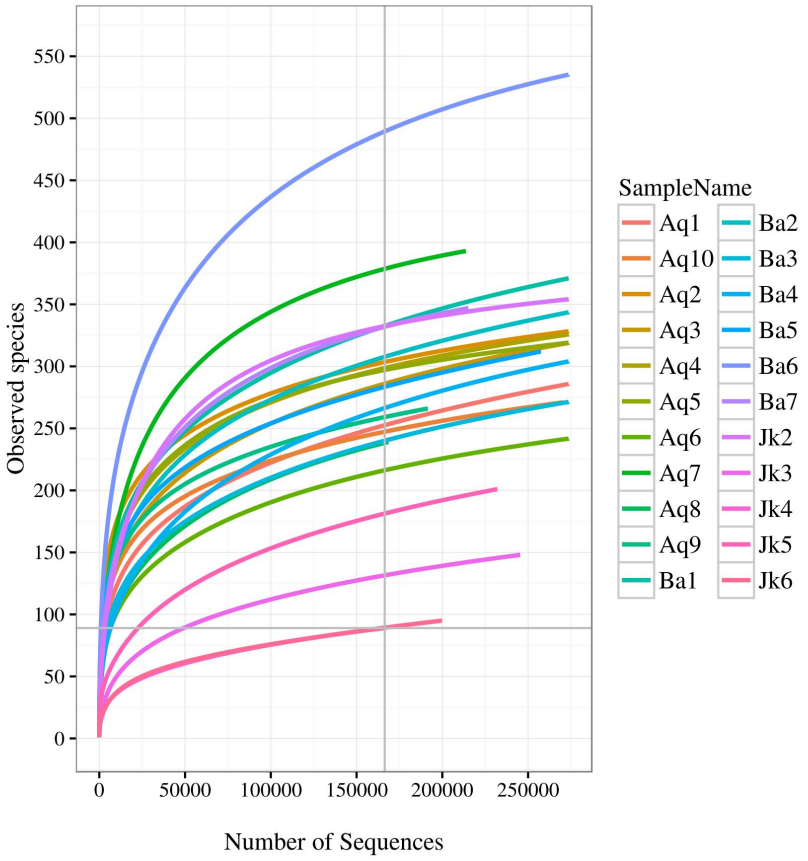
Supplementary



Supplementary Figure 1: **The Illumina MiSeq Sequencing Reads Processing Stages in aquaculture, Bali free-living, and Jakarta free-living samples.** Grey shows the raw reads that were gained from Illumina MiSeq. Shade of grey shows the merged reads from the paired of forward and reverse. Shade of black shows the filtered reads from chimera. Black shows the mapped reads that was assigned to the OTUs. The grey dashed line show the maximum reads and black dashed line show the minimum reads in average. Aq1 – Aq10 are Aquaculture samples, Ba1 – Ba7 are Bali free-living samples. Jk1 – Jk7 are Jakarta free-living samples.



Supplementary Figure 2: **Standard curves for quantification of detected potential pathogen.** Standard curves were established from 10 fold dilution series 1 ng – 1 fg/μl DNA of *Vibrio* sp. (16S rRNA gene), *Vibrio alginolyticus* (*toxR*), and *Photobacterium damselae* (*ppb*) vs. Cycle threshold (Ct). Circle represents *Vibrio* sp. (slope = -3.1243 E = 108.96%), rectangle represents *Vibrio alginolyticus* (slope = -3.3232 E = 99.95%), and triangle represents *Photobacterium damselae* (slope = -3.3539 E = 98.68%).



Supplementary Figure 3: **Rarefaction curves were constructed by counting the number of species and increasing the number of sequence reads** A flatter curve to the right indicate a reasonable number of individual samples was taken. More intensive sequencing will only achieve few new species

***P. monodon* Sample**	**Barcode Number**	**Barcode - Sequence**
Ba3	2	GCTGTACGGATT
Aq7	4	TGGTCAACGATA
Aq8	5	TGGTCAACGATA
Jk6	6	GTCGTGTAGCCT
Ba4	7	AGCGGAGGTTAG
Ba7	8	ATCCTTTGGTTC
Ba1	12	TGCATACACTGG
Ba2	13	AGTCGAACGAGG
Ba5	15	GAATACCAAGTC
Ba6	18	CATTATGGCGTG
Jk1	19	CCAATACGCCTG
Aq1	20	GATCTGCGATCC
Jk2	21	CAGCTCATCAGC
Jk3	25	CGAGCAATCCTA
Jk4	27	GTATCTGCGCGT
Aq9	28	CGAGGGAAAGTC
Aq2	29	CAAATTCGGGAT
Aq10	31	AGTTACGAGCTA
Aq3	32	GCATATGCACTG
Jk7	33	CAACTCCCGTGA
Aq4	34	TTGCGTTAGCAG
Aq5	35	TACGAGCCCTAA
Jk5	36	CACTACGCTAGA
Aq6	37	TGCAGTCCTCGA

Supplementary Table 1: **List of the used barcode sequences**

Sample Name	Origin	longitude, latitude	Environment	sex/gender	weight	height	Number of reads
Aq1	Bali	8.114 S 115.089 E	aquaculture	female	123.4	23.4	277.594
Aq2	Bali	8.114 S 115.089 E	aquaculture	female	26.5	15.5	94.173
Aq3	Bali	8.114 S 115.089 E	aquaculture	male	106.1	23.2	234.534
Aq4	Bali	8.114 S 115.089 E	aquaculture	female	54.1	17.9	241.669
Aq5	Bali	8.114 S 115.089 E	aquaculture	male	75.0	20.5	95.332
Aq6	Bali	8.114 S 115.089 E	aquaculture	female	22.4	15.5	257.711
Aq7	Bali	8.114 S 115.089 E	aquaculture	female	144.5	22.5	220.328
Aq8	Bali	8.114 S 115.089 E	aquaculture	male	79.0	20.5	27.802
Aq9	Bali	8.114 S 115.089 E	aquaculture	female	184.2	25.5	58.019
Aq10	Bali	8.114 S 115.089 E	aquaculture	male	79.0	20.5	101.932
Ba1	Bali	8.101 S 115.087 E	free-living	female	21.8	14.0	91.882
Ba2	Bali	8.101 S 115.087 E	free-living	female	31.5	16.0	119.740
Ba3	Bali	8.101 S 115.087 E	free-living	male	19.1	14.0	89.271
Ba4	Bali	8.101 S 115.087 E	free-living	female	50.5	17.5	268.610
Ba5	Bali	8.101 S 115.087 E	free-living	male	87.7	22.0	186.624
Ba6	Bali	8.101 S 115.087 E	free-living	male	92.0	22.0	336.960
Ba7	Bali	8.101 S 115.087 E	free-living	female	17.8	14.0	122.850
Jk1	Jakarta	6.028 S 106.840 E	free-living	-	-	-	1.944
Jk2	Jakarta	6.028 S 106.840 E	free-living	-	-	-	239.914
Jk3	Jakarta	6.028 S 106.840 E	free-living	-	-	-	87.736
Jk4	Jakarta	6.028 S 106.840 E	free-living	-	-	-	39.429
Jk5	Jakarta	6.028 S 106.840 E	free-living	-	-	-	107.469
Jk6	Jakarta	6.028 S 106.840 E	free-living	-	-	-	65.604
Jk7	Jakarta	6.028 S 106.840 E	free-living	-	-	-	142

Supplementary Table 2: **Sample Metadata Table shows a list of features, regarding the sampling location and environment, as well as the initial RAW sequence number per sample.**

Bacterial type (Target gene)	**Taxonomic rank**	**Master Mix type**	**Amplicon size (bp)**	**Sequences (5' → 3')**	**References**
All bacteria (16S)	superkingdom	200	SYBR®	F: ACTCCTACGGGAGGCAGCAG R :ATTACCGCGGCTGCTGG	[60]
All bacteria (16S)	superkingdom	466	SYBR®	F: TCCTACGGGAGGCAGCAGT R: GACTACCAGGGTATCTAATCCTGTT	[61]
All bacteria (23S)	superkingdom	97	SYBR®	F: TCGCTCAACGGATAAAAG R: GATGANCCGACATCGAGGTGC	[59]
Bacteriodetes (16S)	phylum	126	SYBR®	F: GGARCATGTGGTTTAATTCGATGAT R: AGCTGACGACAACCATGCAG	[62]
Firmicutes (16S)	phylum	126	SYBR®	F: GGAGYATGTGGTTTAATTCGAAGCA R: AGCTGACGACAACCATGCAC	[62]
Pseudomonas sp. (carA)	genus	165	TaqMan®	F: GGCTTTCAGGTAGTCGGACAG R: CAACAGATCGTTACCCTGACTT Probe: FAM-GCCAGTTGCTCGC-BHQ1	[35]
Pseudoalteromonas sp. (16S)	genus	180	TaqMan®	F: CGAACTGGCAAACTAGAGTGTGAG R: CCGAGGCTCCGAGCTTCTA Probe: FAM-CACTGACGCTCATGTAC-BHQ1	[35]
Vibrio sp. (16S)	genus	180	TaqMan®	F: TGAAACTGGTGAACTAGAGTGCTGT R: CTCAAGGCCACAACCTCCA Probe: FAM-CTGACACTCAGATGCGA-BHQ1	[34]
Vibrio cholera (hylA)	species	70	TaqMan®	F: TGCGTTAAACACGAAGCGAT R: AAGTCTTACATTGTGCTTGGGTCA Probe: FAM-TCAACCGATGCGATTGCCCAAGA-BHQ1	[63]
Vibrio cholera (toxR)	species	375	SYBR®	F: CAGGTTTGYTGCACGGCGAAGA R: AGCAGCTTATGACCAATACGCC	[64]
Vibrio vulnificus (vvh)	species	276	TaqMan®	F: CGCTGTTTAACGGCCAGCTA R: GGTTGTCATTCTCGTCGGTG Probe: FAM-ACAGCAACCGAGACGAAATCACTCAAG-BHQ1	[65]
Vibrio vulnificus (toxR)	species	412	SYBR®	F: CAGGTTTGYTGCACGGCGAAGA R: GTACGAAATTCTGACCGATCAA	[64]
Vibrio parahaemolyticus (trh)	species	207	TaqMan®	F: CCATCMATACCTTTTCCTTCTCC R: ACYGTCATATAGGCGCTTAAC Probe: FAM-TATTTGTYGTTAGAAATACAACAAT-BHQ1	[39]
Vibrio parahaemolyticus (tdh)	species	229	TaqMan®	F: GTARAGGTCTCTGACTTTTGGAC R: CTACAGAATYATAGGAATGTTGAAG Probe: FAM-ATTTTACGAACACAGCAGAATGA-BHQ1	[39]
Vibrio alginolyticus (toxR)	species	144	SYBR®	F: ATTGAGAACCCGACAGAAGCGAAG R: CCTAATGCGGTGATCAGTGTTACT	[64]
Photobacterium damselae (ppb)	species	297	SYBR®	F: CCGACTCAACTACAGATCACCCAGTC R: GTGCGGCCTAAATTTCGACGA	[66]

Supplementary Table 3: **Used Primers and Probes for qPCR**

Virus	Amplicon size (bp)	Master Mix type	Sequences (5' → 3')	References
WSSV	154	TaqMan®	F: CCAGTTCAGAATCGGACGTT	[67]
(White Spot Syndrome Virus)			R: AAAGACGCCTACCCTGTTGA	
			Probe: FAM-TCCATAGTTCCTGGTTTTGTAATGTGCCG-BHQ1	
IHHNV	77	TaqMan®	F : AAACTGAACACTGGCCTAGTAACAA	[68]
(Infection Hypodermal and Hematopoietic Necrosis Virus)			R : GGCCAAGACCAAAATACGAA	
			Probe: FAM-AACAGGAGACTCAAACACCTTCCATCT-BHQ1	
MBV	135	TaqMan®	F: ACTACCATAAGCTAGCATACGTCCTTT	[69]
(Manudon Baculovirus)			R: AAAGGTCAGCAAAAAACACTCAATT	
			Probe: FAM-ACCCTCTACCGATATGGTATCAATGTCTGGAGTT-BHQ1	
HPV	140	SYBR®	F: CTACTCCAATGGAAACTTCTGAGC	[18]
(Hepatopancreatic Parto Virus)			R: GTGGCGTTGGAAGGCACTTC	
IMNV	139	SYBR®	F: GGCACATGCTCAGAGACA	[33]
(Infectious Mionecrosis Virus)			R: AGCGCTGAGTCCAGTCTTG	
TSV	79	SYBR®	F: CGTAAATAGACGGCCCACAAA	[70]
(Taura Syndrome Virus)			R: TGCATCTATATATCCAGGGACTTATCC	
YHV	147	SYBR®	F: ACATCTGTCCAGAAGGCGTC	[71]
(Yellow Head Virus)			R: GGGGGTGTAGAGGGAGAGAG	

Supplementary Table 4: **Used Primers and Probes for qPCR in Viral Disease Detection**

References

Amagliani, G., Omiccioli, E., Andreoni, F., Boiani, R., Bianconi, I., Zaccone, R., et al., 2009. Development of a multiplex PCR assay for *Photobacterium damselae* subsp. piscicida identification in fish samples. J. Fish Dis. 32 (8), 645–653.

Aranguren, L.F., Salazar, M., Tang, K., Caraballo, X., Lightner, D., 2013. Characterization of a new strain of taura syndrome virus (TSV) from colombian shrimp farms and the implication in the selection of TSV resistant lines. J. Invertebr. Pathol. 112 (1), 68–73.

Fierer, N., Jackson, J., 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl. Environ. Microbiol. 71 (7), 4117.

Guo, X., Xia, X., Tang, R., Zhou, J., Zhao, H.,Wang, K., 2008. Development of a real-time PCR method for firmicutes and bacteroidetes in feces and its application to quantify intestinal population of obese and lean pigs. Lett. Appl. Microbiol. 47 (5), 367–373.

Jang, I.K., Meng, X.H., Seo, H.C., Cho, Y.R., Kim, B.R., Ayyaru, G., et al., 2009. A TaqMan real-time PCR assay for quantifying white spot syndrome virus (WSSV) infections in wild broodstock and hatchery-reared postlarvae of fleshy shrimp, fenneropenaeus chinensis. Aquaculture 287 (1-2), 40–45.

Lyon, W.J., 2001. TaqMan PCR for detection of vibrio cholerae Ö1, o139, non-o1, and non-o139 in pure cultures, raw oysters, and synthetic seawater†. Appl. Environ. Microbiol.
67 (10), 4685–4693.

Nadkarni, M.A., Martin, F.E., Jacques, N.A., Hunter, N., 2002. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology 148 (1), 257–266.

Tall, A., Teillon, A., Boisset, C., Delesmont, R., Touron-Bodilis, A., Hervio-Heath, D., 2012. Real-time PCR optimization to identify environmental Vibrio spp. strains. J. Appl. Microbiol. 113 (2), 361–372.

Wang, H.Y., Lee, G.H., 2004. Rapid identification of Vibrio vulnificus in seawater by real-time quantitative TaqMan PCR. J. Microbiol. 01, 320–326.

Wijegoonawardane, P.K.M., Cowley, J.A.,Walker, P.J., 2008. Consensus RT-nested PCR detection of yellow head complex genotypes in penaeid shrimp. J. Virol. Methods 153 (2), 168–175.

Xie, Zhixun, Liji, X., Yaoshan, P., Zhaofa, L., Zhiqin, X., Jianhua, S., Xianwen, D., Jiabo, L., Xiaofei, T., M. K., 2008. Development of a real-time multiplex PCR assay for detection of viral pathogens of penaeid shrimp. Arch. Virol. 153 (12), 2245–2251.

Yan, D., Tang, K.F.J., Lightner, D.V., 2009. Development of a real-time PCR assay for detection of monodon baculovirus (MBV) in penaeid shrimp. J. Invertebr. Pathol. 102 (2), 97–100.

4 Discussion

The emergence and evolution of metazoan organisms is strongly connected to bacterial life. The number of bacteria living on the internal and external surfaces of metazoans are thought to outnumber their host cells by a factor of 10 to 1 [13, 14]. This close relationship between the metazoan host and its microbiota has a significant influence on the innate and adaptive immunity of the host organism. It is also widely accepted that numerous metabolic processes vital to host fitness and survival are influenced or facilitated by the microbial community [45]. While the microbiome, which is defined as the collection of genome and genes of microbiota, including the surrounding environmental conditions, has been well studied for a few model organisms, little is known about the microbiome of marine organisms.
The aim of this thesis is to investigate the fecal microbiomes of three important Indonesian fish species, namely the *Atule mate*, *Epinephelus sexfasciatus* and *Epinephelus fuscoguttatus* as well as the shrimp species *Penaeus monodon*, in order to find differences in the bacterial community composition between polluted and non-polluted environments. In addition, the detection of a core microbiome for each host species as well as a shared microbiome between the three fish species is examined. Furthermore, the metagenome of one specific host species (*E. fuscoguttatus*) is investigated and analyzed for the eukaryotic taxa and the functional content of the microbiome. A 16S amplicon and whole metagenome shotgun approach were used to achieve these findings.

4.1 The microbiome of marine organisms

Approximately 71% of the earth's surface is covered by ocean [46], which represents the largest connected ecosystem on earth. This environment and its enormous biodiversity offer a high potential for the discovery of new organisms as well as new bioactive compounds. In particular, the Up to now, existing studies have dealt with the microbiome of marine organisms, with most of them focusing on host species that are important for aquaculture. The intestinal and gut microbiome of different marine organisms, such as fish [47, 48], shrimp [49, 50] and oysters [51], have been investigated. Interactions were found between the microbiome and the metabolism [47] and immunity [52, 53] of the investigated host species.
Investigations of the fecal microbiome of the three fish species, *A. mate*, *E. sexfasciatus* and *E. fuscoguttatus* as well as the shrimp *P. monodon*, have shown *Proteobacteria* to be the predominant phylum, followed by *Firmicutes* and *Actinobacteria* for fish host species and *Bacteriodetes*, *Fusobacteria* and *Firmicutes* for the fecal

bacterial community of shrimp. In addition, *Spirochaetes* have been detected as a predominant phylum for *E. fuscoguttatus*, by a whole metagenomic shotgun sequencing approach. Previous studies on gut and fecal bacterial communities support these findings by showing that most of these phyla are common members of the microbiome of marine organisms. With respect to the bacterial communities of *P. monodon*, a number of studies confirm the predominance of *Proteobacteria*, *Bacteriodetes*, *Fusobacteria* and *Firmicutes* [54, 55, 56], while studies on the gut and fecal microbiome of fish confirm that *Proteobacteria* and *Firmicutes* [47, 57, 58] are common members of the fecal microbiome. In addition, *Actinobacteria* and *Spirochaetes* are not reported in high proportions, whereas *Bacteriodetes*, a common member in other studies, only appeared in lower proportions as a rare phylum.

4.1.1 Microbiome differences between environmental conditions

In addition to an independent investigation of microbiomes, a comparison of the differences in bacterial community composition under different environmental conditions is one of the main objectives of this thesis. To do this, samples of marine host species were collected from different areas of Java and Bali, where the environmental conditions are known. Unfortunately, the cooperation partner from the University of Rostock, who was responsible for collecting the samples, failed to take water samples in order to make a more detailed determination of the level of pollution.
All the marine host species, except for *E. fuscoguttatus*, were collected in the waters of Jakarta Bay for the representation of polluted waters. Jakarta, the biggest city in Indonesia with over ten million inhabitants (2014), is a booming coastal megacity [59]. Enormous amounts of untreated wastewater, mainly from households and industries, are discharged into thirteen rivers, which flow through the city and into the bay area of Jakarta [40, 41]. Various studies have reported on the heavy, multifaceted pollution of the marine environment in the Jakarta Bay area and the resulting marked decrease in corals over the last 30 years [60]. In contrast to these highly polluted waters, the bay area of Cilacap and Bali represent cleaner water bodies. For the fish host species *A. mate* and *E. sexfasciatus*, there is no information about exact sampling locations or water quality available because each specimen was bought from local fishermen at the harbor of Cilacap. Furthermore, more general information about the water qualities in the region around Cilacap could also not be obtained. Most of the literature is specific to regional lagoons and mangrove forests or more than 20 years old. Therefore an evaluation of effective water conditions is difficult for this places. A few information of the water quality conditions (e.g., salinity, water temperature and pH) were obtained for the samples of *P. monodon* from Bali Bay and a traditional aquaculture, but the measured conditions were more

or less similar.

First, the differences in the bacterial community compositions of the highly polluted waters of Jakarta Bay and cleaner water conditions around Cilacap were investigated in the samples of *A. mate* and *E. sexfasciatus*. Due to the low sample size of these host species (*A. mate*: n=2 per sample location, *E. sexfasciatus*: n=3 per sample location), the differences of the samples should be treated with caution. The distribution of the predominant bacteria in *A. mate* does not show any differences between the polluted and non-polluted environment. This result is confirmed by a beta diversity analysis, which showed only one cluster containing each specimen of *A. mate* independent of environment. However, the bacterial community composition of *E. sexfasciatus* showed an increased presence of *Firmicutes* in the samples from Jakarta Bay. While *Firmicutes* was the predominant bacteria in samples es1 and es2 and also showed an increased abundance in es3, the samples from Cilacap were completely dominated by *Proteobacteria*. Only one sample (es5) had a slightly increased abundance of *Firmicutes*. The supporting beta diversity analysis points to a distinct cluster in each environment. Only sample es5, which contained the highest amount of *Firmicutes* in the Cilacap samples, is represented as an outlier. A more detailed investigation of the differences was not possible due to the insufficient depth of the taxonomic annotation level.

In order to study the fecal microbiomes of *Penaeus mondon*, samples were taken from three different environments: free-living specimens from Jakarta Bay and Bali Bay as well as from a traditional aquaculture in Bali. *Proteobacteria*, followed by *Bacteriodetes*, *Fusobacteria* and *Firmicutes* were identified as dominant members of the fecal microbiome. The presence of these phyla as common predominant members of many marine organisms and especially shrimp has also been confirmed by other studies [54, 55, 61]. Differences between the environments could be observed at the order level of taxonomy. While *Vibrionales* and *Alteromonadales*, which belong to the phylum of *Proteobacteria*, are relatively similarly distributed across Bali free-living and Bali aquaculture, three samples from Jakarta Bay show higher numbers of *Vibrionales* and the absence of *Alteromonadales*. The remaining two samples from Jakarta Bay only have a minor presence of *Alteromonadales*. In addition, the remaining bacterial community compositions from the Bali environments were very similar. A subsequent beta diversity analysis supports these findings by showing two distinct clusters for these two environments, with the samples from Jakarta Bay without *Alteromonadales* forming one distinct cluster. Similarities between a free-living environment and the environmental conditions of a traditional aquaculture are not uncommon. Due to the absence of food additives and antibiotics in a traditional aquaculture system, the differences between this environment and a free-living habitat are limited. Thus, similar profiles between free-living and aquaculture specimens have been observed in different model organisms such as *Danio*

rerio [62], *Drosophila melanogaster* [63] and *Hydra* [64]. In the free-living samples from Jakarta Bay, *Alteromonadales* could only be detected in two samples. This could be due to the different environmental conditions in the samples from Jakarta Bay.

In order to show the influence of the environmental conditions on the microbiome of marine host species, an additional study was carried out using samples of *E. fuscoguttatus* taken from a mariculture facility at Thousand Islands and the surrounding reefs. Mariculture is a specialized branch of aquaculture involving the cultivation of marine organisms, such as fish, shellfish or oysters, in their natural environment. The differences in the bacterial community composition of *E. fuscoguttatus* cultivated in net cages and the specimens from the surrounding reef were a main focus of this study. In the first step, an amplicon-based sequencing approach was used to obtain an initial impression of the bacterial community composition. As with the results for *A. mate* and *E. sexfasciatus*, only minor differences were found between the mariculture and free-living conditions. While the distribution of predominant bacteria showed similar compositions, further investigations using a beta diversity analysis revealed that samples from the mariculture site formed a subcluster within the free-living samples. This shows the robustness of bacterial communities within these samples compared to the free-living samples. However, the alpha diversity measurements indicated a higher estimated species richness, as well as a higher mean number of species for the mariculture samples than for the free-living samples. These results confirmed that the number and distribution of phyla were similar. At the same time, however, it also revealed a more diverse bacterial community at deeper levels of taxonomy. In a second study using a whole metagenomic shotgun sequencing approach, this assumed community structure was examined. The results revealed similar bacterial community compositions at the phylum level of taxonomy. At a deeper level of taxonomy, obtained by a different sequencing approach, the bacterial community structure becomes more dissimilar between the two different environments. A comparison of the distributions of bacteria at the order level revealed a more stable distribution of microbial taxa within the mariculture samples compared to the free-living samples. While bacterial community composition is represented by comparatively even distributions in the mariculture samples, each free-living sample revealed a unique bacterial community that was different from all of the other samples. In particular, the influence of environmental conditions on the bacterial communities of host species can be seen in two samples that are highly dominated by *Vibrionales* (free3: 88.77%) and *Bacillales* (free5: 75.89%). As shown in the investigation by Xia *et al.* [47] on the influence of starvation on the intestinal microbiome of fish, different feeding conditions will lead to shifts in the distribution of bacterial communities. This supports the opinion that more controlled feeding as well as controlled environmental conditions of the

mariculture system will stabilize the distributions of bacterial communities inside host species. Each specimen of *E. fuscoguttatus* lives more or less under similar environmental conditions and with nearly identical feeding, which results in comparable microbiomes. In contrast, the feeding and environmental conditions of the free-living specimens are not comparable and therefore results in a unique microbiome for each specimen at order level.
Finally, in order to determine the influence of environmental conditions on microbiomes, the depth of taxonomical annotation seems to be a crucial factor. While the first study only showed minor differences at the phylum level, a deeper level of taxonomy revealed huge differences at the order level.

4.1.2 Core and shared gut microbiomes

The detection of core and shared microbiomes has become a common analysis method of amplicon-based sequencing data. A core microbiome is typically defined as a collection of members of the bacterial composition that are shared among all samples from a similar habitat [65]. The concept of a core gut microbiome has been explored in depth for mammalian hosts, e.g., Roeslers *et al.* [62] applied this concept successfully to a marine host species.
A common approach for obtaining a core microbiome from metagenomic sequencing data is to report the number of species that have been found across localities from a similar habitat or under different environmental conditions based on a presence or absence dataset [65]. This concept of a membership-based core, the proportion of shared and unique OTUs, is presented in this thesis (section 3.1).
The membership-based core microbiomes of the three host species *A. mate*, *E. sexfasciatus* and *E. fuscoguttatus*, as well as a shared microbiome across all three host species, were investigated. Instead of showing the differences between the environmental conditions of each single host species, the core microbiomes were used to show the differences and similarities between the microbiomes of the different host species. The core microbiomes of each host species were formed by six distinct bacteria at the class level, that appeared under different relations. The presence and distributions of the core members seem to be linked to the sampling origin of the host species. The core microbiomes of *A. mate* and *E. sexfasciatus* (both sampled at Jakarta Bay and Cilacap) are dominated by *Betaproteobacteria* (*E. sexfasciatus* : 49.16%, *A. mate*: 69.84%), and the core of *E. fuscoguttatus* is dominated by *Gammaproteobacteria* with over 90%. In addition, *Alphaproteobacteria* is present in the core microbiomes of *A.mate* and *E. sexfasciatus*, whereas in the core microbiome of *E. fuscoguttatus* it is completely missing. On the other hand, the taxonomic relationship also seems to have an influence on the core microbiomes. The two more related host species *E. sexfasciatus* and *E. fuscoguttatus* share *Clostridia* in differ-

ent proportions (*E. fuscoguttatus*: 1.35%, *E. sexfasciatus: 32.67%)*. In *A. mate*, however, *Clostridia* is missing and *Bacilli* is a member of the core microbiome. In turn, this bacteria is completely missing in the core microbiomes of the other two host species. As reported from the literature, the resulting shared microbiome of the three host species is formed by bacterial species that are characteristic for marine organisms and especially for teleost species [45, 47, 58, 57, 51] and belong to a substantial core microbiome [17].
Furthermore, membership-based core microbiomes of *P. monodon* were determined, but not included in the study (see section 3.3). Instead of the described core microbiomes of fish host species, the comparison of the different sampling locations focused on the core microbiomes of *P. monodon*. Due to the low taxonomic resolution at a deeper level of taxonomy for these samples, the obtained core microbiomes did not show a clear diversity. They are completely dominated by *Gammaproteobacteria* (Bali: 95.00%, Aquaculture: 94.85%), with the core of the Jakarta samples consisting only of *Gammaproteobacteria* (99.21%). In addition, the core microbiome of the Bali samples also had an abundance of *Alphaproteobacteria* (1.15%) and *Fusobacteria* (3.07%). With respect to the remaining core microbiome of the aquaculture samples, *Alphaproteobacteria* (1.65%), *Sphingobacteria* (1.52%) and *Flavobacteria* (1.23%) were also present.
In summary, the core and shared microbiomes of the fish host species confirmed the presence of bacterial members of a substantial fecal and gut core microbiome in teleost species, which were also detected in previous studies [45, 47, 58, 57, 51, 17]. In comparison to previous studies, however, the higher sequencing depth also resulted in an increased number of different OTUs belonging to the core members above at the class level of taxonomy. In contrast, the core microbiomes of the *P. monodon* samples did not show any considerable compositions. The poor taxonomic annotation rate combined with the low information content, especially for the Jakarta samples, was one of the main reasons for not presenting this analysis in the study of *P. monodon* (see section 3.3).

4.2 Elucidation of eukaryotic content in *Epinephelus fuscoguttatus*

The analysis of a complete metagenome shotgun sequencing dataset that examines all aspects from microbiome to functional components and eukaryotic content still remains an unresolved challenge [66]. While the majority of metagenomic studies only focus on the bacterial community composition and the related functional components of one specific body part or complete host organism, little is known about the eukaryotic components of the metagenome. This has led to the devel-

opment of a variety of different methods and annotation approaches for eukaryotic sequences. The majority of these methods are based on a specific sequencing step, i.e., amplicon-based profiling using 18S rRNA, developed especially for eukaryotes or the ITS regions for fungi. In order to access the eukaryotic content in the *E. fuscogutattus* samples, a different strategy was applied based on a whole metagenome shotgun sequencing dataset without specialized sequencing steps.

Each sample of *E. fuscoguttatus* is highly dominated by *Chordata* (99.95%), a taxa linked to the host itself as well as other eukaryotic orders of fish representing feed. An initial filtering step for host-specific DNA was not performed before sequencing, which resulted in a high level of contamination with residues of host cells (74.74%) in all samples. In order to solve this problem, the taxon *Perciformes*, which represents all possible residues of the host, was excluded for subsequent analysis. This resulted in the loss of all non-host-related taxonomic annotations in the order of *Perciformes*. In addition, nine additional orders of fish taxa could be identified in the samples, with most of them living mainly in salt or brackish waters around the island of Java. Two orders of fish, *Cichliformes* (3.56%) and *Salmoniformes* (0.17%), could be identified as false annotations due to the high sequence similarity with a wrong sequence hit in the database used for annotation. Neither of the fish orders are endemic to Indonesian waters. All members of *Cichlifomes* are exclusively freshwater fishes and *Salmoniformes* migrate between freshwater and saltwater during their life cycle. Further details on the processing and annotation of the whole metagenome sequencing dataset as well as the resulting difficulties and problems will be explained in more detail in section 4.5.

In order to access rare eukaryotic annotations, the phylum *Chordata* was filtered out. A total of 13 rare phyla could be detected and categorized in three groups: parasites, fungi and food residues. The group of parasites includes seven phyla, *Platyhelminthes*, *Arthropoda* and *Acanthocephala* are present in nearly all samples, independent of environmental conditions, and *Nemertoda*, *Nemertea*, *Annelida* and *Apicomplexa* could only be detected in individual samples. Each of these phyla, except for *Nemertea* and *Apicomplexa*, were also detected by parasitological investigations in the first study (see section 3.3). Since the parasitologically investigated fish and the fish that were used to extract fecal samples for whole metagenome shotgun sequencing are identical, some of the detected phyla could be confirmed by parasitological investigation. For instance, a high number of the parasite taxa detected in the intestines, stomach and stomach wall belong to *Platyhelminthes* and *Nemertoda*, which could be confirmed by the sequencing results. *Hirudinea*, a member of *Annelida*, could be detected in the fins and gill cavity and also in the sequencing data for some samples. While *Acanthocephala* was detected in nearly all samples of the sequencing dataset, is was absent in the parasitological investigations of *E. fuscoguttatus*. The remaining three phyla were only detected in the

sequencing dataset. *Apicomplexa*, a large phylum of parasitic protists, was not investigated by parasitological investigation. Due to the huge amount of organisms combined under the phyla *Arthropoda* and *Nemertea*, these could also be assigned to the food residues group. *Brachiopoda*, *Cnidaria*, *Echinodermata*, *Mollusca* and *Hemichordata* could also be assigned to this group. These findings are not surprising since *E. fuscoguttatus* represents a carnivorous fish species. In addition, two plant taxa, *Chlorophyta* and *Strepthophyta*, could be detected in the rare phyla. These two phyla, as well as several algae taxa, also belong to the group of food residues. *Ascomycota* and *Basidiomycota*, two fungal phyla, were also observed in the sequencing dataset. Since these two fungi (pooled in the subkingdom of *Dikarya*) [67], are known to be terrestrial, these annotations are more likely to be false classifications from the database. These phyla were therefore excluded from further analysis.
A comparison of these findings with similar studies is difficult because similar studies annotating eukaryotes from a whole metagenomic shotgun sequencing dataset had not existed until now. However, some studies have been performed with a target amplicon sequencing approach using the 18S rRNA for the detection of eukaryotic taxa. For instance, Hino *et al.* used the 18S rRNA amplicon to assess the biodiversity of parasites [68] and established a new method for detecting the parasitome of host organisms more easily than with traditional methods. In a related study, Hino *et. al* [69] detected parasitic phyla, that are identical with the detected parasitic phyla of *E. fuscoguttatus*. However, there were significant differences in the proportions due to the specialized technique. Also some studies have determined microbial eukaryotes (mainly in soil) by using an 18S rRNA amplicon-based approach [70, 71].

4.3 Exploration of functional content of the microbiome of *Epinephelus fuscoguttatus*

In addition to an exploration of eukaryotic content, the use of a whole metagenome shotgun sequencing approach also enables an investigation of the functional components in the samples of *E. fuscoguttatus*. To provide a mapping between the taxonomical annotations and the functional annotations the sequences of the predominant bacteria were used for the functional analysis. However, using a subset of data also means limited annotation results. The investigation of the functional components focuses on the biological process ontology of Gene Ontology [72], which describes a series of events accomplished by a number of organized assemblies of molecular functions. A review of the other two existing ontologies [72] resulted in low informational content for the cellular component and molecular function ontology [43].
All of the functional annotations obtained are completely linked to *Proteobacteria*,

regardless of taxonomic level. In the free-living samples of *E. fuscoguttatus*, an enrichment of DNA repair functions could be observed. In particular, the reparation of mismatches, as well as an undefined response to DNA damage, was significantly enriched. In comparison, samples derived from mariculture also showed enriched terms directed to DNA metabolic processing functions, but no enrichment for DNA repair functions due to insufficient resolution. The number of functional annotations obtained was a major problem during this analysis. Only a low number of sequences could be successfully annotated. While free-living samples reached an sufficient annotation level for the interpretation of contained functional terms, the mariculture samples were poorly annotated, thereby allowing only a general interpretation of the results. For instance, while the DNA repair-related functions of free-living samples could be assigned to three taxa at the order level (*Vibrionales*, *Pseudomondales* and *Enterobacteriales*), these functions could only be assigned to the phylum *Proteobacteria* in the mariculture samples. A more detailed explanation of the low resolution achieved is provided in section 4.5.
Nevertheless, the results indicate an increased contamination level in the free-living samples compared to the mariculture samples living under controlled environmental conditions [43]. In the free-living samples an enrichment for antibiotics response could be detected, that is completely linked to *Vibrionales*. The Pfam identifier [27] associated with these enriched terms is assigned to the membrane- and transport-related protein domains.
Finally, a low functional annotation rate, especially in the mariculture samples, allowed only a general overview of the functional components of the predominant taxa. An increased annotation rate could be achieved through a higher sequencing depth combined with the use of more comprehensive databases. At this stage, it is not possible to compare these results with other functional metagenomic studies. One reason is that most studies are not related to marine organisms. Furthermore, studies with a marine focus use other methods to obtain functional annotations, which are not directly comparable with the results presented here.

4.4 Interactions between the microbiome of host species and endoparasites

The combined parasitological results from our cooperation partner at the University of Rostock and the microbial community data revealed interactions between the number of endoparasites in a sample and contamination with specific bacterial genera. Three potentially pathogenic bacterial genera could be identified in the bacterial community composition, i.e., *Vibrio* sp., *Flavobacterium* and *Photobacterium* sp. Each of them are known causes of fish diseases and could also be dangerous

for human consumption. By using a Spearman's rank-order correlation test, a weak negative correlation between the recorded parasite number and the abundance of *Vibrio* sp. as well as *Photobacterium* sp. could be detected. This implies that an increased number of endoparasites result in a reduced abundance of these two bacterial genera. While the highest number of parasites was observed in Cilacap free-living samples of *E. sexfasciatus*, followed by *A. mate* and *E. sexfasciatus* from Jakarta Bay, the abundance of *Vibrio* sp. and *Photobacterium* sp. was quite low in these samples. In comparison, increased abundances for these genera were observed in samples of *E. fuscoguttatus* from net cages and, to a much lower degree, in free-living samples from the surrounding reef. Each of these samples show a reduced number of endoparasites. In contrast, contamination with *Flavobacterium* shows a weak positive correlation between the abundance of bacterial genus and the recorded number of endoparasites. However, the abundance of *Flavobacterium* was not high enough to obtain significant information because only one mariculture sample of *E. fuscoguttatus* showed contamination with this bacterial genus. In summary, the correlations support the assumption that there is a positive influence of metazoan parasite infection on fish health and the occurrences of potential pathogenic bacteria, as confirmed by our cooperation partner.

4.5 Comparison between targeted amplicon and whole metagenome shotgun sequencing strategies

For elucidation of several metagenomic questions, two distinct sequencing approaches were used to generate the metagenomic sequencing data. Each sequencing experiment was performed on an Illumina MiSeq sequencing machine. The MiSeq technology is able to sequence fragments of up to 350 bp in length, with a low error rate as well as a low rate of indel mutations. Due to its bench-top size, MiSeq technology is not able to obtain the same level of sequence coverage as the Illumina HiSeq technology.

Two experiments were perfomed using a targeted amplicon sequencing approach to investigate the bacterial communities in several fecal microbiomes and to detect core microbiomes of *Atule mate*, *E. sexfasciatus*, *E. fuscoguttatus* (see section 3.1) and *P. monodon* (see section 3.3). These techniques are based on the amplification and sequencing of variable regions of highly conserved bacterial genes in order to determine the taxonomic composition of a microbiome [66]. Over time, gene encoding for 16 rRNA has developed as a common marker gene for the analysis of prokaryotic taxonomic composition because it is highly conserved in all prokaryotic organisms. Since it is difficult to sequence the entire 16S rRNA gene with high-throughput

techniques, amplification and sequencing of one or more of the nine variable regions (V1-V9) of a gene is the most a common approach. Particular sets of primers exist to address these variable regions. However, recent studies have shown that these variable regions differ between species in a different way, which has an influence on the clustering of OTUs and the reported richness and evenness of communities. Therefore, the V4 region was used for this thesis because it has been standardized by the Earth Microbiome Project [73] and the available primers are capable of detecting most *Bacteria* and *Archaea* [66]. The popularity of target amplicon-sequencing studies has increased over the last years, and a variety of tools and pipelines have been developed to facilitate analysis of these data. Some of these tools have been used in this thesis to detect the bacterial community composition (usearch [74, 75], phyloseq [76]) and perform a statistical analysis (vegan [77]). In addition, there are also a variety of well-maintained databases.

In contrast to amplicon-based studies, the whole metagenomic shotgun sequencing approach is capable of investigating metagenomic samples in multiple ways. From the first experiment, twelve samples of *E. fuscoguttatus* were selected for an additional metagenomic experiment using a whole metagenomic shotgun sequencing on the Illumina MiSeq platform. During the performance of such an experiment, the whole DNA of a sample is sequenced and analyzed, which enables investigation of the bacterial community composition as well as the detection of genes and gene functions of bacterial communities. Due to a number of reasons, metagenomic analysis is much more challenging than an amplicon study [66]. While coverage of sequencing is not very important for amplicon-based approaches, it becomes one of the most important factors for whole metagenomic sequencing approaches since this technique tries to sequence whole genomes instead of individual phylogenetic marker with limited length. Firstly, high coverage is important for the assembly of short sequence reads to longer contigs because most assemblers rate their assembling results by coverage. Furthermore, coverage is also needed for better detection of the differences between the samples.

Until now, standardized techniques for analyzing metagenomic sequencing data were not available. This fact make it difficult to compare the results between different analyses [66]. However, a variety of tools and pipelines have been recently developed to analyze this kind of data, with most of them using different techniques or concepts. With respect to this thesis, many techniques were evaluated to find the best way of interpreting a whole metagenome sequencing dataset of a little-studied marine organism. The main goal of this evaluation was to increase the detection rates of the bacterial communities as well as the annotation count of genes in order to perform a functional analysis and compensate for the low sequence coverage of the MiSeq platform compared to the more common Illumina HiSeq technology.

For a comparison of the two different metagenomic sequencing approaches, rar-

efaction was examined for each experiment. Rarefaction, which was introduced by Sanders [78], represents a common technique for examining if the achieved species richness for a sequencing experiment is suitable for subsequent analysis or if higher sequencing depth will lead to more expected species richness. This is done by calculating the rarefaction curves, plotting the number of species as a function of the number of sequencing reads, and determining a saturation point. Both experiments using a target amplicon sequencing approach result in satisfactory rarefaction curves. While rarefaction of the first experiment is limited by the sample with the lowest sequencing reads (see SFig 2, page 33), the samples of the *P. monodon* experiment show a saturation point before the sample with the lowest sequence read count (see Supplementary Figure 3, page 61). Notwithstanding, the rarefaction curves of both experiments validate that the obtained sequence coverage of each sample is high enough to perform each possible analysis of the microbiome. However, the metagenomic sequencing experiment also shows good rarefaction curves after the samples were filtered by hierarchical clustering (see Supplementary Figure 1, page 44). Before filtering, one sample with low sequencing and a low annotation rate prevented a saturation point from being reached.

The processing of MiSeq whole metagenome shotgun sequencing data was challenging, because most of the tools and pipelines were developed for Roche 454 reads with complete different biases or optimized for the low sequence read length of the Illumina HiSeq technology. We therefore established our own pipeline [79] and optimized it for the sequencing data of this experiment. It mainly consists of a collection of standard tools for specialized tasks (e.g., quality control, connection of overlapping regions, etc.) and our own scripts linking the different pipeline steps and implementing some additional required functions. It was decided to improve the taxonomic annotation results in order to process the MiSeq sequencing data. One decision was to just combine overlapping reads with continuous regions [80] because each tested metagenomic assembling algorithm resulted in insufficiently assembled reads, thereby giving misleading information and reducing the annotation results to below a rate of 1% of the complete RAW reads. Instead of the metagenomic assembling, the sequence reads were combined, based on the overlapping regions in the forward and reverse strand, which enlength them to a theoretical maximum of 500 bp. This approach has shown to be suitable for each used annotation steps.

A common method for obtaining taxonomic annotations from metagenomic sequencing data is to extract DNA coding for the 16S rRNA gene [81] and perform an amplicon-based pipeline. Due to the low content of 16S rRNA in the samples, this step could not be performed. Instead, the metagenomic reads were taxonomically annotated with a blast search [22] against the non-redundant *nr* database [82]. Afterwards, taxonomic annotations were assigned to the blast results provided by our own least common ancestor algorithm using the ncbi taxonomy database [83] as a

reference. This approach provided reliable annotations for bacterial and eukaryotic organisms. For further processing, the annotation results were converted into a database structure, which was also provided by our own tool [84] and further analyzed with R [85]. The R-packages used for the analysis of the amplicon-based approaches (phyloseq [76], vegan [77]) could be successfully modified and extended by our own R-package (metaR [86]) in order to apply the annotations derived from a whole metagenome shotgun sequencing.

In order to gain an understanding of the functional components of the metagenomic data, parts of the RAMMCAP [87] pipeline were used. Due to the low number of usable sequences for the functional annotations, the application of a complete existing pipeline, e.g., InterproScan 5 [88], RAMMCAP [87] or mg-rast [89], was not possible. Instead, a number of scripts and algorithms were modified to fit underlying data and improve the annotation results. Finally, for a complete functional metagenomic analysis, including pathways, gene calling, etc., the coverage of sequencing was not high enough. Nevertheless, a determination of the most frequently used functions of the predominant bacteria was possible.

As shown in this section, carrying out an analysis of the amplicon-based sequencing data is a well-known method of investigating the bacterial composition of host species or samples taken directly from environment. The availability of standard primers, well-developed protocols and high-performance sequencing techniques as well as the many bioinformatical pipelines and taxonomic databases optimized for this type of data have made amplicon-based analysis a standard metagenomic technique. In contrast, a proper method for analyzing metagenomic data with a whole metagenomic sequencing approach for more in-depth investigation of bacterial compositions and the functional content of a metagenome still remains an unresolved challenge [66]. The computational analysis of metagenomic data in particular represents a significant challenge for the future. A number of tools have been developed in recent years to address important problems. But a standard approach for the processing of whole metagenomic shotgun sequencing data have not been found, leaving enough questions for continuous research.

4.6 Outlook

The results of the fecal microbiome investigations showed a connection between the environment and the stability of microbiomes. It could be demonstrated that the fecal microbiomes of host species living in cleaner water bodies or in a controlled environment, e.g., aquacultures, have similar distributions and predominant presences of bacteria compared to highly polluted environments. However, further metagenomic sequencing studies at a certain taxonomic annotation level are needed in order to support this conclusion. As seen in the first study, a taxonomic annotation at the phylum level is not sufficient enough to investigate the differences in distributions because a variety of bacteria typical for marine species are pooled in a handful of phyla. In general, both metagenomic approaches (i.e., the amplicon-based and whole metagenomic sequencing approach) are capable of obtaining these results.
The eukaryotic content in a marine sample was investigated using a whole metagenomic sequencing dataset for the first time. Although no standard bioinformatical pipeline exists for this kind of analysis, the obtained results give a general overview of the food of *E. fuscoguttatus* and could also confirm the findings of parasitological investigations. Subsequent analysis of this scientific problem will benefit greatly from the continuous improvements in the bioinformatical analysis of whole metagenome sequencing data.
Furthermore, the interactions between the microbiome of host species and viruses or endoparasites could be shown in this thesis. However, the statistical analysis of the interactions with endoparasites were applied on a small scale and need further analysis for verification. The viruses in the metagenomic samples of *P. monodon* were detected and confirmed only by qPCR. The interactions with the microbiome were obvious, but not investigated. For a better understanding of these interactions and how to influence the content of endoparasites or viruses in the host species, further studies are needed.
Apart from the amplicon-based approach, which is a reliable standard metagenomic technique, a proper way of analyzing whole metagenome sequencing data does not exist. However, as stated before, the ongoing improvement and development of bioinformatical pipelines will be a huge benefit for the field of metagenomics and address many of the problems that arose while working on this thesis.

References

[1] Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. The R Journal. 2013;5(1):144–161. Available from: http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

[2] Carpenter KE, Springer VG. The center of the center of marine shore fish biodiversity: the philippine islands. Environmental Biology of Fishes. 2005;72:467–480.

[3] Hutomo M, Moosa M. Indonesian marine and coastal biodiversity: Present status. Indian Journal of Marine Sciences. 2005;34:88–97.

[4] Pet-Soede C, MacHiels M, Stam M, Van Densen W. Trends in an indonesian coastal fishery based on catch and effort statistics and implications for the perception of the state of the stocks by fisheries officials. Fisheries Research. 1999;42:41–56.

[5] Food and Agriculture Organization of the United Nations. Fishery and Aquaculture Country Profiles: The Republic of Indonesia; 2015. Available from: http://www.fao.org/fishery/facp/IDN/en#CountrySector-SectorSocioEcoContribution.

[6] Eng CT, Paw JN, Guarin FY. The environmental impact of aquaculture and the effects of pollution on coastal aquaculture development in Southeast Asia. Marine Pollution Bulletin. 1989;20:335–343.

[7] Dsikowitzky L, et al. Anthropogenic organic contaminants in water, sediments and benthic organisms of the mangrove-fringed Segara Anakan Lagoon, Java, Indonesia. Marine Pollution Bulletin. 2011;62:851–862.

[8] Cooper TF, Gilmour JP, Fabricius KE. Bioindicators of changes in water quality on coral reefs: review and recommendations for monitoring programmes. Coral Reefs. 2009;28:589–606.

[9] Nordhaus I, Hadipudjana FA, Janssen R, Pamungkas J. Spatio-temporal variation of macrobenthic communities in the mangrove-fringed Seagara Anakan lagoon, Indonesia, affected by anthropogenic activities. Regional Environmental Change. 2009;9:291–313.

[10] Kennedy J, Flemer B, Jackson SA, Lejon DPH, Morrissey JP, O'Gara F, et al. Marine Metagenomics: New Tools for the Study and Exploitation of Marine Macrobial Metabolism. Marine Drugs. 2010;8:608–628.

[11] Handelsmann J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology. 1998;5:245–249.

[12] Chen K, Pachter L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Computational Biology. 2005;1:e24.

[13] Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012;486:215–221.

[14] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–214.

[15] The Micro B3 Project; 2016. Accessed: 10 Oct 2016. Available from: `https://www.microb3.eu/`.

[16] Pace NR, Stahl DA, Lane DJ, Olsen GJ. Analyzing natural microbial populations by rRNA sequences. ASM News. 1985;51:4–12.

[17] Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. Genome Research. 2009;19(7):1141–52.

[18] Tringe SG, Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. Current Opinion in Microbiology. 2008;11:442–446.

[19] Wooley JC, Godzwik A, Friedberg I. A Primer on Metagenomics. PLoS Computational Biology. 2010;6:e1000667.

[20] Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational meta'omics for microbial community studies. Molecular Systems Biology. 2013;9(666).

[21] Caporaso JG, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012;6:1621–4.

[22] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–410.

[23] NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Research. 2016;p. D7–D19.

[24] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000;p. 27–30.

[25] Kanehisa M, Sato Y, Kawashima M, Furumichi, M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. 2016;p. D457–D462.

[26] Kanehisa, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research. 2017;p. D353–D361.

[27] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. The Pfam protein families database:. Nucleic Acids Research. 2014;42:D222–D230.

[28] Tatusov R, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research. 2000;(1):33–6.

[29] Froese R, Pauly D. FishBase; 2015. Accessed: 24 Nov 2015. Available from: `www.fishbase.org`.

[30] Kaschner K, Kesner-Reyes K, Garilao C, Rius-Barile J, Rees T, Froese R. AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication. 2015;Available from: `www.aquamaps.org`.

[31] Clarke TA. Reproductive biology and egg abundance of the yellowtail scad or 'omaka, Atule mate (Carangidae), in Kane'ohe Bay, Hawai'i. Pacific Science. 1996;50(1):93–107.

[32] Encyclopedia of Life. Atule mate;. Accessed: 23 Nov 2015. Available from: `eol.org/pages/220547/details`.

[33] Russell B, Cabanban AS, Pollard D. Epinephelus sexfasciatus. The IUCN Red List of Threatened Species. 2008;e.T132780A3449636.

[34] Heemstra PC, Randall JE. Groupers of the world (family Serranidae, subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date. In: FAO Species Catalogue. vol. 16. FAO; 1993. p. 382 p.

[35] Pogonoski JJ, Pollard DA, Paxton JR. Conservation Overview and Action Plan for Australian Threatened and Potentially Threatened Marine and Estuarine Fishes.; 2002.

[36] Sugama K, Rimmer MA, Isma S, Koesharyani I, Suwirya K, Giri NA, et al. Hatchery Management of Tiger Grouper (Epinephelus fuscoguttatus): A Best-Practice Manual. Australian Centre for International Agricultural Research; 2012.

[37] Cornish A. Epinephelus fuscoguttatus. The IUCN Red List of Threatened Species. 2004;e.T44673A10934320.

[38] Encyclopedia of Life. Penaeus monodon;. Accessed: 24 Nov 2015. Available from: eol.org/pages/220547/details.

[39] Food and Agriculture Organization of the United Nations. Penaeus monodon;. Accessed: 24 Nov 2015. Available from: http://www.fao.org/fishery/culturedspecies/Penaeus_monodon/en.

[40] van der Meij SET, Moolenbeek RG, Hoeksema BW. Decline of the Jakarta Bay molluscan fauna linked to human impact. Marine Pollution Bulletin. 2009;59(4-7):101–107.

[41] van der Meij SET, Suharsono, Hoeksema BW. Long-term changes in coral assemblages under natural and anthropogenic stress in Jakarta Bay (1920–2005). Marine Pollution Bulletin. 2010;60(9):1442–1454.

[42] Hennersdorf P, Kleinertz S, Theisen S, Abdul-Aziz M, Mrotzek G, Palm H, et al. Microbial Diversity and Parasitic Load in Tropical Fish under Different Environmental Conditions. PLoS ONE. 2016;11:e0151594.

[43] Hennersdorf P, Mrotzek G, Abdul-Aziz M, Saluz H. Metagenomic analysis between free-living and cultured Epinephelus fuscoguttatus under different environmental conditions in Indonesian waters. Marine Pollution Bulletin. 2016;110:726–734.

[44] Oetama V, Hennersdorf P, Abdul-Aziz M, Mrotzek G, Haryanti H, Saluz H. Microbiome analysis and detection of pathogenic bacteria of Penaeus monodon from Jakarta Bay and Bali. Marine Pollution Bulletin. 2016;110:718–725.

[45] Llewellyn M, Boutin S, Hoseinifar S, Derome N. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front Microbiol. 2014;5.

[46] Kennedy J, Marchesi J, Dobson A. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. Microbial Cell Factories. 2008;7.

[47] Xia J, et al. The intestinal microbiome of fish under starvation. BMC Genomics. 2014;15(1):266.

[48] Xing M, Hou Z, Yuan J, Liu Y, Qu Y, Liu B. Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (Scophothalmus maximus). FEMS Microbiol Ecol. 2013;86:432–443.

[49] Liu H, Wang L, Liu M, Wang B, Jiang K, Ma S, et al. The intestinal microbial diversity in Chinese shrimp (Fenneropenaeus chinensis) as determined by PCR-DGGE and clone library analyses. Aquaculture. 2011;317:32–6.

[50] Zhang M, Sun Y, Chen K, Yu N, Zhou Z, L C, et al. Characterization of the intestinal microbiota in Pacific white shrimp, Litopenaeus vannamei, fed diets with different lipid sources. Aquaculture. 2014;434:449–55.

[51] King GM, Judd C, Kuske CR, Smith C. Analysis of Stomach and Gut Microbiomes of the Eastern Oyster (Crassostrea virginica) from Coastal Louisiana, USA. PLoS One. 2012;7:1–11.

[52] Guarner F, Malagelada JR. Gut flora in health and disease. Lancet. 2003;360:512–9.

[53] Balcazar JL, Blas ID, Ruiz-Zarzuela I, Cunningham D, Vendrell D, Muzquiz JL. The role of probiotics in aquaculture. Vet Microbiol. 2006;114:173–86.

[54] Chaiyapechara S, Rungrassamee W, Suriyachay I, Kuncharin Y, Klanchui A, N K, et al. Bacterial Community Associated with the Intestinal Tract of P. monodon in Commercial Farms. Microbial Ecology. 2012;63:938–53.

[55] Rungrassamee W, Klanchui A, Maibunkaew S, Chaiyapechara S, Jiravanichpaisal P, Karoonuthaisiri N. Characterization of intestinal bacteria in wild and domesticated adult black tiger shrimp (Penaeus monodon). PLoS One. 2014;9.

[56] Vaseeharan B, Sundararaj S, Murugan T, Chen JC. Photobacterium damselae ssp. damselae associated with diseased black tiger shrimp Penaeus monodon Fabricius in India. Lett Appl Microbiol. 2007;45:82–6.

[57] Sullam KE, et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Molecular Ecology. 2012;21:3363–3378.

[58] Sevellec M, Pavey SA, Boutin S, Filteau M, Derome N, Bernatchez L. Microbiome investigation in the ecological speciation context of lake whitefish (Coregonus clupeaformis) using next-generation sequencing. Journal of Evolutionary Biology. 2014;p. 1029–46.

[59] BPS Statistics of Jakarta Province. Jakarta in Figures 2014; 2014.

[60] Neubert K, Yuliantoa I, Theisen S, Kleinertz S, Palm H. Parasite fauna of Epinephelus coioides (Hamilton, 1822) (Epinephelidae) as environmental indicator under heavily polluted conditions in Jakarta Bay, Indonesia. Marine Pollution Bulletin. 2016;110:747–756.

[61] Partida-Arangure BO, Luna-Gonzalez A, Fierro-Coronado JA, Flores-Miranda C, Gonzalez-Ocampo H. Effect of inulin and probiotic bacteria on growth, survival, immune response, and prevalence of white spot syndrome virus (WSSV) in Litopenaeus vannamei cultured under laboratory conditions. African Journal of Biotechnology. 2013;12:3366–75.

[62] Roeselers G, et al. Evidence for a core gut microbiota in the zebrafish. The ISME journal. 2011;5:1595–608.

[63] Cox CR, Gilmore MS. Native microbial colonization of Drosophila melanogaster and its use as a model of Enterococcus faecalis pathogenesis. Infect Immun. 2007;75:1565–76.

[64] Fraune S, Bosch TCG. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. Proc Natl Acad Sci U S A. 2007;104:13146–51.

[65] Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. Environmental Microbiology. 2012;14:4–12.

[66] Di Bella J, Bao Y, Burton J, Reid G. High throughput sequencing methods and analysis for microbiome research. Journal of Microbiological Methods. 2013;95:401–414.

[67] Hibbet DS, et al. A higher-level phylogenetic classification of the Fungi. Mycol Res. 2007;111:509–47.

[68] Hino A, Maruyama H, Kikuchi T. A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. Parasitology International. 2016;65:572–575.

[69] Tanaka R, Hino A, Tsai I, Palomares-Rius J, Yoshida A, et al. Assessment of Helminth Biodiversity in Wild Rats Using 18S rDNA Based Metagenomics. PLoS ONE. 2014;9:e110769.

[70] Aguilar M, Richardson E, Tan B, Walker G, Dunfield P, Bass D, et al. Next-Generation Sequencing Assessment of Eukaryotic Diversity in Oil Sands Tailings Ponds Sediments and Surface Water. Journal of Eukaryotic Microbiology. 2016;63:732–743.

[71] Jacquiod S, Stenbæk J, Santos S, Winding A, Sørensen S, Priemé A. Metagenomes provide valuable comparative information on soil microeukaryotes. Research in Microbiology. 2016;167:436–450.

[72] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

[73] Gilbert J, Jansson J, Knight R. The Earth Microbiome project: successes and aspirations. BMC Biology. 2014;12:69.

[74] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194–2200.

[75] Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods. 2013;10:996–8.

[76] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS ONE. 2013;8:e61217.14.

[77] Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010;11:R106.

[78] Sanders H. Marine Benthic Diversity: A Comparative Study. The American Naturalist. 1968;925:243–282.

[79] Hennersdorf P. meta-pipeline git repository. GitHub; 2014. Available from: `https://github.com/psikon/meta-pipeline`.

[80] Magoc T, Salzberg S. FLASH: Fast length adjustment of short reads to improve genome assemblies. 2011;27:21:2957–63.

[81] Lee J, Yi H, Chun J. rRNASelector: a computer program for selecting ribosomal RNA encoding sequences from metagenomic and metatranscriptomic shotgun libraries. The Journal of Microbiology. 2011;(4):689–91.

[82] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research. 2016;44:D7–D19.

[83] Federhen S. The NCBI Taxonomy database. Nucleic Acids Research. 2012;40:136–143.

[84] Schoefl G, Hennersdorf P. bigBlastParser git repository. GitHub; 2014. Available from: `https://github.com/psikon/bigBlastParser`.

[85] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna; 2014. Available from: `http://www.r-project.org/`.

[86] Hennersdorf P. metaR github repository. GitHub; 2014. Available from: `https://github.com/psikon/metaR`.

[87] Li W. Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinformatics. 2009;(10):359.

[88] Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;.

[89] Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;(9):386.

List of Figures

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbst angefertigt habe, dass ich keine Textabschnitte und Abbildungen ohne Kennzeichnung übernommen habe und alle von mir genutzten Quellen, Hilfsmittel und persönlichen Mitteilungen angegeben sind.
Beim Verfassen dieser Dissertation habe ich weder die Hilfe eines Promotionsberater in Anspruch genommen, noch haben Dritte von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeit erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Bei der Auswahl und Auswertung des Materials haben mich nur die genannten Koautoren und in der Danksagung erwähnten Personen unterstützt.
Ich versichere außerdem, dass die vorliegende Dissertation oder wesentliche Teile davon weder als Prüfungsarbeit für eine staatliche oder andere wissentschaftliche Prüfung noch als Dissertation Gegenstand eines anderen Promotionsverfahren an einer anderen Hochschule ist.
Die Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität ist mir bekannt.

Ort, Datum

Philipp Hennersdorf

Danksagung

Abschließend möchte ich mich bei verschiedenen Personen bedanken, ohne deren Unterstützung die vorliegende Promotionsarbeit nicht möglich gewesen wäre.
Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Hans Peter Saluz, in dessen Abteilung, für Zell- und Molekularbiologie, ich meine Dissertation angefertigt habe. Während der gesamten Zeit stand er mir immer mit zahlreichen Ideen, Tipps und Anregungen zur Seite und hatte stets ein offenes Ohr für meine Probleme. Auch wenn diese aussichtlos erschienen, verstand er es mich zu motivieren, nach einer Lösung zu suchen oder einen geeigneten Ansprechpartner zu finden.
Grit Mrotzek danke ich für ihre zahlreichen Ratschläge. Egal ob fachlicher oder persönlicher Natur, wusste sie oft Rat und hörte sich meine Probleme geduldig an.
Ebenso möchte ich Dr. Frank Hänel für die hervorragende Zusammenarbeit danken.
Auch meinen Koautoren Muslihudeen A. Abdul-Aziz und Vincensius Surya Putera Oetama gilt mein Dank. Ohne ihre Arbeiten an den Veröffentlichungen wäre diese Arbeit nicht möglich gewesen.
Besonderer Dank gilt auch meiner Frau Sophie Hennersdorf. Während der gesamten Arbeit stand sie immer an meiner Seite und hielt mir den Rücken frei. Sie wusste immer wie sie mich motivieren konnte, holte mich aus den tiefsten Löchern der Verzweiflung und opferte viel, um mir diese Arbeit zu ermöglichen. Ohne ihre Unterstützung und Ratschläge wäre diese Arbeit niemals zu Stande gekommen. Danke dafür!
Mein letzter Dank gilt meiner Familie und vor allem meinen Eltern. Ihnen möchte ich für ihre aufopferungsvolle Unterstützung danken, ohne die mir dieser Bildungsweg nicht möglich gewesen wäre.

Appendix

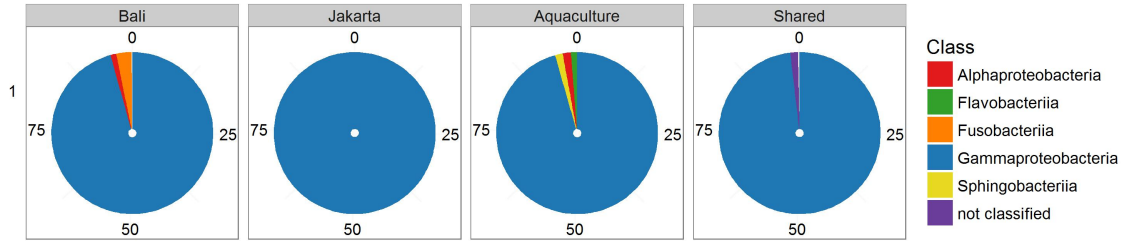


Figure 7: Core microbiomes for each sampling location of *P. monodon* and a shared microbiome at the class level. The membership-based core microbiome is constructed by counting the number of OTUs present in each *P. monodon* sample for a certain sampling condition. The shared microbiome results from the combination of all three core microbiomes, which represent all OTUs found in the samples, independent of sampling location.