Butterflyfish gill mucus metabolome reflects diet preferences and gill parasite intensities.

Miriam Reverter¹,²,³, Pierre Sasal²,³, Bernard Banaigs²,³, David Lecchini²,³, Nathalie Tapissier-Bontemps²,³

1 Institut für Chemie und Biologie des Meeres, Carl von Ossietzky Universität Oldenburg, Wilhelmshaven, Germany
2 PSL Research University, USR EPHE-UPVD-CNRS 3278, CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860, Perpignan Cedex, France
3 Labex Corail, BP1013 Papetoai, 98729 Moorea, French Polynesia

Author Contribution(s): All authors conceived the study. MR and PS captured the butterflyfish. MR, NT-B and BB worked together to set up the fish mucus extraction and LC-MS non-targeted profiling analysis. MR and NT-B analysed the data. All authors commented on the final version of this case study.

Abstract:
Fish mucus is the main surface of exchange between fish and the environment and plays numerous biological and ecological roles such as protection and chemical mediation. In this study, we investigated the influence of phylogeny, reef habitat, diet and gill parasitism levels on the gill mucus metabolome of butterflyfishes in Moorea (French Polynesia). Gill mucus was extracted using a biphasic extraction yielding two fractions (polar and apolar), which were analysed by liquid-chromatography tandem mass spectrometry (LC-ESI-MS). Fish diet explained the highest metabolomic variability of both fractions, with parasitism being the second most important driver in the polar fraction. Ultra-high performance liquid chromatography tandem mass spectrometry (UHPLC-HRMS/MS) analyses allowed to putatively identify several glycerophosphocholines, ceramides and fatty acids related to fish diet and the identification of several peptides specific to non-parasitised fish.

Scope
This work was a field study that investigated the fish gill mucus metabolome of eight butterflyfish species at the Moorea Island (French Polynesia). Butterflyfish species were selected due to their contrasting diet preferences and levels of parasitism and they were sampled at three geographic sites and three different reef habitats (fringing reef, barrier reef and outer slope). This study was part of the Ph.D. project of M. Reverter at the CRIOBE research laboratory supported by a labex Corail doctoral grant.

Aims and objectives:
• Explore the gill mucus metabolome of eight butterflyfish species in Moorea (French Polynesia).
• Investigate the influence of several fish characteristics (habitat, diet, phylogenetic clade and parasitism levels) on the fish mucus metabolome.
• Identify biomarkers for the most important traits in the metabolome variability.

Introduction
All biological processes and systems are regulated by sets of chemical reactions (Metzler 2003). For example, the physiological state of an organism is determined by metabolic reactions.
Chemicals not only regulate functions within organisms, but also shape ecosystem structure by affecting intra- and inter-specific communication (Bornacin et al. 2017). In the marine environment, chemical cues affect species’ foraging behaviour, selection of mates and habitats, competitive interactions, commensal associations and pathogenic invasions (Hay 2009). Therefore, the study of metabolites, the end-products of metabolic reactions, and their variability, does not only provide information on the precise physiological state of an organism, but also useful information on its ecology (Jones 2013).

In fish, the external mucosal layers are the main surface of exchange between them and the environment and play important roles in defence and communication (Reverter et al. 2018). Fish mucus contains a wide-array of compounds, including immune-related molecules, antimicrobial peptides and smaller metabolites such as toxins (Masso-Silva and Diamond 2014; Salinas 2015; Reverter et al. 2018). The use of high-throughput techniques such as proteomics has allowed the identification of many immune-related proteins previously unreported in fish mucus (Rajan et al. 2011; Cordero et al. 2015; Brinchmann 2016). Fish mucus metabolomics studies remain scarce, despite their potential use in varied applications such as non-invasive fish health monitoring (Ivanova et al. 2018) or in biomarker discovery in a wide range of topics (e.g. disease biomarkers or indicators of environmental changes or pollution) (Ekman et al. 2015; Mosley et al. 2018).

Chaetodontids (butterflyfish) are a diverse and emblematic family of coral reef fish, with some species being amongst the most broadly distributed coral reef fish species (diBattista et al. 2015). Most butterflyfish species are associated with live coral cover, with some being particularly vulnerable to coral loss or degradation (Pratchett et al. 2006). Butterflyfish phylogeny, ecology and behaviour are well studied, with their diet varying from consuming only few hard coral taxa to consuming a large variety of invertebrates and algae (Harmelin-Vivien and Bouchon 1983; Harmelin-Vivien 1989; Prachtett et al. 2005; Fessler and Westnec 2007). Butterflyfish are naturally parasitized by monogeneans (ectoparasitic flatworms) belonging to the family Dactylogyridae (Plaisance and Kritsky 2004). In a recent study, we found that phylogenetically close species have very different levels of monogenean parasites. Indeed, the Oval butterflyfish (Chaetodon lunulatus) was the only species out of 34 studied species in the Indo-Pacific to never be parasitized by gill monogeneans (Reverter et al. 2016). This study aimed to investigate the butterflyfish gill mucus metabolome and to explore how this is related to the fish traits, including the presence or absence of parasites, and their environment. We first performed an untargeted metabolomics study of the gill mucus of eight butterflyfish species, from which biomarkers (determined by Variable Important in Projection, VIP) were obtained for the fish traits that explained the highest metabolic variability. We then used high-resolution mass data and fragmentation patterns to identify some of the VIP highlighted metabolites.

**Approach**

**a. Biological sample**

Eight butterflyfish species (genus Chaetodon, family Chaetodontidae, order Perciformes) were sampled by spearfishing around the island of Moorea (French Polynesia) to collect their gill mucus (Table 1). Standard length of all individuals was measured. Fish species were selected to have individuals from different phylogenetic clades and displaying contrasting diets and parasitism intensities, and according to their field abundance (Table 1). All animal handling and sampling protocols were pre-approved by animal experimentation experts from our
b. Environment sampled
Fish were sampled at three geographic locations (N: Ta’ahiamanu, E: Afareiatu and W: Haapiti) in Moorea between April and July 2015 (Figure 1, Table 1). At one of the locations (Ta’ahiamanu), butterflyfish species were sampled in different reef habitats gradually further away from the coastline: fringing reef, barrier reef and outer slope reef. The fringing reef at Ta’ahiamanu was composed of a gentle slope with coral colonies (mostly *Porites rus*) and a sandy floor that reached around 5 m deep. The sampling at the barrier reef was performed on the inner part of the barrier, which was characterised by high coral cover (mainly *Acropora* spp. and *Pocillopora* spp.) at depths varying between 2 and 7 m. Finally, the outer slope reef sampling was performed by scuba-diving at depths between 12-18 m, and the site presented high coral cover and biodiversity. Sampling in Haapiti and Afareiatu was done at the fringing reef, which in Haapiti was constituted by sparse coral colonies (mainly *Acropora* spp.) in a sandy, shallow lagoon with relatively constant depth (2 – 4 m), whereas in Afareiatu was composed of a steep fringing slope with high coral cover and diversity that reached around 10 m deep at less than 10 m from the coast.

c. Sampling protocol
Fish were spearfished and killed immediately by brain spiking, put in individual plastic bags and brought immediately to the laboratory on ice. Gills were aseptically dissected (< 1 hour after spear-fishing them) and gill mucus was carefully scraped with a spatula into sterile tubes, which were then frozen at -80°C, freeze-dried and kept at -20°C until extraction (Figure 2).

d. Analytical protocol
Ten milligrams of freeze-dried mucus were extracted using a two-step biphasic extraction (Reverter et al. 2017). Briefly, we suspended the freeze-dried mucus in 1.5 mL of H₂O/MeOH (2:0.5), we vortexed and added 1.5 mL of MeOH/CH₂Cl₂ (2.5:2). The mixture was extracted for 15 min in an ultrasonic bath and then centrifuged to yield two fractions, the polar and non-polar fractions, which were pipetted carefully into different tubes. Both fractions were dried, resuspended in 300 µL MeOH and filtered through a 20-µm polytetrafluoroethylene filter prior to LC-MS analysis. QC samples were prepared for both fractions by pooling small aliquots (100 µL) from the respective extracts.

In order to avoid systematic errors associated with instrumental drift, we randomised our analytical injection sequence using a Latin square for each of the sample sequences (polar and non-polar). QC samples were analysed at the beginning, end and equidistantly (every 10 samples) throughout the sequence. Methanol blank samples were analysed just before each QC sample to detect column contamination. This resulted in two sequence batches, polar and apolar, of 132 analyses each.

LC-MS analyses for the untargeted metabolomics study were performed using a LC-DAD-ESI-MS system from ThermoScientific (MA, USA) equipped with an Accela PDA detector and a LCQ Fleet 2300 mass spectrometer with an electrospray ionisation source as specified in Reverter et al. (2017). Briefly, a Kinetex C6-Phenyl analytical column (100 x 2.10 mm, 2.6 µm particle size; Phenomenex, CA, USA) was used to achieve chromatographic separation using H₂O (A) and acetonitrile (B) as carriers, both containing 0.1% v/v formic acid. For the polar fraction, a linear biphasic gradient from 2 to 100% of B in A was applied over a period of 30
For the non-polar fraction, a linear gradient was conducted from 30 to 70% of B in 5 min, then from 70 to 80% of B in 20 min and from 80 to 100% of B in 5 min.

Raw LC-MS datafiles were converted to NetCDF files with Excalibur software and processed using the XCMS package for R (R version 3.3.1, XCMS version 1.50.0, Smith et al. 2016) as specified in Reverter et al. (2017). XCMS analysis provided a matrix containing the retention time, m/z value and integrated peak area of the identified features. Data was normalised prior to statistical analysis by log transformation (R program) to reduce heteroscedasticity and correct for skewed data.

Adonis function (vegan package from R v2.3-3, Oksanen et al. 2016), which is equivalent to permutational analysis of variance using distance matrices, was used to identify the factors that played a significant role in the metabolic variability observed. We used the following co-variables in the Adonis test: diet, intensity of parasitism, phylogenetic clade, geography of the sampling site, reef habitat and fish species. Principal component analysis (PCA) was used to visualise the gill metabolome variation on significant fish traits (Reverter et al. 2017). Multiple regression trees (MRT, package mvpart from R v1.6-2, Therneau and Atkinson 2014) were used to rank the different factors (diet, intensity of parasitism, phylogeny, habitat and site) according to their contribution to the metabolic variability. MRT produces a tree where at each split data is partitioned into two mutually exclusive groups, each of which is as homogeneous as possible, displaying this way the main contributor factors to variability (De’ath 2002; Vignon and Sasal 2010; Reverter et al. 2017). Only fish traits with a contribution higher than 0.1% were plotted for each of the fractions. Main contributor factors for each of the fractions were selected and a partial least squares discriminant analysis (PLS-DA) was performed to obtain information on the metabolites driving the differences (VIP) in these factors (diet for the non-polar fraction and parasitism in the polar fraction) as specified in Reverter et al. (2017). PLS-DA model accuracy was calculated using a two-fold cross-validation (2CV) and the resulting number of misclassifications (NMC) was compared to its permuted null distribution (999 permutations) to test for model significance (P value < 0.05). VIP features higher than 1.5 for diet and 1 for parasitism were retained, and a Kruskal-Wallis test (after checking for normality and heteroscedasticity of data) was used to identify significant differences in abundance of metabolites between the factor categories (dietary groups or parasitized vs. non-parasitized).

Ultra-high-performance liquid chromatography tandem mass spectrometry (UHPLC-HRMS/MS) analyses were performed on selected samples (polar and non-polar fractions of representative groups such as different diets or different parasitism levels) to obtain a higher mass accuracy of the discriminant compounds obtained by the metabolomic analyses as specified in Reverter et al. (2017). Analyses were performed on a Dionex UltiMate 3000 UHPLC system (ThermoScientific) equipped with a Corona Veo detector and interfaced with a hybrid quadrupole Orbitrap mass spectrometer (Q Exactive Focus, ThermoScientific). Putative identification of metabolites was performed by searching the following libraries for known molecules: the Global Natural Products Social Molecular Network (GNPS) (https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp), Lipid Maps Structure Database (LMSD) (http://www.lipidmaps.org/tools/ms/LMSD_search_mass_options.php) and METLIN (http://metlin.scripps.edu/). Mass tolerance for the m/z value of precursor ions was set at ±0.005 Da (5 ppm). Molecular formulas were assigned using the accurate mass (or m/z) of the analytes and their associated isotopologues, by comparison with isotopic distribution profile reported on the LMSD. MS/MS fragmentation patterns were further used to identify the class or sub-class of the metabolites detected in the non-polar fractions. This was done by matching the MS/MS spectra with those recorded in the publicly available metabolite-curated database METLIN and
Main findings

Despite spatial metabolome variability has been reported before in organisms such as marine invertebrates (He et al. 2014) in this study geographic location and habitat did not significantly affect the metabolite profiles of fish gill mucus (Adonis, $P_{\text{polar-location}} = 0.07$, $P_{\text{non-polar-location}} = 0.26$, $P_{\text{polar-habitat}} = 0.06$, $P_{\text{non-polar-habitat}} = 0.25$). This could be related to the low environmental variability of waters around Moorea island (Lamy et al. 2015), compared to the variability between other fish traits. Butterflyfish diet was the main factor influencing the gill mucus metabolome (Adonis $P_{\text{polar}} = 0.001$, $P_{\text{non-polar}} =0.001$), with differences between obligate corallivores and facultative corallivores/omnivorous fish accounting for 7% of the metabolic variability of the non-polar fractions (Figure 3). Butterflyfish feed on a wide variety of invertebrates (Prachett et al. 2005) and thus different diet preferences (e.g. scleractinian corals, algae or polychaetes) can influence the metabolome by differently affecting the fish physiology or by transferring some of the prey metabolites to the external gill mucus. Parasitism and phylogeny were also significantly linked to the metabolic variability in both fractions (Adonis, $P_{\text{polar-parasitism}} = 0.001$, $P_{\text{non-polar-parasitism}} =0.001$, $P_{\text{polar-phylogeny}} = 0.001$, $P_{\text{non-polar-phylogeny}} =0.001$). Fish phylogeny was the second most important factor in the non-polar fraction, explaining the differences between facultative corallivores/omnivorous fish (3%) (Figure 3). This can be explained by the fact metabolic profiles are highly species-specific (e.g. Jaramillo et al. 2018, Reverter et al. 2018), especially specialized metabolite profiles, which are predominantly non-polar compounds. In the polar fraction, the difference between parasitized and non-parasitized fish was the second most important factor and explained 7.4% of the metabolic variability (Figure 3). Different metabolic profiles in differently parasitized fish can have several explanations. Firstly, parasites can affect the metabolic profiles by either modulating expression of mucus molecules for their benefit or by eliciting fish responses (Low et al. 2017). Secondly, some fish mucus molecules such as anti-parasitic molecules, or recognition molecules that can enhance parasitism, could be directly responsible for respectively absence or presence of parasites (Kallert et al. 2011).

After visualisation of the metabolic variability using PCA analysis (Reverter et al. 2017) and identifying the most significant factors (adonis and MRT), we then applied PLS-DA models to obtain the metabolites that drove the differences between the fish with different dietary groups. PLS-DA models of both factions accurately predicted metabolite differences among fish with different diets (NMC$_{\text{polar}} = 0.08$, P value$_{\text{polar}} = 0.001$; NMC$_{\text{non-polar}} = 0.04$, P value$_{\text{non-polar}} = 0.001$). We obtained 44 features in the non-polar fraction and 39 in the polar fraction with VIP scores higher than 1.5, all of which presented significantly higher expressions in one of the fish dietary groups (Kruskal-Wallis $p$ value$< 0.05$). Metabolites responsible for the differences between dietary groups in the non-polar fractions included glycerophosphocholines, ceramides and free fatty acids (Table 2). However, due to low concentrations of VIP in the polar fractions, no fragmentation data could be obtained for these metabolites, and therefore, no putative identifications were possible.

Since parasitism was the second main factor influencing the gill mucus metabolome in the polar fraction and the third in the non-polar fraction, we also performed PLS-DA models to investigate what metabolites drive the differences between fish with different parasite levels. Only the PLS-DA model for the polar fraction accurately predicted differences between the metabolomics profiles of parasitized and non-parasitized fish (NMC$_{\text{polar}} = 0.043$, P value$_{\text{polar}} =$...
This was probably due to the lower importance of parasitism in the non-polar fraction (2.2%) compared with the polar fraction (7.4%) (Figure 3). After cleaning the matrix, 69 features with a VIP score higher than 1 and significantly over-expressed (Kruskal-Wallis, P < 0.001) in non-parasitised fish were retained. Seven out of 69 VIPs were identified as peptides due to their characteristic multicharged ions and typical fragmentations observed from the high-resolution mass spectrometry data, which included some of the most discriminant compounds (Table 3).

### What went right

#### i) Large collection of wild biological samples

Organising marine field metabolomics studies is often challenging, especially in the tropics, due to the logistics and limitations in accessing different sampling sites and ensuring sample preservation. In this study, we were able to gather a large collection of samples from butterflyfishes (110 individuals) in different habitats at the island of Moorea, which was only possible due the availability of both terrestrial and aquatic means of transport and the availability of in situ -80°C freezer and freeze-drier. Previous butterflyfish sampling episodes were performed to develop the extraction and analytic method and to study the parasitism intensities and prevalence of different species (Reverter et al. 2016, 2017). Such previous knowledge was highly helpful in the selection of the fish species for the current study, in terms of not only species abundance and frequency at different sites, but also on their gill mucus abundance. For example, some abundant species such as *Chaetodon citrinellus* were disregarded for this study due to the small amount of mucus obtained. The interdisciplinarity of the team, consisting of both field ecologists and chemists was also crucial in the design of the study, which required good knowledge of the fish species and their ecology as well as analytical method development.

#### ii) Metabolomic pipeline

In this study we used a biphasic extraction, in order to capture as much of the small molecule (< 3,000 Da) variation as possible. With this extraction method, we were able to extract both lipophilic compounds such as ceramides and fatty acids and hydrophilic compounds such as peptides, showing the wide variety of polar and non-polar metabolites that were extracted simultaneously in a highly complex matrix, composed mainly of large macromolecules (e.g. mucins). This work proved the complementarity of LC-MS techniques in identifying biomarkers of ecological relevance from low freeze-dried mucus quantities (10 mg). Low-resolution LC-MS (HPLC-ESI-MS) was used successfully in the metabolite fingerprinting and allowed highlighting of the main features behind the observed metabolomics variability. Then, high-resolution mass and fragmentation data (UHPLC-ESI-HRMS/MS) were used to retrieve more information on the molecular formula and fragmentation patterns, which allowed us to putatively identify the families of most of the important VIP. The metabolomics pipeline used allowed the identification of features with different levels of expression between the studied groups (e.g. absence/presence of monogenean parasites) that were not amongst the major compounds, and therefore would probably have not been characterised using other approaches such as bio-guided fragmentation. We could therefore recommend the use of LC-MS metabolomics in biomarker identification of fish mucus and the use of biphasic extraction as a comprehensive extraction method.

### What went wrong

#### i) Samples limitations

Despite sampling was performed just after the rainy season, in order to collect the samples in
the shortest period possible, fish were sampled at different times of the day and at different dates, which could have inevitably induced some variability. Furthermore, due to the limited amount of mucus found in each individual fish, all gill arches were used for mucus sampling, and reference parasite intensities were considered from a previously published study. Similarly, coarse diet categories were established from previous articles (Table 1). Ideally each of the parameters should have been measured in each of the individual samples, in order to provide more specific categories and maybe better correlation between the metabolome and the studied factors. However, this is not biologically realistic on most butterflyfish species, especially the ones with low parasitism intensities.

ii) Limited feature annotation

Despite the fact we were able to characterise the families of some of the most important VIP compounds, many features remained unknown. High number of unknowns is often a drawback of environmental metabolomics studies, where non-model species are investigated. In this study, a low-resolution spectrometer was used for the untargeted metabolomics analysis, using only HRMS/MS experiments to obtain accurate masses and fragmentation patterns of VIP compounds for database matching. If we would have performed HRMS/MS analysis of all samples, we could have used innovative approaches such as MS/MS molecular networking (Wang et al. 2016), which combined with MS/MS database matching would have allowed a perhaps more efficient metabolite annotation. Combining different analytic techniques such as LC-MS, GC-MS (gas chromatography – mass spectrometry) and NMR (nuclear magnetic resonance) would have also offered a more comprehensive coverage of the metabolome. However, in this study we decided to focus on non-volatile compounds based on the defence molecules reported in the literature. We acknowledge, however, that if neither time or cost constrained, coupling LC-MS with other techniques such as GC-MS or NMR would be highly interesting, especially since some volatile compounds might play a role in parasite attraction (Mordue and Birkett 2009).

Summary and Future Perspectives.

In summary, we studied the gill metabolome of different butterflyfish species in Moorea island (French Polynesia) and found that the metabolome fingerprint is related to their diet preferences and their parasitism intensities. Despite the challenges of metabolite annotation in non-model species, we identified several lipidic compounds related to fish diet and several peptides related to parasitism. Peptide class could not be assigned from mass data but their purification and sequencing is currently being performed at the laboratory. Overall, this was a descriptive study that allowed the identification of biomarkers of some factors (i.e. diet or parasitism) but further biological tests are needed to confirm a causative relationship. For example, we identified peptides that were only present in *C. lunulatus*, but whether they are directly responsible for the absence of parasites in this species remains unknown. Therefore, isolation, characterisation, synthesis and in vitro tests need to be performed to confirm whether they do play an active role in the absence of monogenean parasites in *C. lunulatus*. Furthermore, peptide origin remains also unknown, and therefore a multi-omics approach integrating genomics and transcriptomics could better elucidate whether the peptides are from fish, or bacterial origin.

References


He, Q., Sun, R., Liu, H., Geng, Z., Chen, D., Li, Y., Han, J., Lin, W., Du, S.,& Deng, Z., 2014. NMR-based metabolomic analysis of spatial variation in soft corals. Marine Drugs, 12, 1876–1890.


Table 1. Details on the ecology and biological traits of butterflyfish sampled in the island of Moorea to study the gill mucus metabolome.

<table>
<thead>
<tr>
<th>Species</th>
<th>Phylogenetic clade</th>
<th>Diet</th>
<th>Parasitism intensity (mean ± SD)</th>
<th>Geographic location</th>
<th>Reef habitat</th>
<th>N</th>
<th>Standard length (cm) (mean ± SD)</th>
<th>Weight of freeze-dried mucus (mg) (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. auriga</em></td>
<td>Clade 4</td>
<td>Invertebrates (e.g., polychaetes 40-95%), hard corals (3–60%),4*</td>
<td>43 ± 45 Low</td>
<td>Ta’ahiamanu</td>
<td>Fringing</td>
<td>1</td>
<td>14.2 ± 0.5</td>
<td>29.0 ± 16.7</td>
</tr>
<tr>
<td><em>C. lunula</em></td>
<td>Clade 4</td>
<td>Hard corals (50%), polychaetes (21%), gastropods (6%)4* Facultative corallivore</td>
<td>87 ± 90 Low</td>
<td>Ta’ahiamanu</td>
<td>Outer slope</td>
<td>7</td>
<td>15.9 ± 2.3</td>
<td>43.6 ± 20.6</td>
</tr>
<tr>
<td><em>C. lunulatus</em></td>
<td>Clade 3</td>
<td>Hard corals (&gt;99%), Facultative corallivore</td>
<td>0 Non-parasitized</td>
<td>Ta’ahiamanu</td>
<td>Fringing</td>
<td>1</td>
<td>11.8 ± 0.7</td>
<td>24.6 ± 4.4</td>
</tr>
<tr>
<td><em>C. ornatis-simus</em></td>
<td>Clade 3</td>
<td>Hard corals (&gt;99%) Facultative corallivore</td>
<td>178 ± 143 High</td>
<td>Ta’ahiamanu</td>
<td>Fringing</td>
<td>2</td>
<td>9.9 ± 1.3</td>
<td>37.7 ± 9.6</td>
</tr>
<tr>
<td><em>C. quadrimaculatus</em></td>
<td>Clade 2</td>
<td>Soft coral (34%), hard corals (28%), filamentous algae (26%) Facultative corallivore</td>
<td>216 ± 163 High</td>
<td>Afareiatu</td>
<td>Fringing</td>
<td>8</td>
<td>10.7 ± 0.6</td>
<td>21.5 ± 3.9</td>
</tr>
<tr>
<td><em>C. reticulatus</em></td>
<td>Clade 3</td>
<td>Hard corals &gt;99% Facultative corallivore</td>
<td>211 ± 155 High</td>
<td>Ta’ahiamanu</td>
<td>Fringing</td>
<td>11</td>
<td>11.6 ± 0.4</td>
<td>35.8 ± 7.6</td>
</tr>
<tr>
<td><em>C. ulietensis</em></td>
<td>Clade 4</td>
<td>Soft corals (27%), scleractinian corals (20%), other invertebrates (&gt;10%)Omnivorous</td>
<td>18 ± 12 Low</td>
<td>Ta’ahiamanu</td>
<td>Outer slope</td>
<td>3</td>
<td>17.4 ± 1.1</td>
<td>43.3 ± 27.9</td>
</tr>
<tr>
<td><em>C. vagabundus</em></td>
<td>Clade 4</td>
<td>Filamentous algae (22%), scleractinians (18%), polychaetes (21%), molluscan eggs (20%)Omnivorous</td>
<td>149 ± 92 High</td>
<td>Ta’ahiamanu</td>
<td>Fringing</td>
<td>9</td>
<td>9.8 ± 0.4</td>
<td>24.8 ± 5.6</td>
</tr>
</tbody>
</table>

1 Fessler and Westneat (2007)  
2 Reverter et al. (2016)  
3 Pratchett (2005)  
4 Harmelin-Vivien and Bouchon-Navarro (1983)  
5 Harmelin-Vivien (1989)  
* indicates at least one of the studies evaluated the butterflyfish diet in Moorea (French Polynesia).  
Images kindly given by Laurent Puyaud (Lengguru, IRD)
Table 2. VIP metabolites for diet identified in the non-polar fraction of the butterflyfish gill mucus using UPLC-HRMS/MS and the METLIN and Lipid Maps Structure Database. Chain positions and double bond regiochemistry remain undetermined.

<table>
<thead>
<tr>
<th>VIP Id</th>
<th>VIP score</th>
<th>m/z</th>
<th>Molecular formula</th>
<th>Class/sub-class</th>
<th>Common name</th>
<th>Delta Da (ppm)</th>
<th>Fragmentation patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>M706T687</td>
<td>1.69</td>
<td>706.5451</td>
<td>C&lt;sub&gt;38&lt;/sub&gt;H&lt;sub&gt;77&lt;/sub&gt;N&lt;sub&gt;8&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;P</td>
<td>Diacylglycerophosphocholines</td>
<td>PC (30:0)</td>
<td>.0069 (9.8)</td>
<td>184.0731; 526.4821; 508.4716; 300.2890; 282.2785; 264.2680</td>
</tr>
<tr>
<td>M732T917</td>
<td>1.64</td>
<td>732.5836</td>
<td>C&lt;sub&gt;41&lt;/sub&gt;H&lt;sub&gt;83&lt;/sub&gt;N&lt;sub&gt;7&lt;/sub&gt;O&lt;sub&gt;6&lt;/sub&gt;P</td>
<td>Diacylglycerophosphocholines</td>
<td>PC(33:1) ou</td>
<td>.0066 (9.0)</td>
<td>184.0731</td>
</tr>
<tr>
<td>M510T559</td>
<td>1.54</td>
<td>510.3548</td>
<td>C&lt;sub&gt;25&lt;/sub&gt;H&lt;sub&gt;53&lt;/sub&gt;O&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Monoacylglycerophosphocholines</td>
<td>LPC(17:0)</td>
<td>.0006 (1.2)</td>
<td>184.0731; 339.2883; 339.2945</td>
</tr>
<tr>
<td>M706T956</td>
<td>1.60</td>
<td>706.5372</td>
<td>C&lt;sub&gt;38&lt;/sub&gt;H&lt;sub&gt;77&lt;/sub&gt;N&lt;sub&gt;8&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;P</td>
<td>Diacylglycerophosphocholines</td>
<td>PC (30:0)</td>
<td>.0010 (1.4)</td>
<td>184.0731</td>
</tr>
<tr>
<td>M370T810</td>
<td>1.62</td>
<td>369.3506</td>
<td>C&lt;sub&gt;23&lt;/sub&gt;H&lt;sub&gt;45&lt;/sub&gt;O&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Fatty acids</td>
<td>FA</td>
<td>.0143 (38.7)</td>
<td>neutral loss 14 amu</td>
</tr>
<tr>
<td>M381T714</td>
<td>1.20</td>
<td>381.3509</td>
<td>C&lt;sub&gt;23&lt;/sub&gt;H&lt;sub&gt;45&lt;/sub&gt;O&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Fatty acids</td>
<td>FA</td>
<td>.0146 (38.3)</td>
<td>neutral loss 14 amu</td>
</tr>
<tr>
<td>M690T722</td>
<td>1.68</td>
<td>689.5578</td>
<td>C&lt;sub&gt;38&lt;/sub&gt;H&lt;sub&gt;77&lt;/sub&gt;N&lt;sub&gt;8&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;P</td>
<td>Ceramide Phosphocholines</td>
<td>PE-Cer(d36:1)</td>
<td>.0014 (2.0)</td>
<td>184.0731</td>
</tr>
<tr>
<td>M556T948</td>
<td>1.77</td>
<td>556.5287</td>
<td>C&lt;sub&gt;53&lt;/sub&gt;H&lt;sub&gt;98&lt;/sub&gt;N&lt;sub&gt;4&lt;/sub&gt;O&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Ceramide</td>
<td>Cer(34:0)</td>
<td>.0013 (2.3)</td>
<td>264.2680; 282.2787</td>
</tr>
<tr>
<td>M700T811</td>
<td>1.74</td>
<td>698.5545</td>
<td>C&lt;sub&gt;60&lt;/sub&gt;H&lt;sub&gt;102&lt;/sub&gt;N&lt;sub&gt;4&lt;/sub&gt;O&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Ceramide</td>
<td>HexCer(d34:2)</td>
<td>.0021 (3.0)</td>
<td>264.2680; 282.2785</td>
</tr>
<tr>
<td>M721T717</td>
<td>1.35</td>
<td>720.5599</td>
<td>C&lt;sub&gt;38&lt;/sub&gt;H&lt;sub&gt;77&lt;/sub&gt;N&lt;sub&gt;8&lt;/sub&gt;O&lt;sub&gt;5&lt;/sub&gt;P</td>
<td>Diacylglycerophosphocholines</td>
<td>PC(31:0)</td>
<td>.0069 (9.6)</td>
<td>184.0731</td>
</tr>
</tbody>
</table>
Table 3. VIP metabolites for parasitism identified in the polar fraction of the butterflyfish gill mucus using UHRMS/MS.

<table>
<thead>
<tr>
<th>VIP Id</th>
<th>VIP score</th>
<th>VIP position</th>
<th>m/z</th>
<th>Ion charge</th>
<th>Peptide exact mass calculated from multi-charged ions (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M683T693</td>
<td>1.81</td>
<td>30&lt;sup&gt;th&lt;/sup&gt;</td>
<td>682.8862</td>
<td>4</td>
<td>2727.5132</td>
</tr>
<tr>
<td>M712T744</td>
<td>2.03</td>
<td>7&lt;sup&gt;th&lt;/sup&gt;</td>
<td>711.7073</td>
<td>3</td>
<td>2132.0982</td>
</tr>
<tr>
<td>M645T780</td>
<td>2.23</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>644.8723</td>
<td>2</td>
<td>1287.7288</td>
</tr>
<tr>
<td>M657T836</td>
<td>2.27</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>657.1578</td>
<td>5</td>
<td>3280.5495</td>
</tr>
<tr>
<td>M782T855</td>
<td>2.00</td>
<td>11&lt;sup&gt;th&lt;/sup&gt;</td>
<td>782.1692</td>
<td>4</td>
<td>3124.6452</td>
</tr>
<tr>
<td>M757T860</td>
<td>2.06</td>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>756.9081</td>
<td>4</td>
<td>3023.6008</td>
</tr>
<tr>
<td>M760T942</td>
<td>2.07</td>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>759.4188</td>
<td>6</td>
<td>4550.4654</td>
</tr>
</tbody>
</table>
Figure 1. Map of the Moorea Island (French Polynesia) with the sampling sites at three geographic locations (Ta’ahiamanu, Haapiti and Afareiatu) and three reef environments (fringing reef, barrier reef and outer slope) at the Ta’ahiamanu site with their coordinates.
Figure 2. Workflow of the study investigating the metabolomics variability in butterflyfish gill mucus from the island of Moorea.
Figure 3. Multiple regression tree of the gill mucus metabolome from the polar (a) and non-polar (b) fractions with the % of variance explained by each split. O corallivore = obligate corallivore, F corallivore = facultative corallivore, High = high parasite intensities, Low = low parasite intensities, No = never parasitized

Error : 0.791  CV Error : 0.915  SE : 0.0567

Error : 0.889  CV Error : 1.06  SE : 0.0788