High bacterial diversity in near-shore and oceanic biofilms and their influence on larval settlement by *Hydroides elegans* (Polychaeta)

Lema, Kimberley A.; Constancias, Florentin; Rice, Scott A.; Hadfield, Michael G.

2019

Lema, K. A., Constancias, F., Rice, S. A., & Hadfield, M. G. (2019). High bacterial diversity in nearshore and oceanic biofilms and their influence on larval settlement by *Hydroides elegans* (Polychaeta). *Environmental Microbiology*, 21(9), 3472-3488.
doi:10.1111/1462-2920.14697

<https://hdl.handle.net/10356/139838>

<https://doi.org/10.1111/1462-2920.14697>

© 2019 Society for Applied Microbiology and John Wiley & Sons Ltd. All rights reserved. This paper was published in *Environmental Microbiology* and is made available with permission of Society for Applied Microbiology and John Wiley & Sons Ltd.

Downloaded on 26 Apr 2023 22:04:47 SGT

High bacterial diversity in near-shore and oceanic biofilms and their influence on larval settlement by *Hydroides elegans* (Polychaeta)

Journal:	<i>Environmental Microbiology and Environmental Microbiology Reports</i>
Manuscript ID	EMI-2019-0705
Journal:	Environmental Microbiology
Manuscript Type:	EMI - Research article
Date Submitted by the Author:	21-May-2019
Complete List of Authors:	Lema, Kim; University of Hawai'i at Manoa Hawaii Research Center for Futures Studies, PBRC Constancias, Florentin ; Nanyang Technological University, 1. Singapore Centre for Environmental Life Sciences Engineering Rice, Scott; Nanyang Technological University, The Singapore Centre for Environmental Life Sciences Engineering Hadfield, Michael; University of Hawaii, Kewalo Marine Laboratory
Keywords:	Microbial communities, Biofilms, <i>Hydroides elegans</i> , Metacommunity sequencing, Biofouling

SCHOLARONE™
Manuscripts

High bacterial diversity in near-shore and oceanic biofilms and their influence on larval settlement by *Hydroides elegans* (Polychaeta)

Running title: Marine biofilm diversity and settlement of *H. elegans*

Authors: Kimberley A. Lema^{1*}, Florentin Constancias², Scott A. Rice^{2,3,4}, Michael G. Hadfield¹

Addresses of authors:

¹ Kewalo Marine Laboratory, University of Hawai‘i Mānoa, Honolulu, HI 96813, USA

² The Singapore Centre for Environmental Life Sciences Engineering and ³The School of Biological Sciences, Nanyang Technological University, Singapore 637551

⁴The iThree Institute, The University of Technology Sydney, Sydney, NSW Australia

*** Corresponding Author:**

Address: The University Of New South Wales
School of Biological, Earth and Environmental Sciences
Sydney, NSW 2052 Australia
Telephone: + 61 29385 87 23
Current email: kimlema85@gmail.com

Key Words: Microbial communities, Biofilms, *Hydroides elegans*, Metacommunity sequencing, Biofouling

22 **Originality Significance statement**

23 Surface microbial films (biofilms) are known to induce recruitment of larvae of many
24 marine invertebrate species and thus are a major key to the establishment and maintenance of
25 benthic communities. This article explores and compares the bacterial diversity in marine
26 biofilms from distant and diverse locations and tests the capacity of the biofilms to induce
27 settlement of larvae of a circum-global polychaete worm, *Hydroides elegans*. Thus, it provides
28 novel, in-depth understanding of the compositions of marine biofilms from nearshore and
29 offshore oceanic sites and their roles in differential benthic community development.

30

31 **ABSTRACT**

32 Settlement of many benthic marine invertebrates is stimulated by bacterial biofilms,
33 although it is not known if patterns of settlement reflect microbial communities that are specific
34 to discrete habitats. Here, we characterized the taxonomic and functional gene diversity (16S
35 rRNA gene amplicon and metagenomic sequencing analyses), as well as the specific bacterial
36 abundances, in biofilms from diverse nearby and distant locations, both inshore and offshore, and
37 tested them for their ability to induce settlement of the biofouling tubeworm *Hydroides elegans*,
38 an inhabitant of bays and harbors around the world. We found that compositions of the bacterial
39 biofilms were site specific, with the greatest differences between inshore and offshore sites.
40 Further, biofilms were highly diverse in their taxonomic and functional compositions across
41 inshore sites, whilst relatively low diversity was found at offshore sites. *H. elegans* settled on all
42 biofilms tested, with settlement strongly correlated with bacterial abundance. Bacterial density in
43 biofilms was positively correlated with biofilm age. Our results suggest that the localized

distribution of *H. elegans* is not determined by "selection" to locations by specific bacteria, but is more likely linked to the prevailing local ecology and oceanographic features that affect the development of dense biofilms and the occurrence of larvae.

1. INTRODUCTION

Complex microbial films (biofilms) rapidly form on surfaces placed in the sea with a variety of possible consequences. They provide refuges for many bacteria within their secreted layers of extracellular polymeric substances and they are responsible for a variety of other effects, depending on the nature and location of the surface. Bacteria residing in marine biofilms are increasingly recognized as stimuli for the settlement and metamorphosis of the larvae of many marine invertebrates across a broad spectrum of phyla (reviewed in Hadfield, 2011). There is growing evidence that in most cases the capacity of bacteria to induce settlement of invertebrate larvae is species-specific on both sides. That is, for any given larval species, only a few bacterial strains isolated from biofilms induce settlement (e.g., Unabia and Hadfield, 1999; Dobretsov and Qian, 2006, Bao et al., 2007, Tran and Hadfield, 2011). For example, biofilms of *Pseudoalteromonas luteoviolacea* HI1 (genbank accession numbers EU072022 as KMB1 for the 16S sequence and JWIC01000000 as HI1 for the full genome) induce rapid settlement of *H. elegans*, while biofilms of a closely related strain, *P. luteoviolacea* (ATCC X82144.1) fail to induce settlement (Huang et al. 2012). Metamorphosis of *H. elegans* is also triggered by taxonomically distant strains of *Cellulophaga lytica*, *Bacillus aquimaris* and *Staphylococcus warneri*, suggesting that there might be conserved mechanisms that induce settlement and metamorphosis (Freckelton et al., 2017). This is also the case for other marine organisms such as

macroalgal zoospores (e.g., *Ulva*) that respond to chemical signals from specific bacteria for settlement (Tait et al., 2009).

Larval settlement is the basis for the creation and maintenance of benthic marine communities and thus is important in both mariculture and maritime industries. Because larval recruitment is necessary for sustaining fisheries and aquaculture industries, and because biofouling has significant, negative impacts on shipping, the drivers of larval settlement need to be better understood.

The serpulid polychaete worm *Hydroides elegans* is a primary biofouling animal in warm-water bays and harbors around the world, including our field site in Pearl Harbor, Hawai'i. The worm is easily maintained in the laboratory, and its gametes can be readily obtained to achieve fertilization and larval rearing through metamorphosis (see Nedved and Hadfield, 2008). For these reasons, *H. elegans* has emerged as a major model organism for investigations of the factors and processes that trigger larval settlement. Environmental biofilms from Pearl Harbor, Hawaii or laboratory aquaria that receive a continuous flow of unfiltered seawater have been shown to induce the settlement of *H. elegans*, as well as biofilms of single bacterial species originally isolated from these complex biofilms. In all experiments, settlement success increased as the age and bacterial density in the biofilms increased (Hadfield et al., 1994; Huang and Hadfield, 2003; Shikuma and Hadfield, 2005; Huggett et al., 2009). These prior discoveries suggested that settlement success was not related to a shift in bacterial communities as the biofilms matured but, rather, was due to the presence of specific and ubiquitous bacterial strains that increased in the biofilms over time (Shikuma and Hadfield, 2005; Huggett et al., 2009). For example, the highly inductive bacterium *P. luteoviolacea* has been cultured from biofilms in multiple locations, e.g., Hawai'i, Australia and Hong Kong (Huang and Hadfield, 2003; Lau et

al., 2005; Huggett et al., 2006). Other studies have shown similarly that the species composition of the bacterial community affects larval settlement for species such as barnacles (Lau et al., 2005; Lee et al., 2014), bryozoans (Dobretsov and Qian, 2006) and mussels (Bao et al., 2007). However, metamorphosis of *H. elegans* is also triggered by taxonomically distant strains of *Cellulophaga lytica*, *Bacillus aquimaris* and *Staphylococcus warneri*, suggesting that there might be conserved mechanisms that induce settlement and metamorphosis (Freckelton et al., 2017). Thus, it remains to be determined if *H. elegans* responds to specific bacterial biofilm communities or to some specific community members with conserved functional genes. Indeed functional redundancy can be found between distantly related bacteria (Burke et al., 2011; Lema et al., 2014).

To date, the isolated bacterial strains that induce settlement of *H. elegans*, as well as the environmental biofilms that have been the subject of settlement assays, have come from bays and harbors where the tubeworm resides (e.g., Hawai'i, Unabia and Hadfield, 1999; Hong Kong, Jin and Qian, 2005; India, Arumugam, 2012; California, Shikuma et al., 2014). This leaves the significant question: are the bacteria that induce settlement uniquely present in these harbor biofilms compared to biofilms from locations not known to harbor dense populations of *H. elegans*?

In this study, we compared the bacterial diversity and settlement of *H. elegans* in response to biofilms collected from sites where *H. elegans* is known to be present and sites where it has not been observed. For sites inhabited by *H. elegans*, biofilms were collected from Pearl Harbor, HI, the source of *H. elegans* for many published studies, aquaria of the Kewalo Marine Laboratory, where the development and recruitment of *H. elegans* have been investigated for over 20 years, and a distant location in Changi Lagoon, Singapore. Study sites where *H.*

elegans had not been found included a coral reef and a harbor mouth with strong wave action both near O‘ahu, Hawai‘i, an open coastal site on Hawai‘i Island, an island site in Singapore Straight, and a deep-ocean site in the north Pacific Ocean. To test for reproducibility, we also compared settlement on biofilms collected 6 weeks apart at one site, Pearl Harbor, and from biofilms on slides that had been suspended in Kewalo Marine Laboratory aquaria for 7 d and approximately 3 months. We characterized the taxonomic and functional gene diversity through 16S rRNA gene amplicon and metagenomic sequencing analyses, as well as the bacterial abundance in these biofilms. We also compared the taxonomic diversity of bacteria in the water column adjacent to the biofilm surfaces. In addition, we searched the genomes of bacterial species known induce settlement of *H. elegans*.

2. RESULTS

Bacterial community (amplicon sequencing) analysis of biofilms and seawater

A total of 3,558,890 bacterial gene sequences from biofilms and seawater samples were retrieved after analysis and quality filtering. Comparative diversity analysis (rarefied to 5,436 sequences per sample, 138 samples in total) identified a total of 35,572 OTUs (at 97% DNA similarity threshold). Offshore samples (Station ALOHA) had significantly lower diversity than inshore samples, and, among inshore sites, biofilm and seawater samples were also significantly different (Figure 1; Supplementary Figure 1; Supplementary Table 1). At offshore sites, the average number of OTUs was 131 (+/- SE 2.8) for biofilm samples and 120 (+/- SE 7.8) for seawater samples (Figure 1). Significantly higher numbers of OTUs were found for inshore

samples, 1,440 (+/- SE 18.8) and 685.7 (+/- SE 112.4) OTUs in biofilm and seawater samples, respectively. Shannon diversity indices supported similar patterns (Supplementary Figure 1, Supplementary Table1). Rarefaction curves demonstrated that, at a sampling effort of 5,436 sequences per sample, the number of OTUs sampled was sufficient to represent the diversity in samples (Supplementary Figure 2).

Diversity of bacteria

Taxonomic analysis (16S rRNA amplicon) revealed that all biofilms were composed primarily of *Alphaproteobacteria*, *Gammaproteobacteria*, *Flavobacteria* and *Sphingobacteria*, collectively representing $\geq 60\%$ of OTUs in all samples. Although the relative abundances of these taxa differed between samples and particularly when comparing inshore and offshore Station ALOHA sites, the taxonomic composition of offshore biofilm samples was significantly less diverse than that of inshore biofilms. Offshore samples were largely dominated by *Gammaproteobacteria* ($>50\%$ of sequences), while many groups found in inshore samples, such as *Cyanobacteria*, *Acidimicrobiia*, *Verrumicrobiae*, *Opitutae*, *Cytophagia* and *Planctomycetes*, were present in lower relative abundances or were absent in offshore samples (Supplementary Figure 3).

Inshore biofilm samples differed from inshore seawater samples in the relative abundance of many groups, e.g., *Sphingobacteria*, *Planctomycetacia*, *Cytophagia*, *Bacteroidetes* and *Opitutae*, which were much lower or absent in inshore seawater samples. While *Cyanobacteria* represented the third most abundant group (after *Alphaproteobacteria* and *Gammaproteobacteria*) across inshore seawater samples (with a mean abundance of $17.8\% \pm 3.7$ SE), this group was lower or absent in inshore biofilms samples ($1.9\% \pm 0.4$ SE). In contrast, offshore biofilm and

offshore seawater (Station ALOHA) samples were similar in the taxonomic groups present and their relative abundances (Supplementary Figure 3).

At the genus level, inshore biofilm samples were diverse (340 genera) and different across sites. The offshore biofilms were less diverse (150 genera) but highly similar to each other, despite being separated by 400-500 m as well as being similar to their corresponding offshore seawater samples (Figure 2). Offshore samples were dominated by only a few genera: *Alteromonas* spp. (42.6% +/- SE 1.8), *Shimia* spp. (9.6 +/- SE 0.3), *Tenacibaculum* spp. (11.4% +/- SE 1.2) and *Marinobacter* spp. (4.6 +/- SE 0.7), all of which were absent or in remarkably low abundance in inshore biofilms samples (<1% off all biofilm inshore sequences). Inshore biofilms were diverse, including many bacterial genera in low abundance, although most samples harbored relatively high abundances of *Rhodobacteraceae* (~8.8% +/- SE 0.6) and *Thalassobius* spp. (8.1 +/- SE 1.5). Samples from the Kewalo Marine Channel had higher abundances of *Rubritalea* spp. and *Photobacterium* spp. Some bacterial genera, e.g., *Epibacterium*, *Hyphomonas* sp., *Rhodobacteraceae*, *Vibrionaceae*, *Saprospiraceae* and *Erythrobacter*, were found across all biofilm samples including inshore and offshore samples, although with different and lower relative abundances (Figure 2). Inshore seawater samples had high abundances of *Synechococcus* spp. (15.9% +/- SE 4.2), SAR116 (10.9% +/- SE 4) and SAR11 clade (with *Candidatus Pelagibacter* representing 6.3% +/- SE 1.6 and other members belonging to the SAR11 clade representing 3.8% +/- SE 0.02). The later were absent or almost (<0.001%) in all other samples including inshore biofilm samples and offshore seawater and biofilm samples.

The 100 most abundant OTUs and their average percent abundance for different samples (representing ~43% of all sequences) are shown in Supplementary Table 2. The 20 most abundant OTUs are also represented in Figure 3A. Offshore biofilms were dominated by OTU1

179 that was most closely related to an *Alteromonas* sp. with >39 % mean abundance in each sample
 180 and this OTU was absent or almost in inshore biofilm samples. The second and third most
 181 abundant OTUs: OTU2 and OTU3 *Shimia* sp. and *Epibacterium* sp., respectively, were
 182 primarily found in offshore seawater samples. However, these were also relatively abundant in
 183 offshore biofilm samples while absent or almost in inshore biofilm and seawater samples
 184 (representing 19.3 and 20.6% in offshore seawater samples and an average of 9.6% +/- SE 0.3
 185 and 2.6 % +/- SE 0.2 respectively in offshore biofilm samples). Similarly, several OTUs were
 186 found in offshore samples while being almost entirely absent in inshore samples (relatively
 187 abundances ranging from 0.3%-4.7%): OTU8 (*Tenacibaculum* sp.), OTU10 (*Tenacibaculum*
 188 sp.), OTU12 (*Arenicella* sp.) and OTU16 (*Pseudoalteromonas* sp.). On the opposite, OTU5
 189 (*Thalassobius* sp.) was only found in inshore biofilms and was absent in offshore biofilms.
 190 OTU4 (*Hyphomonas* sp.), and 3 OTUs related to *Rhodobacteraceae* (OTU6, OTU9 and OTU13)
 191 were found across in inshore and offshore biofilms though were absent or almost in older
 192 positive control settlement biofilms formed in the sea-water table at the Kewalo Marine
 193 Laboratory ('Sett_Cont'). OTU7 (*Synechococcus* sp.), OTU15 (*Candidatus Pelagibacter* sp.),
 194 OTU18 (*Rhodobacteraceae*) and OTU19 (*Psychrobacter* sp.) were representatives of inshore
 195 seawater samples. OTU17 (*Saprospiraceae* sp.) and OTU20 (*Pseudahrensia* sp.) were solely
 196 found in inshore biofilm samples and were abundantly low across samples (Figure 3A,
 197 Supplementary Table 2).

198 ***Bacteria community composition***

199 Non-metric, multidimensional analysis (NMDS Bray-Curtis distance) and ANOSIMs
 200 statistical analyses of bacterial community structure revealed that compositions of bacterial

communities were particularly diverse and significantly different among samples from deployment sites, inshore vs offshore sites, as well as between biofilms and seawater samples (ANOSIMs $p < 0.001$; Supplementary Table 3). Indeed, the distribution of the samples based on NMDS ordination corresponded to the sites where the biofilms were collected. The bacterial biofilm communities were also different from the seawater samples collected from the same sites as the biofilms (Figure 3; Supplementary Table 3). Offshore samples, biofilms and seawater, clustered together and were significantly different from inshore samples. Cluster analysis using UPGMA dendrogram confirmed that bacterial communities from inshore seawater differed from offshore samples and inshore biofilm samples, with distances of ~80% among those clusters (Figure 4). Bacterial communities from biofilms differed according to sample location with dissimilarities ranging from ~57% (Coconut Island vs Pearl Harbour to ~75% (Singapore vs Other samples)(Figure 4, Supplementary Figure 4). In addition, statistical analysis on biofilm-OTU abundances (MVAbund) showed that the first 100 OTUs contributed 96.2% of the differences in multivariate analysis with OTU1 contributing 18% of the total difference. The bacterial community composition also changed according to the age of biofilms. For example, biofilms formed in the Kewalo Marine Laboratory aquaria for 7d were significantly different from biofilms 3months old PERMANOVA analysis for samples collected at the same site but at different times (i.e., experiments T1 and T2) showed that both duration of biofilm formation and location were significantly different (Supplementary Table 4).

Gene functional community composition

A total of 25,498 functions were identified in the metagenomes (Supplementary Table 5). These functions were mainly related to carbohydrates, amino acids and derivatives and protein metabolism SEED subsystem level 1 groups (Supplementary Figure 5, Supplementary Table 6).

Functional gene compositions for the communities were significantly different between sites (Figure 3B, PERMANOVA ($R^2 = 0.90$ and $p = 0.004$)). Similar patterns between sites were observed at the subsystem level 2 (Figure 5), and functions involved in toxins and superantigens, RNA processing and modification, lipoic acid and protein folding, were more abundant in the metagenomes from Pearl Harbor (PH1, PH2) as compared to the other samples. In contrast, functions involved in TCA (tricarboxylic acid cycle), biosynthesis of phenylpropanoids, reverse electron transport, and siderophores were more abundant in the settlement control (Sett_Cont) metagenome than in the other samples (Figure 5).

Presence of previously retrieved sequences in settlement studies of *H. elegans*

Four bacterial species previously isolated from Pearl Harbor or Kewalo Marine Laboratory sea-table biofilms have been found to induce settlement by larvae of *H. elegans* in single species biofilms: *P. luteoviolacea*, *C. lytica*, *S. warneri* and *B. aquimaris* (Freckelton et al., 2017). Each of these bacteria was rare or absent in both amplicon and metagenomic analyses (Supplementary Table 7 and 8). For example, *P. luteoviolacea* HI1 (a strain that is extremely inductive to *H. elegans* in monospecific biofilms) was present only in a few samples (Station ALOHA 8, 10 and 11; KC3, KST_L, PH1_T2, PH2_T1, PH2_T2, PH3_T3, Sett_Cont_T2) with only 1 - 9 sequences in the 16S dataset (Supplementary Table 7). This was confirmed with the metagenome dataset where less than 2,000 of a total of 10,562,820 to 12,620,925 reads aligned to the reference genome for *P. luteoviolacea* (ASM175040v1) (Supplementary Table 8). Two sequences matching *S. warneri* (KML) were recovered from Puako Bay, Hawai'i Island. No 16S sequences matched *C. lytica* or *B. aquimaris*. Based on metagenome-read mapping, both *S. warneri* (ASM202642v1) and *C. lytica* (CL8139) were detected in all communities, although at extremely low proportions, ranging from 0.002% to 0.013% of all reads (Supplementary Table

8). Some 16S sequences were retrieved from inductive biofilms that were previously identified in another study based on DGGE analysis (Shikuma and Hadfield 2006). Particularly, a sequence related to a *Flavobacteriaceae* (Acc. Num DQ226062) was present as 4,000 reads in one offshore biofilm (Station ALOHA) sample (Supplementary Table 7).

Settlement assays of *H. elegans* on biofilms

Results of settlement assays on all positive settlement controls (Sett_Cont) showed 90 – 96% settlement, demonstrating that the larval cohorts used were fully competent to metamorphose (Figure 6, Table 1). In the absence of biofilms, on clean glass slides in filtered seawater, settlement ranged from 0 – 8% (Figure 6, Table 1), demonstrating that larvae were not prone to spontaneous settlement.

Biofilms from all O‘ahu field sites stimulated the settlement of larvae of *H. elegans* (Figure 6) (experiment T1). The mean settlement for 7-day old biofilms from Coconut Island (CI), Kewalo Channel (KC) and Pearl Harbor (PH) sites in July 2016 were 63.5% ($\pm$ SE 10.4), 59.5% ($\pm$ SE 5.4) and 60.5% ($\pm$ SE 10.4), respectively. Slides kept in a dark aquarium (Kewalo Sea Table; KST_D) showed 50.4% ($\pm$ SE 11.3) settlement. Settlement did not differ significantly across these sites (Kruskal Wallis $p > 0.05$), or when comparing each site to the positive settlement control biofilms (Kewalo Marine Laboratory aquaria, ‘Sett_Cont’, 3 months old) (Dunn's test for multiple comparisons with one control $p > 0.05$) (Figure 6). The lowest settlement was observed on biofilms from the Kewalo Marine Laboratory sea tables exposed to normal light (KST_L; 7 days old) with only 18.2% settlement; this was also the only sample that was statistically different from the other sites.

Larval settlement on biofilms from Pearl Harbor collected from August 25 to September 1, 2016 (experiment T2) was overall 50.3% ($\pm$ SE 4.4) and of 40.4% ($\pm$ SE 4.6), 60.7% ($\pm$ SE 4.6) and 49.8% ($\pm$ SE 4.1) for samples PH1, PH2 and PH3, respectively (Table 1; Figure 6). Settlement was neither significantly different from the settlement on biofilms collected in July 2016 (T1) (Kruskal Wallis $p > 0.01$), nor from positive settlement controls (Dunn's test for multiple comparisons with one control $p > 0.05$).

Settlement on biofilms formed at the deep-water oceanic, off-shore sites around Station ALOHA was lower than for most biofilms collected near shore ($24.7\% \pm$ SE 11.4) and lower than for biofilm samples of the same age deployed in the Kewalo Marine Laboratory sea tables ($51.8\% \pm$ SE 1.85). While the settlement on two of the four offshore Station ALOHA samples was significantly lower than controls 'Sett_Cont' (3 month old biofilms) (Dunn's test for multiple comparisons to one control $p < 0.01$; Figure 6), none of the Station ALOHA biofilms was significantly different from biofilms of the same age formed in the Kewalo Marine Laboratory sea tables (KST_T3) (Kruskal Wallis $p > 0.01$). Settlement assays were not conducted on biofilms collected at the non-O'ahu sites at Puako on Hawai'i Island and near Singapore, because larval rearing facilities were not available.

Bacterial densities on collected biofilms

There was a positive correlation between bacterial densities in biofilms and larval settlement ($r^2=0.77$) (Supplementary Figure 6) (i.e., positive settlement controls and Pearl Harbor site 2); the lowest larval settlement was associated with biofilms showing the lowest bacterial densities (offshore Station ALOHA and KST_L samples) (Figure 6). The average bacterial

abundances for 7-d biofilms from different sites (experiment T1) were similar, ranging between 22,776 (KST_D) and 32,612 (CI) bacteria mm⁻², except for one sample, KST_L, which had only 4,293 ± SE 677 bacteria mm⁻². The average bacterial densities were also similar between biofilms from Pearl Harbor at both collection times (25,168 ± SE 2,221 and 24,754 ± 1,998 bacterial cells mm⁻² for slides from T1 and T2, respectively). The biofilms formed at the offshore samples collected around Station ALOHA (experiment T3, 12 d biofilms) had the lowest bacterial densities, averaging 2,092 bacterial cells mm⁻² (± 335). Bacterial densities from biofilms of similar exposure times collected from sea tables at the Kewalo Marine Laboratory were higher than those from Station ALOHA (KST_T3, 17,012 ± 1,900 cells mm⁻²). Bacterial densities were highest on mature, positive settlement control biofilms ('Sett_Cont', ~3 months exposure) with ~57,000 bacteria mm⁻² (Figure 6).

Non-parametric statistical tests revealed that densities of bacteria across all tested biofilms were significantly lower than their corresponding positive settlement control biofilms from the Kewalo Marine Laboratory aquaria (Sett_Cont T1 and T2 and KST_T3, $p < 0.001$ Kruskal-Wallis nonparametric ANOVA with consecutive Dunn's post-hoc test), although they were not significantly different from each other (i.e., samples within each experiment analyzed separately). The only exceptions were biofilms from the Kewalo sea table kept in the light (KST_L) that were different from the other samples from that experiment (T1) ($p < 0.001$; Kruskal-Wallis nonparametric ANOVA with consecutive Dunn's post-hoc test) (Figure 6). Bacterial densities were significantly different between inshore and offshore samples; i.e., Station ALOHA biofilms vs. all others ($p < 0.001$ Kruskal-Wallis).

4. DISCUSSION

This study sought to expand understanding of the bacterial communities of surface biofilms from nearshore sites in Hawai'i and Singapore, as well as for offshore deep-ocean sites northeast of the Hawaiian Islands. Further, we sought to determine if there are correlations between species composition and settlement rates for larvae of a circum-global polychaete worm, *Hydroides elegans*. The study was designed to determine whether settlement is a reflection of microbial communities that are specific to bays or whether microbial biofilms that originate outside of harbors and bays can also induce its settlement. More generally, comparisons of biofilm diversity across distant sites are scarce, thus this study also sought a trans-oceanic perspective of early biofilm diversity.

Comparison of biofilm community composition

The comparison of biofilm communities within inshore sites revealed that location played a prominent role in bacterial community composition patterns. Indeed, there were significant similarities between biofilm samples collected within O'ahu sites that differentiated them from biofilms collected from Singapore. However, some biofilms from the O'ahu sites appeared to be more different from each other than they were from the Singapore sites. For example, Kewalo Channel biofilms harbored high abundances of *Photobacterium* spp. and *Rubritalea* spp., which were remarkably low in all other O'ahu sites and Singapore samples. The Kewalo Channel site was directly exposed to strong wave action, while all other sites were in somewhat protected areas with calm harbor waters. Thus, it is possible that wave action contributed to differences in community composition between the Kewalo Channel and other sites, although this needs to be addressed in future work. Importantly, inshore marine biofilms from O'ahu, Hawai'i islands and Singapore were composed of diverse bacterial species (OTUs) that revealed site specific

patterns. Indeed, only a few OTUs were shared across these samples and the shared OTUs had very low abundances. These results add evidence that the location, and therefore local environment, plays a prominent role in shaping the distribution of bacterial taxa over space. Therefore, further analyses with higher replication and longer time series might show repeatable patterns when comparing different locations. We also compared samples from the same location after different periods of exposure (i.e., different experiments, T1, T2 and T3 for the Kewalo Marine Laboratory samples at 7 d, 14 d and 3 months old; Figure 3). For this analysis, biofilms collected after similar times of development (i.e., same-age biofilms) from the same location but sampled at different times appeared to be more similar to each other on ordination plots; however, these apparent similarities were not statistically significant (i.e., there were no strong similarities between time or age; see Supplementary Table 3 and 4). Nonetheless, it is possible that further replication might highlight that biofilms of the same age from the same location tend to form similar communities.

Despite the high diversity of inshore biofilms, some bacterial taxa were commonly found across the inshore samples. At the genus level, members of the *Rhodobacteraceae*, *Saprospiraceae*, *Thalassobius* spp., *Hyphomonas* spp., *Epibacterium* spp. and *Erythrobacter* spp. *Rhodobacteraceae* were found in all inshore biofilms. At the species (OTU) level, however, only a few were present across most (not all) inshore samples and at very low abundances (e.g., *Hyphomonas* spp. OTU4; *Thalassobius* spp. OTU5). At the class level, we found that all inshore biofilms collected were predominantly composed of *Alphaproteobacteria*, *Gammaproteobacteria*, although they also all harbored *Flavobacteriia*, *Sphingobacteria*, *Bacteroidetes*, *Cytophagia*, *Deltaproteobacteria* and *Verrucomicrobiae* at different relative abundances. Previous studies of biofilms have similarly found that *Alphaproteobacteria*,

Gammaproteobacteria and *Bacteroidetes* (comprised mainly of *Flavobacteria*, *Sphingobacteria* and *Cytophagia*) dominate in young coastal marine biofilms and also in freshwater stream biofilms (Elifantz et al., 2013; Battin, 2016; Pollet et al., 2018). *Rhodobacteraceae* species are also generally found as primary colonizers in coastal marine biofilms (e.g., Dang et al., 2008; Elifantz et al., 2013; Pollet et al., 2018) and are ubiquitous and abundant microbial components in diverse marine samples (Buchan et al., 2005). Our results are in agreement with other studies of environmental marine biofilms where at higher taxonomical levels (phyla and class) similar taxa were reported, suggesting that they may represent primary biofilms colonizers, while they are generally distinctive at the species level (Elifantz et al., 2013; Battin, 2016; Pollet et al., 2018).

In contrast to inshore biofilms, offshore marine biofilms (samples collected near Station ALOHA) had a low diversity and were homogenous even though the sites of biofilm collection were separated by several hundred meters. Offshore biofilms were predominantly composed of *Gammaproteobacteria*, *Alphaproteobacteria* and *Flavobacteriia*, (i.e., representing over 90% of the bacterial taxa) and were composed of only a few genera; *Alteromonas* spp., *Shimia* spp. *Tenacibaculum* spp., *Arenicella* spp. and *Marinobacter* spp. made up over 70% of the OTUs observed in all samples. These samples were also dominated by a particular species (OTU1) related to *Alteromonas* spp., with >39 % mean abundance in each sample. Importantly, these taxa were not recovered from inshore marine biofilms. Thus, biofilm diversity in off-shore sites were significantly different from coastal biofilms.

Comparison of biofilms with water-column samples

Inshore biofilm samples were largely different from seawater samples collected from the same sites, while offshore biofilms and seawater samples were more similar to each other. Differences between inshore biofilms and seawater were due mainly to the abundance of typical members of global seawater samples such as *Synechococcus* spp. and members of the SAR116 and SAR11 clades (mainly *Candidatus Pelagibacter* spp.) (Morris et al., 2002; Rappé et al., 2002; Flombaum et al., 2013). SAR11 and SAR116 contribute to approximately 25% and 10% of bacterioplankton communities leaving in the euphotic zone, respectively (Fuhrman et al., 2006; Giovannoni 2017). We found that SAR11 and SAR116 contributed ~10% each of inshore seawater samples. In contrast, they were not found in offshore seawater samples that were instead composed of taxa similar to those recovered in offshore biofilms. Offshore sites were highly oligotrophic, which may be one factor explaining the low diversity found in both biofilm and seawater samples; i.e., nutrient availability is a main factor shaping bacterial diversity (e.g., Hewson et al., 2003; Sipura et al., 2005; Walsh et al., 2016). Furthermore, offshore, deep-water samples are not exposed to the complex and varied factors that affect coastal waters such as urbanization, river effluents, structures and biota all which likely contribute to the diversification of coastal biofilm and seawater samples. It was somewhat surprising that members of the SAR11 and SAR116 clades were not recovered in offshore seawater samples, because they are common in oligotrophic gyres. Further work, with increased numbers of samples, different locations and depths around Station ALOHA will be needed to better understand their distribution at this offshore site.

Functional comparisons of bacterial biofilms

We hypothesized that even though inshore biofilms were diverse taxonomically, they could have some similarity in their gene functional composition, as has been suggested from

previous studies (Burke et al., 2011; Lema et al 2014). The functional gene analysis, like the 16S analysis, showed significant patterns that were site specific. Functions implicated in pathways such as FeS clusters, protein translocation across cytoplasmic membrane, transcription, electron transport and photophosphorylation appeared to be systematically more abundant in the metagenomes from the Kewalo Sea Table (KC1, KC2) compared to other sites. A higher proportion of identified genes Coconut Island biofilms (CI1, CI2) was classified to the branched chain amino acids, organic acids and urate degradation according to SEED subsystem level 2. Finally, the positive settlement control (Sett_Cont) exhibited a higher proportion of functions related to type III secretion system extended, ABC transporter, selenoproteins, and lipid-derived mediators. Whilst a higher number of metagenomes samples should be investigated to statistically confirm the highlighted trends, they may suggest that settlement by larvae of *H. elegans* can be induced by multiple mechanisms rather than any specific function that might be conserved in the communities from the various sites. Indeed, it has been found that the common marine biofilm bacterium *P. luteoviolacea* (genbank accession numbers EU072022 and) produces complex aggregates of phage-tail-derived structures that trigger metamorphosis of *H. elegans* in cell-free preparations, while mutants lacking those genes did not (Huang et al., 2012; Shikuma et al., 2014). The metamorphosis of *H. elegans* is also triggered by other taxonomically distant bacteria, including *C. lytica*, *B. aquimaris* and *S. warneri*, that lack the genes that produce bacteriocins (Freckelton et al., 2017). Outer membrane vesicles (OMVs) produced by these diverse bacteria, separated from the bacterial cells, induce *H. elegans* to settle in proportions similar to those when larvae of *H. elegans* are exposed to the bacterial biofilm (Freckelton et al., 2017). It is interesting to note that only about 1% of the bacteria that we can isolate in the laboratory from marine biofilms are inductive of larval settlement (Unabia and Hadfield, 1999;

Lau et al., 2002). Further work is needed to identify inductive species in natural biofilms and determine what fraction of the bacterial species present, even the uncultivable, contributes to larval recruitment.

Larval settlement of *H. elegans* across diverse biofilms

There were no significant differences in *H. elegans* settlement on inshore O'ahu island or Station ALOHA biofilms. For biofilms collected inshore larvae showed similar settlement (>50%), and settlement was positively correlated with bacterial densities. Lower settlement was observed for offshore Station ALOHA biofilms (25%). While the offshore samples were significantly different and less diverse in their bacterial composition, these samples also had the lowest bacterial densities, which may be linked to the lower settlement.

The highest percentage settlement was consistently observed on older biofilms (3 months old), and settlement also correlated with the highest amounts of biofilm biomass. Metagenomic analyses also supported that, while there were differences in functional gene diversity between biofilms, there was no obvious correlation with settlement efficiency. For example, no specific functional trait could be attributed to older positive settlement control biofilms ('Sett_Cont') that had the highest settlement. In addition, we could not relate settlement to any specific bacterial species, because species were so different between samples. Overall, our study shows that larvae of *H. elegans* did not show a significant preference for biofilms from specific sites. This contrasts with some invertebrates such as corals that appear to be specific in their settlement preferences requiring the presence of specific bacteria and chemical compounds (Negri et al., 2001; Tebben et al., 2015). Thus, we conclude that settlement by *H. elegans* was not strongly influenced by biofilm species composition, but rather by total biofilm biomass. This is supported

by previous observations showing that settlement by *H. elegans* as well as other marine invertebrates, including mussels and barnacles, increases with bacterial density (Huang and Hadfield, 2003; Jin and Qian, 2005; Shikuma and Hadfield, 2005; Bao et al., 2007; Hung et al., 2008; Huggett et al., 2009). Given the high diversity of the bacteria found in biofilms and the absence of a specific bacteria or functional traits that could be attributed to settlement efficiency, we suggest that the localized distribution of *H. elegans* is not determined by the presence or absence of specific biofilm community members, but is instead favored by local environmental and hydrographic conditions typically found in bays and harbors.

In conclusion, our analyses revealed that compositions of bacterial biofilms were site specific with the strongest differences between the one offshore and inshore sites. Inductive biofilms were highly diverse in their taxonomic and functional composition within coastal, inshore compositions, suggesting that *H. elegans* may respond to a range of different cues. Further work focusing on environmental data (e.g., DOM, POM, NO₃⁻, Oxygen, etc.) may aid in further understanding selection of specific microbial communities in the near shore biofilms studied here. Importantly, our study suggests that *H. elegans* is more of a generalist in its settlement preferences since it responded to vastly diverse microbial communities. Thus, its localized distribution may be less determined by "selection" to locations with specific bacteria, than by prevailing local ecology and oceanography.

3. MATERIALS AND METHODS

Biofilm sites and collection

Marine biofilms samples on glass slides hung in plastic glass slide holders (21 slides per holder, Supplementary Figure 7) were collected in 2016 and 2017 from seven inshore sites, five in Hawai‘i and two in Singapore, and deep ocean at Station ALOHA site 100 km NE of O‘ahu, Hawai‘i (USA) (see Supplementary methods and Table 1). Replicate biofilm slides from each site were collected and used for DNA extractions (swabbed and stored at -80°C), bacterial density estimations (fixed in PFA), and for larval settlement assays with larvae of *H. elegans* (slides were randomly picked for each treatment) for 3 different sets of settlement assays (T1, T2, and T3; see details on samples used and experiment dates in Supplementary methods and Supplementary Table 1). Seawater samples were also collected at each site and stored at -80°C for DNA analysis in order to investigate differences in bacterial communities between seawater and biofilms (Table 1; Supplementary methods)

DNA extraction and sequencing

DNA extractions were carried out on swabs (biofilms) and Sterivex filters (seawater) using the MasterPure DNA purification kit (Epicenter, Illumina, WI, USA) and following the manufacturer’s instructions (see Supplementary methods). DNA was quantified by Qubit 2.0 (Invitrogen, Cal, USA) and sent for gene-specific 16S rRNA gene (variable region V3-4), paired-end amplicon sequencing (Illumina MiSeq 500 cycle: 2 X 251 bp) by the Argonne National Laboratory (IL, USA) (primers: 515F 5’-GTGYCAGCMGCCGCGGTAA-3’ and 806R 5’-GGACTACNVGGGTWTCTAAT- 3’) (Walters et al., 2016).

Library preparation

For metagenomic library preparation, biofilm DNA samples from replicate slides from each slide holder (see Table 1) were pooled to increase the DNA concentration, and library

preparation was performed according to Illumina's TruSeq Nano DNA Sample Preparation protocol (nb. only inshore biofilm samples from the first experiment, T1, were used in this analysis). The samples were sheared on a Covaris S220 or E220 to ~450 bp, following the manufacturer's recommendation and uniquely tagged with one of Illumina's TruSeq LT DNA barcodes. The finished libraries were quantified using Invitrogen's Picogreen assay, and the average library size was determined using a Bioanalyzer 2100 with a DNA 7500 chip (Agilent). Library concentrations were then normalized to 4 nM and validated by qPCR on a ViiA-7 real-time thermocycler (Applied Biosystems) using the Kapa library quantification kit for Illumina platforms (Kapa Biosystems). The libraries were then pooled at equimolar concentrations and sequenced on an Illumina HiSeq2500 sequencer in rapid mode at a read-length of 250 bp paired-ends.

Bacterial community analysis by amplicon sequencing

Sequence analysis

The 16S rRNA gene amplicon sequences were analyzed using MOTHUR (version 1.38.0.; Department of Microbiology and Immunology, The University of Michigan [<http://www.mothur.org/>]) following the pipeline and recommendations for MiSeq data (Kozich et al. 2013) (See Supplementary methods for details). The raw amplicon sequence datasets for both genes have been uploaded to the short read archive under bioproject accession number PRJNA495084.

Statistics on generated sequences

To compare bacterial diversity among samples, sequences were randomly normalized to 5,610 sequences per sample (the most abundant number of sequences possible to cover the

majority of samples) and assigned to operational taxonomic units (OTUs) at the 97% similarity level through clustering. Where possible, OTUs were identified to the genus level using the SILVA reference taxonomy. Class and genus classification patterns were analyzed in R with the “shared” and “taxonomy” files generated in MOTHUR. For taxonomic analysis, the average abundance of taxonomic groups across replicate slides (from each slide holder) was calculated, and the relative abundances were plotted (percentage abundances) for groups that had a mean relative abundance across slides greater than 0.5%.

Alpha and beta diversity statistics were calculated using the sets of random sub-samples to ensure sequencing effort did not affect the diversity values calculated at the 97% sequence similarity (MOTHUR v.1.38.0). Shannon diversity indices, total observed species (OTUs) and rarefaction estimations were calculated in MOTHUR and plotted in R. Statistics on diversity indices were performed using a dual ANOVA linear model to test differences between sample type (biofilm vs seawater) and site (offshore vs inshore) (the normality of data was checked and data were log transformed; See Supplementary Table 1).

To identify patterns emerging from beta-diversity community structure, a PHYLIP-formatted distance matrix was generated using the Bray-Curtis distance. Changes in community structure were then visualized through (i) non-metric multidimensional scaling (NMDS) and (ii) UPGMA (unweighted pair group method with arithmetic mean) cluster analysis. Analysis of similarities (ANOSIMs) based on 10,000 permutations, was used to determine significant differences in bacterial community structures: (1) between biofilm and seawater; (2) within biofilm samples only across all sites; (3) between inshore and offshore biofilms; and (4) across biofilms of different ages (7d, 14 d and 3 mo) for samples collected at the same location (Kewalo Sea Table samples: KST and Sett_Cont); (See Supplementary Table 3). Additional

PERMANOVA analysis statistical tests ('adonis' function) were performed in order to test temporal versus location differences for samples collected at the two different time points (Pearl Harbor and Kewalo Sea Table samples for experiments T1 and T2) (Supplementary Table 4). To find OTUs matching to previous sequences from KML (Supplementary Table 7), a detailed second analysis without normalization of the data was performed (all sequences analyzed). Additional statistical tests were performed on biofilm OTUs to calculate the percent involvement of the OTUs in patterns of dissimilarity using the R package MVAbund and negative binomial GLMs on OTUs counts (Wang et al., 2012).

Analysis of gene functional community composition of metagenomes

Read preprocessing

Illumina adaptors, 5' or 3' bases with quality scores lower than 20, and read pairs having a mate with an ambiguous base or shorter than 150 bp, were trimmed using 'cutadapt' in paired-end mode (version 1.10, -a -A --max-n 0 -q 20,20 -- --minimum-length 150 -O 6). Overall, quality-trimmed reads represented 83.3% (82.9% - 83.0%) of reads.

*Detection of strains that induced settlement by larvae of *H. elegans**

To estimate the presence of bacteria previously cultured and found to induce larval settlement in *H. elegans*, quality-filtered reads were aligned to reference genomes using Bowtie2 (Supplementary Table 8b; Langmead and Salzberg, 2012, default parameters)

Gene functional community composition profiling

To characterize the functional potential of the metagenomes, quality-trimmed sequences were processed with SUPER-FOCUS (Silva et al., 2016). Briefly, reads were aligned using DIAMOND (--sensitive ref, v.0.9.22) against a 90% similarity reduced protein SEED database (Overbeek et al., 2005). Only alignments of more than 20 amino acids with an identity > 60% and an e-value <1e⁻⁵ were used for analysis (superfocus_app/superfocus.py -q -a diamond -db DB_90 -m 0 -mi 60 -ml 20 -f 0 -n) to generate the normalized relative abundance tables.

Statistical analysis

The functional SEED tables including identified functions and functional categories from levels 3 to 1 classification (i.e., SEED functional hierarchy) were imported into R. The Bray-Curtis distance was then computed using the relative-abundance table to characterize the beta-diversity of the functional community profile, which was visualized using PCo Analysis (i.e., PCoA, 'plot_ordination' function) and heatmap (pheatmap R package, 'pheatmap custer_rows=T, clustering_method = "average", clustering_distance_rows= "correlation", clustering_distance_cols=bray-curtis'). Differences in beta diversity among locations were tested using PERMANOVA statistical tests ('adonis' function).

Settlement assays with larvae of *H. elegans* on biofilms

Settlement assays with 6 d old larvae of *H. elegans* were performed on biofilms collected from the inshore O'ahu Island sites and offshore Station ALOHA sites for three different sets of settlement assays (T1, T2, and T3; see details on samples used, experiments dates and rearing of larvae in Supplementary information and Table 1). Slides were first thoroughly rinsed in 0.22 µm DFASW (Double Filtered Autoclaved Sea Water), and 3-4 slides were randomly selected from each slide holder, cut into smaller similar sized pieces (~ 12-15mm²) and transferred into

24 well plates (Cellstar, Bioexpress, VWR, PA, USA) with each well containing 450 μ L DFASW for settlement assays. Larvae of *H. elegans* were washed in DFASW and uniformly distributed in 50 μ L DFASW into each well (500 μ L total volume). Negative controls consisted of larvae exposed to clean glass slides and positive settlement controls consisted of mature biofilms on glass slides that were incubated for approximately 3 months in KML aquaria ('Sett_Cont' samples, see Table 1). After 24 h of initial exposure, the proportion of settled juveniles in each well was determined by microscopic inspection (where 100% settlement corresponded to all juveniles settling).

Non-parametric analysis of variance with Dunn's multiple comparison test and Bonferroni adjustment (Siegel and Castellan, 1988) was performed with the Pairwise Multiple Comparison of Mean Ranks Package for R (PMCMR; 'dunn.test.control' 'p.adjust=Bonferroni', R version 1.0.136). The significance of differences across sampling sites, inshore and offshore Station ALOHA sites (i.e., excluding positive settlement control) and temporal factor between Pearl Harbor experiments of July and late August were assessed using non-parametric Kruskal-Wallis and Dunn's multiple comparisons of rank sums post-hoc tests (post-hoc Kruskal Dunn test) (R version 1.0.136).

Biofilm bacterial density estimations

Replicate glass slides from each slide holder at the inshore O'ahu Island sites and offshore Station ALOHA sites were used for bacterial density counts (Table 1). Slides were fixed in 4% (w/v) paraformaldehyde (PFA) (in Phosphate Buffer Saline (PBS) at 3% NaCl, pH 7.4), stained with Hoechst and examined with a Zeiss 701 laser-scanning confocal microscope. Captured images were then processed in FIJI (Image J version 2.0.0) where they were converted into 1 bit images using the 'analysis of particles' option (see Supplementary methods for details).

Significant differences among samples (all sites, and inshore vs. offshore sites) were tested using non-parametric Kruskal Wallis analysis of variance, and specific differences amongst samples were calculated using post hoc Dunn's multiple comparisons of rank sums ('posthoc.kruskal.dunn.test' PMCMR package for R).

ACKNOWLEDGEMENTS

Research at the University of Hawai'i was supported by grants no. ORIO-13-013 from the Oak Foundation and N00014-15-1-2658 from the U.S. Office of Naval Research to MGH. Research at the Singapore Centre for Environmental Life Sciences Engineering was supported by the Singapore National Research Foundation (MSRDP-05) and the Singapore Ministry of Education. We would like to acknowledge Dr. David Karl, Dr. Sara Ferrón and the SCOPE operations team from the Center for Microbial Oceanography Research and Education (CMORE) at the University of Hawai'i at Mānoa who collected the offshore samples during the SCOPE cruise ID KM1709.

This is publication number _____ from the School of Ocean and Earth Science and Technology at the University of Hawai'i at Mānoa.

The authors declare they have no conflict of interest.

REFERENCES

- 628 Arumugam, S. (2012) Studies on the effects of Flucloxacillin (antibiotic) on fertilization
629 and early development of a sedentary polychaete *Hydroides elegans*. Intern. J. Environ. Sc. 3(1);
630 616-130.
- 631 Bao, W.Y., Satuito, C.G., Yang, J.-L., and Kitamura, H. (2007) Larval settlement and
632 metamorphosis of the mussel *Mytilus galloprovincialis* in response to biofilms. Mar. Biol. 150:
633 565–574.
- 634 Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., and Packmann, A.I. (2016)
635 The ecology and biogeochemistry of stream biofilms. Nat. Rev. Microbiol. 14: 251–263.
- 636 Buchan, A., González, J.M., and Moran, M.A. (2005) Overview of the marine *Roseobacter*
637 lineage. Appl. Environ. Microbiol. 71: 5665–5677.
- 638 Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial
639 community assembly based on functional genes rather than species. Proc. Natl. Acad. Sci. USA
640 108: 14288–14293.
- 641 Dang, H., Li, T., Chen, M., and Huang, G. (2008) Cross-ocean distribution of
642 Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. Appl.
643 Environ. Microbiol. 74: 52–60.
- 644 Dobretsov, S. and Qian, P.-Y. (2006) Facilitation and inhibition of larval attachment of the
645 bryozoan *Bugula neritina* in association with mono-species and multi-species biofilms. J. Exp.
646 Mar. Biol. Ecol. 333: 263–274.
- 647 Dunn, O.J. (1964) Multiple comparisons using rank sums. Technometrics 6: 241–252.

- 648 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME
649 improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194–2200.
- 650 Elifantz, H., Horn, G., Ayon, M., Cohen, Y., and Minz, D. (2013) Rhodobacteraceae are
651 the key members of the microbial community of the initial biofilm formed in Eastern
652 Mediterranean coastal seawater. *FEMS Microbiol. Ecol.* 85: 348–357.
- 653 Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., et al.
654 (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and
655 *Synechococcus*. *Proc. Natl. Acad. Sci. USA* 110: 9824–9829.
- 656 Freckelton, M.L., Nedved, B.T., and Hadfield, M.G. (2017) Induction of invertebrate larval
657 settlement; different bacteria, different mechanisms? *Sci. Rep.* 7: 42557.
- 658 Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S.
659 (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc.*
660 *Natl. Acad. Sci.* 103: 13104–13109.
- 661 Giovannoni, S.J. (2017) SAR11 Bacteria: The Most Abundant Plankton in the Oceans.
662 *Annu. Rev. Mar. Sci.* 9: 231–255.
- 663 Hadfield, M.G., Unabia, C.R.C., Smith, C.M., and Michael, T.M. (1994) Settlement
664 preferences of the ubiquitous fouler *Hydroides elegans*. In, *Recent Developments in Biofouling*
665 *Control*. AA Balkema, Rotterdam, The Netherlands, pp. 65–74.
- 666 Hewson, I., Vargo, G.A., and Fuhrman, J.A. (2003) Bacterial diversity in shallow
667 oligotrophic marine benthos and overlying waters: effects of virus infection, containment, and
668 nutrient enrichment. *Microb. Ecol.* 46: 322–336.

- 669 Huang, S. and Hadfield, M.G. (2003) Composition and density of bacterial biofilms
670 determine larval settlement of the polychaete *Hydroides elegans*. Mar. Ecol. Prog. Ser. 260: 161–
671 172.
- 672 Huang, Y., Callahan, S., and Hadfield, M.G. (2012) Recruitment in the sea: bacterial genes
673 required for inducing larval settlement in a polychaete worm. Sci. Rep. 2:.
- 674 Huggett, M.J., Nedved, B.T., and Hadfield, M.G. (2009) Effects of initial surface
675 wettability on biofilm formation and subsequent settlement of *Hydroides elegans*. Biofouling 25:
676 387–399.
- 677 Huggett, M.J., Williamson, J.E., de Nys, R., Kjelleberg, S., and Steinberg, P.D. (2006)
678 Larval settlement of the common Australian sea urchin *Heliocidaris erythrogramma* in response
679 to bacteria from the surface of coralline algae. Oecologia 149: 604–619.
- 680 Hung, O.S., Thiagarajan, V., and Qian, P.Y. (2008) Preferential attachment of barnacle
681 larvae to natural multi-species biofilms: Does surface wettability matter? J. Exp. Mar. Biol. Ecol.
682 361: 36–41.
- 683 Jin, T. and Qian, P.Y. (2005) Amino acid exposure modulates the bioactivity of biofilms
684 for larval settlement of *Hydroides elegans* by altering bacterial community components. Mar.
685 Ecol. Prog. Ser. 297: 169–179.
- 686 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013)
687 Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon
688 sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79: 5112–
689 5120.

- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat Meth 9: 357–359.
- Lau, S.C.K., Mak, K.K.W., Chen, F., and Qian, P.-Y. (2002) Bioactivity of bacterial strains isolated from marine biofilms in Hong Kong waters for the induction of larval settlement in the marine polychaete *Hydroides elegans*. Mar. Ecol. Prog. Ser. 226: 301–310.
- Lau, S.C.K., Thiagarajan, V., Cheung, S.C.K., and Qian, P.-Y. (2005) Roles of bacterial community composition in biofilms as a mediator for larval settlement of three marine invertebrates. Aquat. Microb. Ecol. 38: 41–51.
- Lee, O.O., Chung, H.C., Yang, J., Wang, Y., Dash, S., Wang, H., and Qian, P.-Y. (2014) Molecular techniques revealed highly diverse microbial communities in natural marine biofilms on polystyrene dishes for invertebrate larval settlement. Microb. Ecol. 68: 81–93.
- Lema, K.A., Willis, B.L., and Bourne, D.G. (2014) Amplicon pyrosequencing reveals spatial and temporal consistency in diazotroph assemblages of the *Acropora millepora* microbiome. Environ. Microbiol. 16: 3345–3359.
- Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420: 806–810.
- Nedved, B.T. and Hadfield, M.G. (2008) *Hydroides Elegans* (Annelida: Polychaeta): A model for biofouling research. In, SpringerLink, Springer Series on Biofilms. Springer, Berlin, Heidelberg, pp. 1–15.

- 710 Negri, A.P., Webster, N.S., Hill, R.T., and Heyward, A.J. (2001) Metamorphosis of
711 broadcast spawning corals in response to bacteria isolated from crustose algae. *Mar. Ecol. Prog.*
712 *Ser.* 223: 121–131.
- 713 Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M., et al.
714 (2005) The subsystems approach to genome annotation and its use in the project to annotate
715 1000 genomes. *Nuc. Acids. Res.* 33: 5691–5702.
- 716 Pollet, T., Berdjeb, L., Garnier, C., Durrieu, G., Poupon, C.L., Misson, B., and Briand, J.-
717 F. (2018) Prokaryotic community successions and interactions in marine biofilms: the key role of
718 *Flavobacteriia*. *FEMS Microbiol. Ecol.*
- 719 Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner,
720 F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal
721 RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35: 7188–7196.
- 722 Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the
723 ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418: 630–633.
- 724 Shikuma, N.J. and Hadfield, M.G. (2005) Temporal variation of an initial marine biofilm
725 community and its effects on larval settlement and metamorphosis of the tubeworm *Hydroides*
726 *elegans*. *Biofilms* 2: 231–238.
- 727 Shikuma, N.J., Pilhofer, M., Weiss, G.L., Hadfield, M.G., Jensen, G.J., and Newman, D.K.
728 (2014) Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures.
729 *Science* 343: 529–533.

- 730 Silva, G.G.Z., Green, K.T., Dutilh, B.E., and Edwards, R.A. (2016) SUPER-FOCUS: a
731 tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* 32: 354.
- 732 Siegel, S. and Castellan, N.J. (1988) *Nonparametric statistics for the behavioral sciences*.
733 McGraw-Hill, New York.
- 734 Sipura, J., Haukka, K., Helminen, H., Lagus, A., Suomela, J., and Sivonen, K. (2005)
735 Effect of nutrient enrichment on bacterioplankton biomass and community composition in
736 mesocosms in the Archipelago Sea, northern Baltic. *J. Plankton Res.* 27: 1261–1272.
- 737 Tait, K., Williamson, H., Atkinson, S., Williams, P., Cámara, M., and Joint, I. (2009)
738 Turnover of quorum sensing signal molecules modulates cross-kingdom signaling. *Environ.*
739 *Microbiol.* 11: 1792–1802.
- 740 Tebben, J., Motti, C.A., Siboni, N., Tapiolas, D.M., Negri, A.P., Schupp, P.J., et al. (2015)
741 Chemical mediation of coral larval settlement by crustose coralline algae. *Sci. Rep.* 5: 10803.
- 742 Tran, C. and Hadfield, M.G. (2011) Larvae of *Pocillopora damicornis* (Anthozoa) settle
743 and metamorphose in response to surface-biofilm bacteria. *Mar. Ecol. Prog. Ser.* 433: 85–96.
- 744 Unabia, C.R.C. and Hadfield, M.G. (1999) Role of bacteria in larval settlement and
745 metamorphosis of the polychaete *Hydroides elegans*. *Mar. Biol.* 133: 55–64.
- 746 Walsh, E.A., Kirkpatrick, J.B., Rutherford, S.D., Smith, D.C., Sogin, M., and D'Hondt, S.
747 (2016) Bacterial diversity and community composition from seafloor to subseafloor. *ISME J.*
748 10: 979–989.
- 749 Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012) mvabund– an R package
750 for model-based analysis of multivariate abundance data. *Meth. Ecol. Evol.* 3: 471–474.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016) Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* 1: e00009-15.

FIGURE LEGENDS

Figure 1. Alpha diversity based on the number of OTUs from 16S rRNA amplicon sequences comparing inshore and offshore sites for biofilms and seawater samples. Significant differences are highlighted with and asterix (*)

Figure 2. Average abundance of different bacteria (amplicon 16S rRNA gene sequences) retrieved from biofilms and seawater samples at different sites (Table 1). Here, organisms were classified to the Genus level, unless this assignment could not be made, in which case, the organism were classified to the lowest taxonomic rank possible.

Figure 3. A) Non-metric, multidimensional analysis plot (NMDS Bray-Curtis distance) of bacterial (amplicon 16S rRNA 97% sequence identity cut-off) community composition from biofilm and seawater samples from distant sites. The 20 most abundant OTUs are represented by vectors. The effects of location, age and type of sample were performed with ANOSIMs (Supplementary Table 2) **B)** Principal Coordinates Analysis plot (PCoA) using Bray-Curtis

distance computed from identified SEED functions in metagenomes depicting gene functional community composition from biofilm samples from distant sites. The effect of location was tested using PERMANOVA ($R^2 = 0.90$ and $p = 0.004$).

Figure 4. Dendrogram representing differences bacterial community composition in (amplicon 16S rRNA 97% sequence identity cut-off) between samples included in the study. The dendrogram was built using UPGMA method based on a Bray-Curtis distance.

Figure 5. Heatmap displaying proportion of SEED subsystem level 2 categories identified in biofilm metagenomes. The sample dendrogram is based on Bray-Curtis distance summarizing samples dissimilarities. Functional subsystems are clustered according to Pearson's correlation.

Figure 6. Settlement proportions (%) of *Hydroides elegans* larvae and bacteria densities on environmental young biofilms from different sites around and Station ALOHA (Hawai'i, USA). Positive settlement controls ('Sett_Cont') correspond to older biofilms (~3 months old) collected at the Kewalo Marine Laboratory sea tables. Negative controls correspond to 'clean' (no biofilm) glass slides. Results from experiments conducted at different times (T2) and (T3) (see Table 1 for dates) are highlighted in pink and green respectively. Significantly different samples are highlighted (*).

Table 1. Sample information table. Biofilms and seawater locations, age, depth, dates, and replicate numbers for different analysis (DNA extractions and fixed in PFA for bacterial abundance microscope counts) and settlement experiments with *Hydroides elegans* (in July T1 and September T2 2016; and July T3 2017). KST_L, KST_D, and Sett_Cont stand for samples from sea tables at the Kewalo Marine Laboratory kept at light and no light (kept at dark) and positive settlement controls respectively. Samples used for additional metagenomic analyses are highlighted with and asterix (*).

Type of sample	Slide holder name (#replicates slides DNA/ PFA)	Location	Slide holder GPS coordinates	Deployment	Collection	Biofilm age / and depth (m)	Settlement Experiment (#replicate slides)
Biofilm	CI1 (x6/5) *	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.8"N 157°47'10.8"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	CI2 (x5/5)*	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.9"N 157°47'10.8"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	CI3 (x5/5)	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.8"N 157°47'10.9"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	KC1 (x6/5)	Kewalo channel, Honolulu, Hawaii	21°17'24.0"N 157°51'40.6"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	KC2 (x6/5) *	Kewalo channel, Honolulu, Hawaii	21°17'24.7"N 157°51'40.3"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	KC3 (x5/5) *	Kewalo channel, Honolulu, Hawaii	21°17'24.9"N 157°51'40.2"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	KST_L (x6/5)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	07/07/16	14/07/16	7 days/~0.3m	T1 (x4)
Biofilm	KST_D (x6/5)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	07/07/16	14/07/16	7 days/~0.3m	T1 (x4)
Biofilm	KST_T3 (x3)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	02/07/17	16/07/17	14days/~0.3m	T3 (x3)
Biofilm	PH1_T1 (x6/5)	Pearl Harbour, Oahu, Hawaii	21°21'27.2"N 157°57'37.2"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	PH2_T1 (x6/5) *	Pearl Harbour, Oahu, Hawaii	21°21'26.9"N 157°57'36.6"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	PH3_T1 (x5/5) *	Pearl Harbour, Oahu, Hawaii	21°21'26.6"N 157°57'36.2"W	07/07/16	14/07/16	7 days / ~2m	T1 (x4)
Biofilm	PH1_T2 (x6/5)	Pearl Harbour, Oahu, Hawaii	21°21'27.2"N 157°57'37.2"W	25/08/16	01/09/16	7 days / ~2m	T2 (x4)
Biofilm	PH2_T2 (x6/5)	Pearl Harbour, Oahu, Hawaii	21°21'26.9"N 157°57'36.6"W	25/08/16	01/09/16	7 days / ~2m	T2 (x4)
Biofilm	PH3_T2 (x6/5)	Pearl Harbour, Oahu, Hawaii	21°21'26.6"N 157°57'36.2"W	25/08/16	01/09/16	7 days / ~2m	T2 (x4)
Biofilm	Sett_Cont_T1(x6/5)*	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	04/16	14/07/16	~3mo/ ~2m	T1 (x4)
Biofilm	Sett_Cont_T2 (x6/5)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	07/16	01/09/16	~3mo / ~2m	T2 (x4)
Biofilm	Changi (x6)	Changii Fish Farm, Singapore	1°23'16.5"N 103°56'37.5"E	27/07/16	04/08/16	8 days / ~2m	/
Biofilm	St Jhon (x6)	St Jhon Island, Singapore	1°12'56.4"N 103°51'05.3"E	27/07/16	04/08/16	8 days / ~2m	/
Biofilm	BI (x6)	Puako, Big Island, Hawaii	19°58'24.8"N 155°50'11.5"W	09/06/16	08/07/16	29 days / ~2m	/
Biofilm	St8 (x3/3)	Pacific Ocean near ALOHA Station	22°17.580"N 157° 44.622"W	02/07/17	16/07/17	14 days / ~5m	T3 (x3)
Biofilm	St9 (x3/3)	Pacific Ocean near ALOHA Station	22°17.580"N 157° 44.863"W	02/07/17	16/07/17	14 days / ~5m	T3 (x3)
Biofilm	St10 (x3/3)	Pacific Ocean near ALOHA Station	22°17.580"N 157° 45.027"W	02/07/17	16/07/17	14 days / ~5m	T3 (x3)
Biofilm	St11 (x3/3)	Pacific Ocean near ALOHA Station	22°17.580"N 157° 45.148"W	02/07/17	16/07/17	14 days / ~5m	T3 (x3)
Seawater	SW_AlSt (x3/0)	Pacific Ocean near ALOHA Station	22°17.580"N 157° 45.148 "W	/	16/07/17	/ ~5m	T3 (x3)
Seawater	SW_CI1 (x1/0)	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.8"N 157°47'10.8"W	/	14/07/16	/ ~2m	T1
Seawater	SW_CI2 (x1/0)	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.9"N 157°47'10.8"W	/	14/07/16	/ ~2m	T1
Seawater	SW_CI3 (x1/0)	Coconut Island, Kaneohe Bay, Oahu, Hawaii	21°26'00.9"N 157°47'10.8"W	/	14/07/16	/ ~2m	T1
Seawater	SW_KC1 (x1/0)	Kewalo channel, Honolulu, Hawaii	21°17'24.0"N 157°51'40.6"W	/	14/07/16	/ ~2m	T1
Seawater	SW_KC2 (x1/0)	Kewalo channel, Honolulu, Hawaii	21°17'24.7"N 157°51'40.3"W	/	14/07/16	/ ~2m	T1
Seawater	SW_KC3 (x1/0)	Kewalo channel, Honolulu, Hawaii	21°17'24.9"N 157°51'40.2"W	/	14/07/16	/ ~2m	T1
Seawater	SW_KST_L (x1/0)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	/	14/07/16	/ ~0.3m	T1
Seawater	SW_KST_D (x1/0)	Kewalo Marine Laboratory, Honolulu, Hawaii	21°17'26.6"N 157°51'39.8"W	/	14/07/16	/ ~0.3m	T1
Seawater	SW_PH1_T1 (x1/0)	Pearl Harbour, Oahu, Hawaii	21°21'27.2"N 157°57'37.2"W	/	14/07/16	/ ~2m	T1
Seawater	SW_PH2_T1 (x1/0)	Pearl Harbour, Oahu, Hawaii	21°21'26.9"N 157°57'36.6"W	/	14/07/16	/ ~2m	T1

Seawater	SW PH3 T1 (x1/0)	Pearl Harbour, Oahu, Hawaii	21°21'27.2"N 157°57'37.2"W	/	14/07/16	/~2m	T1
Seawater	SW PH2 T2 (x1/0)	Pearl Harbour, Oahu, Hawaii	21°21'26.6"N 157°57'36.2"W	/	01/09/16	/~2m	T2
Seawater	SW PH3 T2 (x1/0)	Pearl Harbour, Oahu, Hawaii	21°21'27.2"N 157°57'37.2"W	/	01/09/16	/~2m	T2

For Peer Review Only

Figure 1. Alpha diversity based on the number of OTUs from 16S rRNA amplicon sequences comparing inshore and offshore sites for biofilms and seawater samples. Significant differences are highlighted with and asterix (*)

Figure 2. Average abundance of different bacteria (amplicon 16S rRNA gene sequences) retrieved from biofilms and seawater samples at different sites (Table 1). Here, organisms were classified to the Genus level, unless this assignment could not be made, in which case, the organism were classified to the lowest taxonomic rank possible

Figure 3. A) Non-metric, multidimensional analysis plot (NMDS Bray-Curtis distance) of bacterial (amplicon 16S rRNA 97% sequence identity cut-off) community composition from biofilm and seawater samples from distant sites. The 20 most abundant OTUs are represented by vectors. The effects of location, age and type of sample were performed with ANOSIMs (Supplementary Table 2) B) Principal Coordinates Analysis plot (PCoA) using Bray-Curtis distance computed from identified SEED functions in metagenomes depicting gene functional community composition from biofilm samples from distant sites. The effect of location was tested using PERMANOVA ($R^2= 0.90$ and $p = 0.004$).

Figure 4. Dendrogram representing differences bacterial community composition in (amplicon 16S rRNA 97% sequence identity cut-off) between samples included in the study. The dendrogram was built using UPGMA method based on a Bray-Curtis distance.

Unable to Convert Image

The dimensions of this image (in pixels) are too large to be converted. For this image to convert, the total number of pixels (height x width) must be less than 40,000,000 (40 megapixels).

Figure 5. Heatmap displaying proportion of SEED subsystem level 2 categories identified in biofilm metagenomes. The sample dendrogram is based on Bray-Curtis distance summarizing samples dissimilarities. Functional subsystems are clustered according to Pearson's correlation.

Figure 6. Settlement proportions (%) of *Hydroides elegans* larvae and bacteria densities on environmental young biofilms from different sites around and Station ALOHA (Hawai'i, USA). Positive settlement controls ('Sett_Cont') correspond to older biofilms (~3 months old) collected at the Kewalo Marine Laboratory sea tables. Negative controls correspond to 'clean' (no biofilm) glass slides. Results from experiments conducted at different times (T2) and (T3) (see Table 1 for dates) are highlighted in pink and green respectively. Significantly different samples are highlighted (*).