

RESEARCH

Open Access



# Microbial hydrocarbon degradation potential of the Baltic Sea ecosystem

Joeselle M. Serrana<sup>1,2\*</sup> , Benoît Dessirier<sup>1,3</sup> , Francisco J. A. Nascimento<sup>1,4</sup> , Elias Broman<sup>1,3,4</sup> and Malte Posselt<sup>1,2</sup>

**Background** The Baltic Sea receives petroleum hydrocarbons from various point sources. The degradation of these contaminants in the environment is typically facilitated by a variety of microorganisms that possess a range of genes and metabolic functions related to the degradation of various hydrocarbon substrates. However, our understanding of natural attenuation and the microbial capacity to degrade these contaminants within the Baltic Sea ecosystem remains limited. In this study, we compiled metagenomes from the benthic and pelagic ecosystems across the Baltic Sea to identify microorganisms and characterize their genes and metabolic functions involved in the degradation of hydrocarbon compounds.

**Results** Known hydrocarbon-degrading phyla, i.e., Pseudomonadota, Myxococcota A, Actinomycetota, and Desulfobacterota, were identified within the Baltic Sea metagenome-assembled genomes (MAGs). Notably, 80% of the MAGs exhibited multiple hydrocarbon degradation gene annotations (> 10 reads per kilobase million). Aerobic degradation was the predominant pathway for hydrocarbon degradation across environmental samples. Hydrocarbon degradation gene abundances varied among samples and Baltic Sea subbasins, with long-chain alkanes and dibenzothiophene compounds being the preferred substrates. Species richness and diversity of both benthic and pelagic microorganisms positively correlated with hydrocarbon degradation gene diversity, with the pelagic ecosystem exhibiting significantly higher richness and diversity compared to the benthic ecosystem. Additionally, the composition of the hydrocarbon degradation genes across the Baltic Sea subbasins was influenced by oil spill history, with areas that experienced higher spill volumes showing lower microbial diversity, suggesting potential enrichment of specific hydrocarbon degraders. Among the environmental factors assessed, depth played a significant role in shaping the composition of genes involved in hydrocarbon degradation within the Baltic Sea.

**Conclusions** Using metagenomics, we profiled the native microorganisms associated with hydrocarbon degradation in the Baltic Sea. This knowledge will aid in understanding the natural capacities of microbial communities, potentially linked to the natural attenuation of hydrocarbon pollutants in the area. Insights into microbial degradation potential can enhance predictions of petroleum pollutant persistence and accumulation, support mitigation strategies for marine pollution, and reveal the ecological resilience of native microbial communities in marine ecosystems.

**Keywords** Baltic Sea, Benthic sediments, Environmental microbiome, Hydrocarbon degradation, Metagenomics, Pelagic water

\*Correspondence:

Joeselle M. Serrana

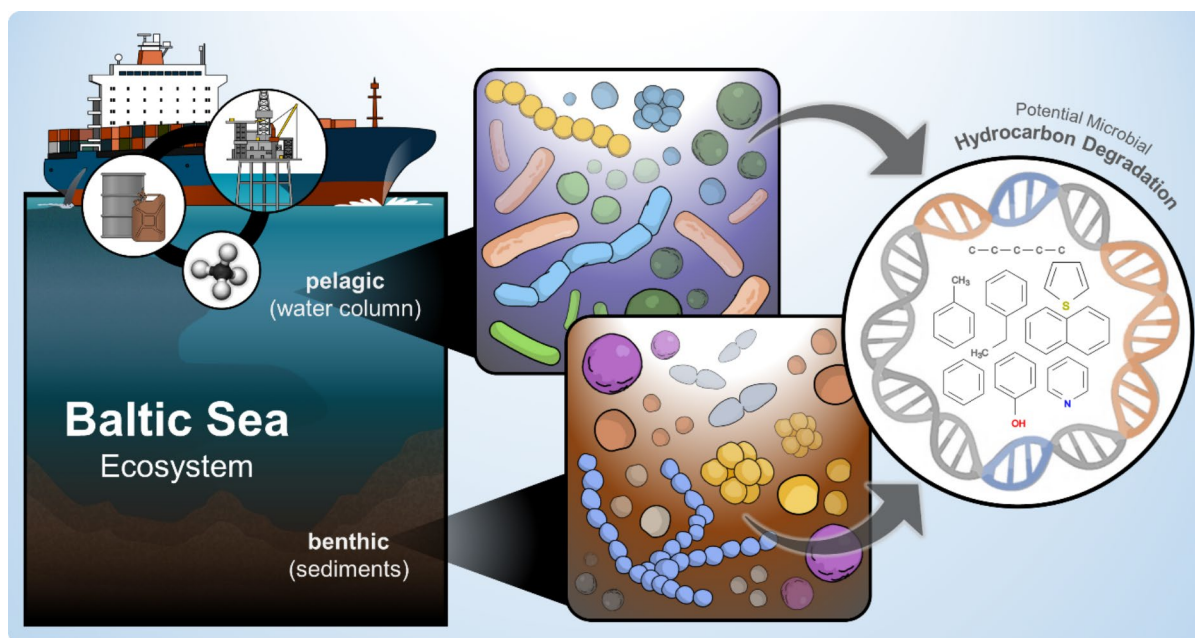
joeselle.serrana@aces.su.se

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Graphical Abstract



## Background

Petroleum hydrocarbons are one of the major organic pollutants in marine ecosystems [1, 2]. These compounds disrupt ecological diversity and impair ecosystem functions, posing substantial threats to global health and sustainability [3]. For example, the influx of hydrocarbons into marine ecosystems can have devastating effects on marine life mortality [4], causing a chain reaction of negative impacts on ecosystem functions, e.g., nutrient cycling, carbon sequestration, and water purification [5, 6]. While some hydrocarbons come from natural sources (e.g., oil seeps and the breakdown of organic matter) [7–9], more concerning issues arise from anthropogenic activities. This includes sudden and significant releases from oil spills, high-volume industrial waste discharges, atmospheric deposition, and extensive marine oil and gas exploration activities [2, 9, 10], estimated to introduce 8.3 million tons into marine ecosystems each year [11].

For instance, the Baltic Sea, one of the largest brackish water bodies, is also among the busiest waterways in the world [12, 13]. Hazardous substances entering the Baltic catchment at elevated levels due to anthropogenic activities, including agriculture, municipal sewage, and industrial discharges [14], remain among the most widespread and impactful pressures on the Baltic Sea [15, 16]. The recent HELCOM holistic assessment for the 2016–2021 period (HOLAS 3) [17] reported a continued overall

decrease in both the frequency and volume of oil spills in the Baltic Sea, with accidental spills near port areas from cargo vessels or passenger ships rarely causing pollution. Despite this, recent reports have documented negative impacts of oil spills and other petroleum products on the Baltic Sea's marine and coastal ecosystems [12, 18, 19]. This is because the Baltic Sea basin remains vulnerable to the harmful effects of petroleum hydrocarbons and other hazardous substances [19, 20]. These compounds can accumulate in sediments and marine organisms, leading to long-term ecological impacts, e.g., toxicity to marine life and disruption of local ecosystems [21, 22]. Furthermore, potential future contamination continues to pose a risk due to various factors, e.g., increased maritime traffic and transportation, climate change, and the possibility of large-scale accidental spills [17]. For example, tankers traveling the Baltic Sea route have increased in both size and length, accompanied by growing shadow fleet activities [23–25]. This puts the area particularly at risk of oil spills and other maritime incidents. Therefore, while the situation has improved, ongoing monitoring and preventive measures remain critical to protect the Baltic Sea ecosystem.

The removal of petroleum hydrocarbons from the environment primarily depends on native microorganisms that degrade them by utilizing hydrocarbons as a source of carbon and energy [26–28]. The breakdown of these

contaminants is generally facilitated by taxonomically diverse microorganisms that display a wide range of functions and substrates involved in hydrocarbon degradation [27–30]. Microbial degradation can occur under both aerobic and anaerobic conditions, with aerobic pathways employing oxygen to activate hydrocarbons and anaerobic pathways utilizing alternative electron acceptors, e.g., sulfate or nitrate, in oxygen-deprived environments [31]. Microbial degradation of hydrocarbons is a crucial ecosystem service vital for maintaining ecosystem health [27, 32]. Assessing the natural microbial degradation of petroleum hydrocarbons is essential for understanding their environmental impacts on marine ecosystems and their inhabitants [31, 33]. Previous studies have evaluated the ability of microorganisms to degrade hydrocarbons across various ecosystems [34–37].

In the case of the Baltic Sea, a handful of studies have assessed the natural hydrocarbon degradation potential of the region [38–44]. For example, Miettinen et al. [38] investigated the hydrocarbon degradation potential of bacterial, archaeal, and fungal microbiomes, comparing long-term oil-exposed and less-exposed coastal sites. Using targeted gene identification (*alkB* and PAH ring hydroxylating deoxygenase) and taxonomic profiling via amplicon sequencing, they found similar microbial diversity and degradation potential at polluted and pristine sites in the Baltic Sea. However, our understanding of marine ecosystems remains limited, as most previous studies have primarily focused on a few substrates or cultured microorganisms [38, 39] or have only focused on specific regions of the Baltic Sea [40–43]. Advantageously, the application of advanced molecular biology approaches, e.g., metagenomics, allows for broader and more detailed insights into environmental microbiomes.

In this study, we profiled hydrocarbon-degrading microorganisms in both benthic and pelagic ecosystems of the Baltic Sea and characterized the associated genes and metabolic functions involved in the degradation of various hydrocarbon compounds. To the best of our knowledge, this is the first thorough profiling of the functional potentials of the Baltic Sea concerning hydrocarbon degradation. We utilized publicly available metagenomic datasets from environmental microbiomes sampled in the benthic [45, 46] and pelagic [47, 48] ecosystems across the Baltic Sea. Through this analysis, we identified potential microbial hydrocarbon degraders from both environments and screened for genes associated with the degradation of diverse hydrocarbon substrates. We also assessed the potential influence of environmental factors, i.e., water depth, salinity, and temperature gradients, on the degradation capacities of these environmental microbiomes. Additionally, we hypothesize that oil spills in the area have influenced the adaptive

capacity and diversity of hydrocarbon-degrading microorganisms exposed to these contaminants. Hence, we investigated the potential link between the recorded annual average total oil spills in each subbasin of the Baltic Sea and their microbiomes' ability to degrade hydrocarbons. Understanding the degradation potential of environmental microbiomes exposed to hydrocarbon contamination is essential for refining estimates of pollutant persistence and predicting the fate of petroleum hydrocarbons in marine and coastal environments. Our findings contribute to a deeper understanding of the functional resilience of Baltic Sea microbiomes and provide a scientific basis for developing improved management and mitigation strategies that address hydrocarbon pollution in aquatic ecosystems.

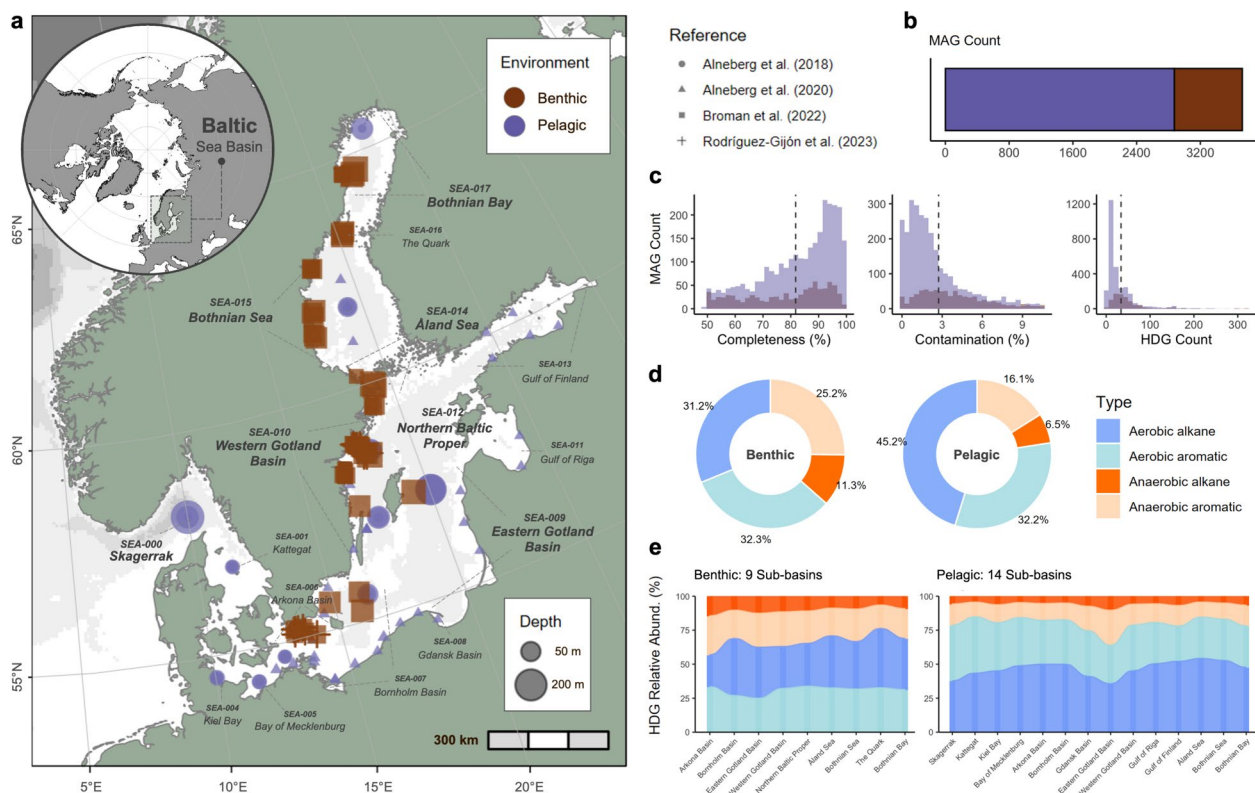
## Methods

### Compiling Baltic Sea metagenomes

A total of 203 Baltic Sea metagenomes from both benthic sediments ( $n=108$ ; PRJEB41834) [45, 46] and water column samples ( $n=95$ ; PRJEB22997 and PRJEB34883) [47, 48] were compiled from public metagenomic data across four published studies (Fig. 1a). The benthic samples were collected from 2016 to 2019, while the pelagic samples were from a 2011 to 2015 collection. Sequence information and sample coordinates can be found in Supplementary Table S1. For further details on site description, collection dates, and sequencing information, please refer to the respective publications of the public datasets [45–48]. The environmental parameters, i.e., depth (m), salinity (PSU), and temperature (°C), compiled by Rodríguez-Gijón et al. [46], were utilized in this study. Following the HELCOM subdivisions of the Baltic Sea nomenclature, the metagenomes were assigned to Baltic Sea subbasins based on their location [49]. Two hundred samples were distributed across fifteen subbasins, while the remaining three pelagic samples from Alneberg et al. [48] collected from Skagerrak basin that were not included in the HELCOM subbasins nomenclature were referred to as SEA-000 in this study (Supplementary Fig. S1).

### Metagenomic assembly, binning, and phylogenomic analysis

The raw paired-end sequences of the 203 metagenomic datasets were adapter-trimmed and quality-filtered using fastp v0.23.4 [50], followed by the removal of human and PhiX spike-in reads with Kraken2 v2.1.2 [51]. The quality-filtered reads were then error-corrected using bbcm version 38.61b from BBTools [52], with parameters set to a minimum count of 2 and a high-count fraction of 0.6. Individual assemblies of the 203 quality-filtered metagenomic datasets were conducted with MEGAHIT v1.2.9



**Fig. 1** Baltic Sea metagenomics data and metagenome-assembled genome (MAG) information. **a** Location and sample environment of the metagenomics data compiled from Baltic Sea microbiome studies (i.e., Alneberg et al., 2018, 2020; Broman et al., 2022; Rodríguez-Gijón et al., 2023). **b** MAG counts and **c** histograms illustrating the distribution of MAG quality based on % completeness and % contamination, and the predicted hydrocarbon degradation gene (HDG) counts based on reads per kilobase million (RPKM). The relative abundance of HDGs by type based on **d** environmental samples and **e** across the Baltic subbasins

[53]. Genes were then predicted on contigs using Prodigal v2.6.3 [54], employing the -p meta option. The read counts for each predicted gene were estimated by mapping the quality-filtered paired-end reads to the contigs using Minimap2 v2.24-r1175 [55] and summarizing the read abundances into count tables with featureCounts v2.0.3. [56]. Read processing counts at each step are listed in Supplementary Table S2.

Binning was performed using the single\_easy\_bin module of SemiBin v2.1 [57]. Binning parameters were set to default with the environment flag set to global. The bins were then evaluated for contamination and completeness using CheckM v1.1.3 [58] with lineage\_wf and default parameters. A total of 10,017 bins were recovered from the individual assemblies. From this, 13.6% (1,318) were classified as high-quality (completeness of >90% and contamination of <5%), while 37.2% (3,726) were medium-to-high quality (completeness of >50% and contamination <10%) (Supplementary Table S3 and Fig. 1a,b). The 193 samples with medium-to-high quality bins were retained in the downstream

analysis, as their inclusion broadened the dataset, enabling a more extensive exploration of microbial capabilities and referred to as metagenome-assembled genomes (MAGs). The average guanine and cytosine (GC) base content was 52%, with an average contig (N50) length of 19,833 and a predicted gene count of 2230, respectively. The average genome size was 2.20 million base pairs (Mbp), with the largest measuring 8.89 Mbp (ERR2206771.bin\_3) (Supplementary Fig. S2).

Taxonomic assignment was performed on the MAGs using GTDB-Tk v2.4.0 with R220 [59] through the classify\_wf function. The MAG quality parameters and taxonomic information are presented in Supplementary Table S3. The protein sequence alignments produced from GTDB-Tk were subsequently trimmed using trimAl v1.4.rev15 [60] with the heuristic “-automated1” method and were used as inputs to construct phylogenomic trees with IQTree2 multicore v2.2.0.3 COVID-edition [61]. The resulting trees were visualized and annotated using the Interactive Tree Of Life (iTOL) v4.4.2 [62].

### Screening of hydrocarbon degradation genes and their substrates

Hydrocarbon degradation genes were screened using the Calgary approach to ANnotating HYDrocarbon degradation genes (CANT-HYD) [63]. Genes involved in hydrocarbon degradation were identified using Hidden Markov Model (HMM) profiles from the CANT-HYD database, with a noise cut-off E-value of 0.01 for the binned metagenomes. This approach annotates 37 key marker genes associated with anaerobic and aerobic degradation pathways for both aliphatic and aromatic hydrocarbons, based on experimental and in silico-inferred functional enzymes. The marker genes were categorized into the following groups: 4 types (i.e., anaerobic aromatic, anaerobic alkane, aerobic alkane, and aerobic aromatic), 5 enzymatic groups (i.e., anaerobic carboxylase, anaerobic hydroxylation, monooxygenase, dioxygenase, and anaerobic fumarate addition), 18 substrates (e.g., C5-C13 alkane, polyaromatics, toluene), and 23 enzyme categories (e.g., benzene carboxylase, p-cymene dehydrogenase, propane 2-monooxygenase), as listed in Supplementary Table S4. Gene abundances were expressed as reads per kilobase million (RPKM) to normalize gene coverage values, correcting for differences in sample sequencing depth and gene length [64]. The RPKM counts for each gene per sample are provided in Supplementary Table S5.

### Statistical analysis

All visualizations and statistical analyses were conducted using R v4.4.3 [65]. A Pearson correlation analysis was used to assess the relationship between genome size and the number of annotated hydrocarbon degradation genes (HDGs). The microeco v1.13.0 package [66] was utilized to calculate both alpha (i.e., Observed, Chao1 richness, Shannon diversity, and Pielou evenness) and beta diversities (i.e., Jaccard dissimilarity index) of the microbial community and HDG datasets. A *t*-test or a one-way analysis of variance (ANOVA) with Duncan's multiple comparison tests was employed to examine significant differences in alpha diversity across categories. Principal coordinate analysis (PCoA) was performed to visualize the Jaccard dissimilarity matrices, and a non-parametric multivariate analysis of variance (PERMANOVA) was utilized to infer significant differences in composition for both the taxonomic and HDG datasets. A Pearson correlation analysis was conducted to investigate the relationships between species and HDG diversity indices, and to assess the potential connection between microbial diversity and HDG composition. A random forest analysis coupled with a differential abundance test [67] was performed on the HDG dataset to identify differentially abundant HDGs between environments. The annual average total oil spill volumes (m<sup>3</sup>) during the assessment

period of 2016–2021, as reported by HELCOM [17] (Supplementary Table S6), were used to correlate oil spills in each Baltic Sea subbasin with their hydrocarbon degradation gene profile. Spearman's correlation analysis was conducted to assess the relationships between environmental factors and HDGs using the microViz v0.12.1 package [68]. Skagerrak samples were excluded from this analysis since the area is not covered in the oil spill reports recorded by the commission.

## Results

### Microbial hydrocarbon degraders from the benthic and pelagic ecosystems

We recovered 3726 medium-to-high quality MAGs, with 854 MAGs identified in benthic samples, and 2872 MAGs were found in pelagic samples (Fig. 1b). The average genome size of the MAGs was 2.20 Mbp, a completeness of 81.8%, and 2.7% contamination (Fig. 1c). The average genome size of benthic MAGs was 2.73 Mbp, while the pelagic MAGs had 2.04 Mbp (Supplementary Table S7). The difference in average genome sizes between the environments may be linked to the distinct evolutionary pressures and ecological roles these microorganisms fulfill within their respective habitats.

Using a validated HMM profiling approach, genes encoding hydrocarbon-degrading enzymes were identified in MAGs. All 37 key hydrocarbon degradation genes (HDGs) were identified from the compiled Baltic Sea metagenomic samples. Multiple copies of HDGs can be found within a single MAG, with an average of 34 HDG reads per kilobase million (RPKM) (Fig. 1c). Except for one bacterial MAG (ERR2206774.bin\_159) and one archaeal MAG (ERR2206790.bin\_40), most of the MAGs had at least one HDG (Supplementary Table S3), with a total of 213 MAGs having >100 multiple HDGs annotated (Supplementary Table S3 and Fig. 1c). For example, one bacterial MAG (ERR5010709.bin\_8) had 320 RPKM, which had the highest HDG annotation. Benthic MAGs had higher HDG annotations, with an average of 49 RPKM, compared to 29 in pelagic MAGs. We also found a statistically significant but moderately positive linear relationship between genome size and HDG count for all MAGs ( $r=0.43$ ,  $t=29.39$ ,  $df=3724$ ,  $p<0.001$ ) (Supplementary Fig. S3). By environment, the genome size of benthic MAGs showed a stronger relationship with HDG count ( $r=0.59$ ,  $t=21.424$ ,  $df=852$ ,  $p\text{-value}<0.001$ ) compared to the pelagic samples ( $r=0.35$ ,  $t=19.852$ ,  $df=2870$ ,  $p<0.001$ ).

Based on four metabolic categories, the hydrocarbon degradation potential of benthic and pelagic MAGs differed across the assessed pathways. HDGs under the aerobic alkane metabolism category were notably higher in the pelagic MAGs (45.2%) compared to benthic MAGs

(31.2%) (Fig. 1d). Conversely, HDGs involved in anaerobic alkane metabolism were more pronounced in the benthic MAGs (11.3%) than in the pelagic MAGs (6.5%). For the aromatic degradation pathways, HDGs involved in anaerobic metabolism were higher in benthic MAGs (25.2%) compared to pelagic (16.1%), while HDGs for aerobic metabolism exhibited similar levels, with 32.3% in benthic and 32.2% in pelagic MAGs. Moreover, the relative abundance of the four types for each subbasin is shown in Fig. 1e. Based on the frequency of HDG annotation, each subbasin exhibited varying levels of abundance (Supplementary Fig. S4), with the Western Gotland Basin samples (SEA-010) having the highest number of identified HDGs. In support of this, the observed and HDG Chao1 richness and Shannon diversity of each subbasin in both environments were significantly different ( $p < 0.01$ ) (Supplementary Fig. S5). In contrast, Pielou's evenness index was not significantly different among subbasins for both environments (benthic,  $p = 0.78$ ; pelagic,  $p = 0.08$ ). According to the HMM annotations, 25 hydrocarbon degradation marker gene annotations were identified across all Baltic Sea subbasins (Supplementary Fig. S6 and Table S8).

Of the MAGs with hydrocarbon degradation potential, 79 were archaeal genomes (Supplementary Fig. S7), while 3645 were classified as bacteria according to the GTDB-tk classification (Fig. 2a). All MAGs had classifications at least up to the order level. Three MAGs had unclassified families, and 146 MAGs lacked classified genera. In total, 2526 MAGs (67.8%) had species-level identifications (Fig. 2b). The MAGs represented a diverse array of prokaryotes, encompassing 35 phyla (5 archaeal and 30 bacterial), 243 families (9 archaeal and 234 bacterial), 467 genera (15 archaeal and 452 bacterial), and 657 species (22 archaeal and 634 bacterial, including unclassified species). For the archaeal annotations, the MAGs showed that 67.9% (53) were Thermoproteota, followed by Thermoplasmatota (26.9%), Nanoarchaeota (2.6%), Micrarchaeota (1.3%), and Halobacteriota (1.3%).

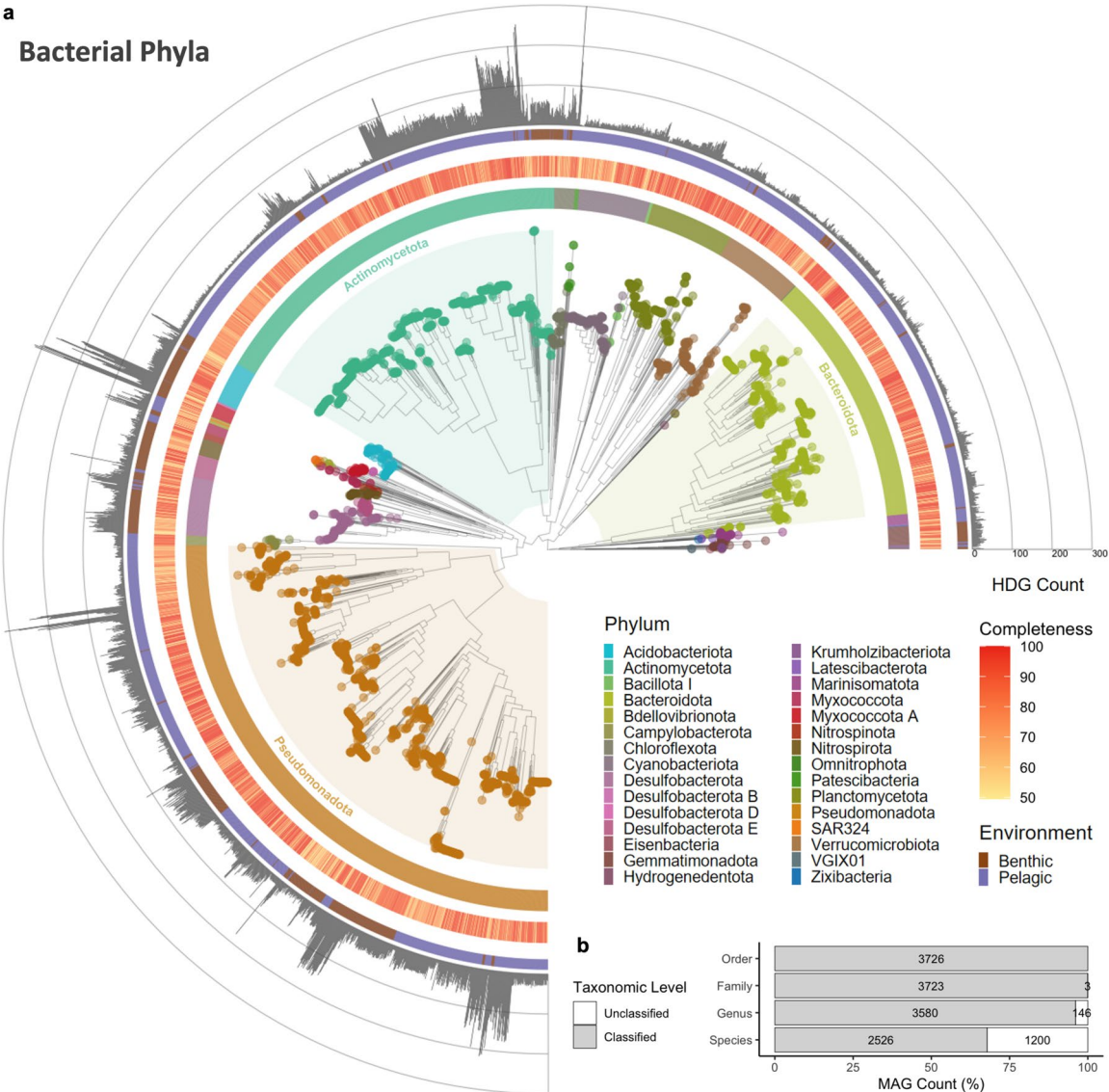
In the bacterial phyla, MAGs assigned to Pseudomonadota (33.6%) were the most prevalent, followed by Actinomycetota (22.3%), Bacteroidota (14.9%), Planctomycetota (5%), and Desulfobacterota (4.9%). The remaining bacterial phyla accounted for less than 5% of the MAGs. The relative abundance of the top 10 phyla across each environment is presented in Fig. 3a. Pseudomonadota predominated most samples, except for the benthic Eastern Gotland Basin, where the MAGs were mostly assigned to Gemmatimonadota. The relative abundance of Desulfobacterota in benthic samples decreased from the Arkona and Bornholm Basins to the more freshwater subbasins, e.g., Åland Sea, Bothnian Sea, and Bothnian Bay. Conversely, Actinomycetota exhibited an increasing

trend from the more marine subbasins, e.g., Skagerrak and Kattegat, to freshwater sites.

At the species level, only six species (0.9%) were found to be present in both environments (Fig. 3b). Including MAGs with unclassified species annotations, the benthic samples comprised 147 species, whereas the pelagic samples included 504 species. The top 10 most abundant species based on MAG relative abundance for each environment are depicted in Supplementary Fig. S8. The species with the most annotations was *UBA1847 sp022866645* (Pseudomonadota, Woeseiaceae) with a total of 61 MAGs (1.64%), found exclusively in benthic samples, followed by *D2472 sp002358345* (Pseudomonadota, D2472) with 59 MAGs (1.58%), *MAG-120802 sp030730425* (Actinomycetota, Nanopelagicaceae) with 53 MAGs (1.42%), and *UBA9145 sp001438145* (Pseudomonadota, Pseudohongiellaceae) with 52 MAGs (1.40%), which were only found in pelagic samples. The counts and total relative abundance of all species are listed in Supplementary Table S8, and a visualization of the shared and unique species per sub-basin is presented in Supplementary Fig. S9. Three MAGs identified as *JAJDAM01 sp.* (Myxococcota A, SMWR01) (i.e., ERR5010709.bin\_8, ERR5010720.bin\_14, and ERR5010722.bin\_46), and two *CALGIA01 sp937916005* (Pseudomonadota, MPNO01) had multiple HDG annotations > 300 RPKM (Supplementary Table S3).

The species composition of the two environments was distinctly separated from one another based on Jaccard dissimilarity (Fig. 3c). PERMANOVA revealed significant effects of both the environment and subbasin on species composition. The environment explained 11.87% of the variance ( $R^2 = 0.12$ ,  $F = 29.36$ ,  $p = 0.001$ ), while the subbasin accounted for 16.94% ( $R^2 = 0.17$ ,  $F = 2.79$ ,  $p = 0.001$ ). The analysis of similarities (ANOSIM) confirmed strong differentiation between the pelagic and benthic samples ( $R = 0.69$ ,  $p = 0.001$ ). The homogeneity of multivariate dispersion (betadisper) test also indicated no significant differences in dispersion among the groups ( $F = 2.35$ ,  $p = 0.137$ ), supporting the assumption of equal variances across the groups. Pairwise *t*-tests further revealed significant differences in group distances between benthic and pelagic communities ( $p_{\text{adj}} < 0.001$ ), suggesting a clear separation between these groups (Supplementary Table S9).

Additionally, the alpha diversity comparison of the two environments further supports the differentiation between the benthic and pelagic samples (Fig. 4a). The observed species ( $30 \pm 1$  vs.  $9 \pm 7$ ), Chao1 richness ( $30 \pm 1$  vs.  $9 \pm 7$ ), and Shannon diversity ( $2.9 \pm 0.5$  vs.  $1.6 \pm 0.8$ ) in the pelagic environment were significantly higher than in the benthic environment ( $p_{\text{adj}} < 0.001$ ). Moreover, Pielou's evenness metric ( $0.99994 \pm 0.0003$



**Fig. 2** Taxonomic assignment of MAGs. **a** Maximum-likelihood phylogenomic tree of bacterial MAGs recovered from Baltic Sea metagenomic samples. **b** Percentage of MAGs assigned at various taxonomic ranks based on GTDB-tk classification

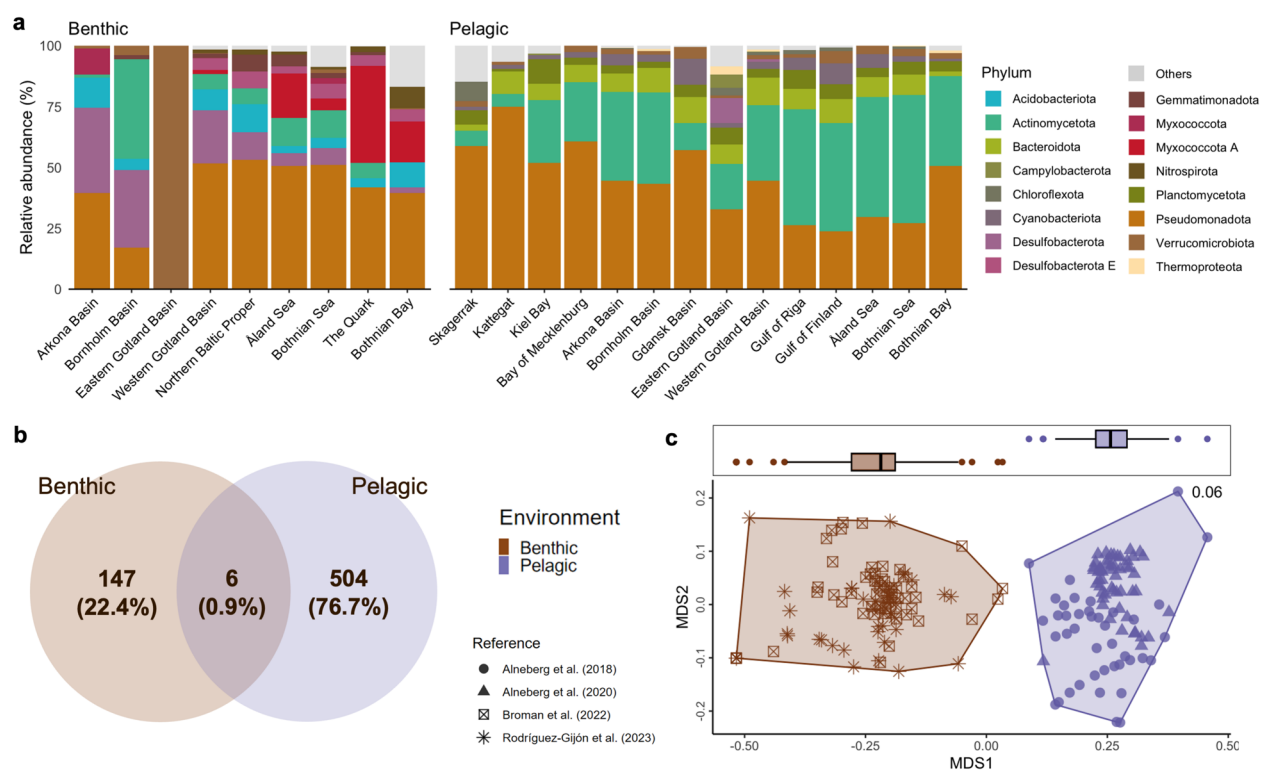
vs.  $0.99997 \pm 0.0002$ ) demonstrated a significant difference in the evenness of the microbial taxa between environments.

**Genes and metabolic pathways associated with the degradation of various hydrocarbon substrates**

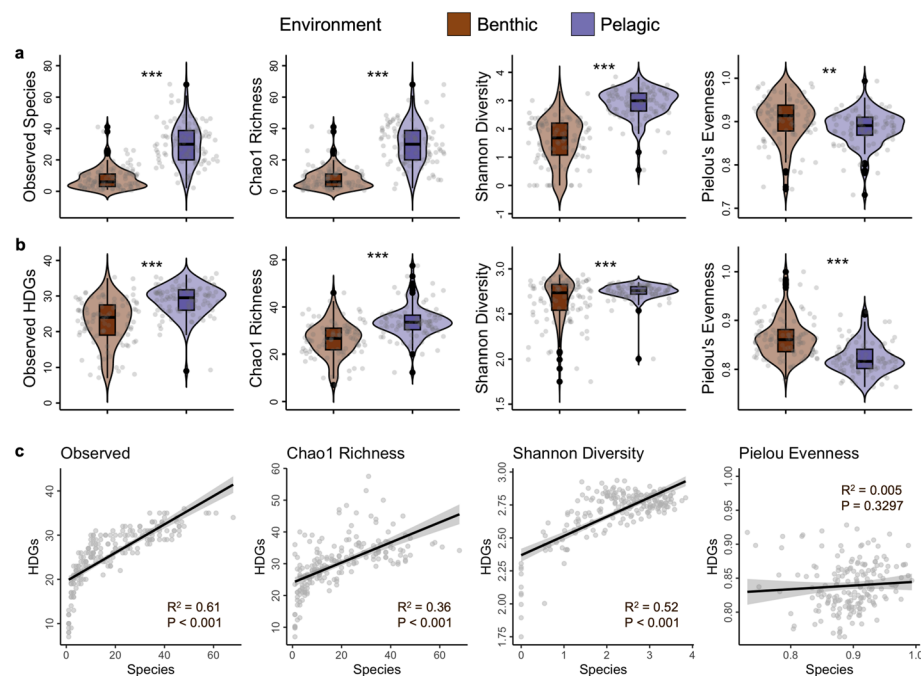
The alpha diversity estimates of the HDGs displayed similar trends to species alpha diversity, with pelagic samples showing higher diversity and richness values compared to benthic samples, and a significantly greater evenness in benthic samples (Fig. 4b). Interestingly, the correlation between their alpha diversity values (Fig. 4c) revealed a strong positive relationship between observed

species and HDGs ( $R^2=0.61$ ,  $p<0.001$ ), indicating that as HDGs increased, the number of observed species also increased significantly. Similarly, the Shannon diversity index showed a moderate positive correlation with HDGs ( $R^2=0.515$ ,  $p<0.001$ ), suggesting a link between increasing HDGs and species richness and evenness. In contrast, the species Chao1 richness exhibited a weaker correlation with HDGs ( $R^2=0.358$ ,  $p<0.001$ ). Notably, Pielou's evenness demonstrated a negligible correlation with HDGs ( $R^2=0.005$ ,  $p<0.001$ ), indicating that the species had minimal influence on the observed HDG evenness.

The top ten most abundant HDGs found in the benthic samples included the dibenzothiophene desulfurization



**Fig. 3** Taxonomic composition of the benthic and pelagic microbial communities. **a** Relative abundance of the top microbial phyla in each sub-basin and environment. **b** Shared and unique species per environment. **c** Non-metric multidimensional scale (NMDS) plot of species composition based on Jaccard dissimilarity

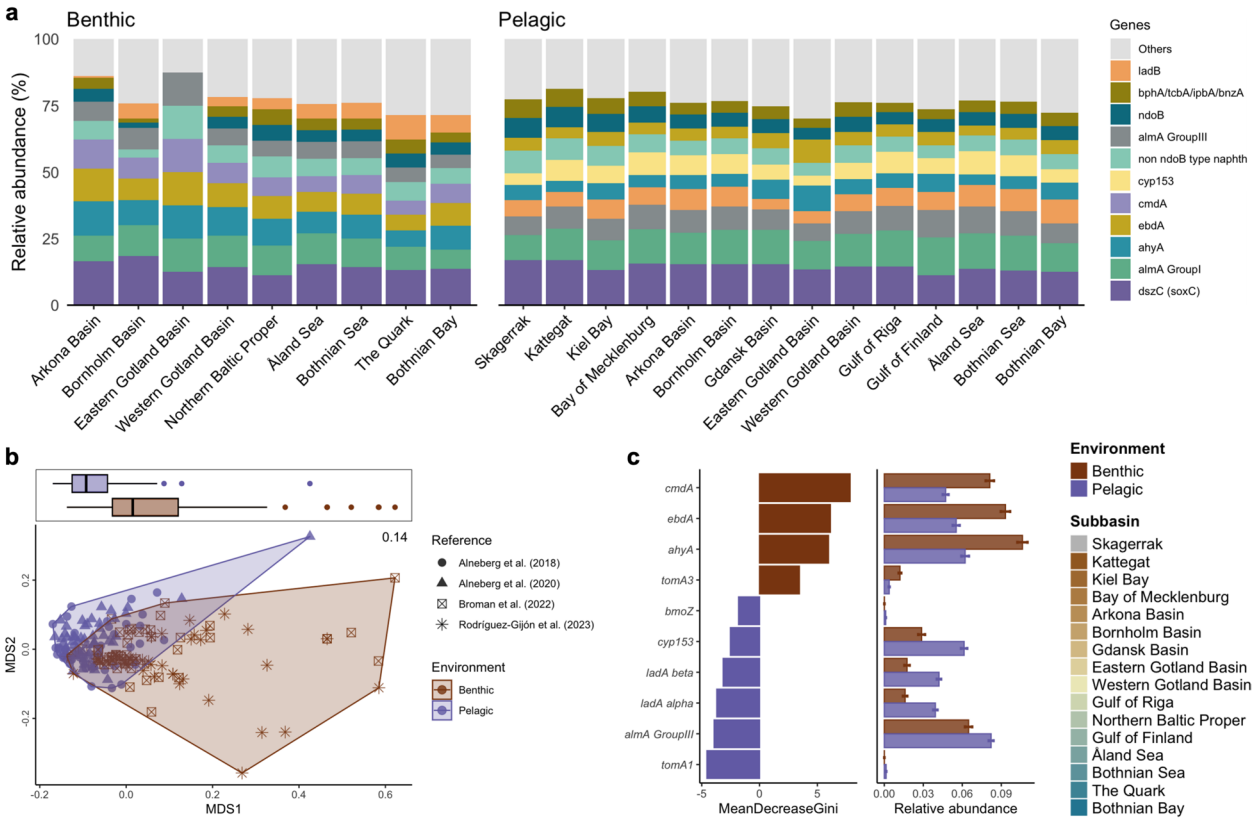


**Fig. 4** Taxonomic and hydrocarbon degradation gene (HDG) diversity. Alpha diversity estimates of the **a** species and **b** HDGs. "\*\*\*\*" indicates differences between groups tested by *t*-test at  $p < 0.001$ , "\*\*\*\*" at  $p < 0.01$ . **c** Linear regression analysis between the species and HDG alpha diversity estimates

enzyme C [*dszC* (*soxC*)], the flavin-binding monooxygenase (*alma* Group I and III), the putative alkane C2 methylene hydroxylase (*ahyA*), ethylbenzene dehydrogenase (*ebdA*), p-cymene dehydrogenase (*cmdA*), naphthalene dioxygenases (*non* *ndoB* type and *ndoB*), the monoaromatic dioxygenase (*bphA/tcbA/ipbA/bnzA*), and the long-chain alkane monooxygenase (*ladB*) (Fig. 5a). Similarly, except for the *cmdA* gene, these HDGs were also abundant in the pelagic samples with the addition of the cytochrome P450 alkane hydroxylase (*cyp153*). The occurrence of each HDG per phylum is presented in Supplementary Fig. S10. Pseudomonadota had the most prevalent and abundant HDGs, followed by Myxococcota A, Chloroflexota, Actinomycetota, and Desulfobacterota.

For the beta diversity of HDG composition, Jaccard dissimilarity visualized using an NMDS plot revealed differences between the benthic and pelagic samples (Fig. 5b). PERMANOVA demonstrated significant effects from both environmental factors and subbasin grouping on HDG composition. The environment accounted for 15.13% of the variance ( $R^2=0.15$ ,  $F=38.17$ ,  $p=0.001$ ),

while subbasin explained 15.12% ( $R^2=0.1512$ ,  $F=2.54$ ,  $p=0.001$ ). These findings were further supported by ANOSIM, which indicated moderate differentiation among the environmental samples ( $R=0.24$ ,  $p=0.001$ ). A betadisper test also showed significant differences in dispersion among groups ( $F=6.73$ ,  $p=0.01$ ), with pairwise comparisons indicating differences between benthic and pelagic samples ( $p=0.01$ ). Finally, the group distance comparison revealed highly significant differences in HDG composition between benthic and pelagic samples ( $p_{adj}<0.001$ ), emphasizing their distinct separation. These results highlight the crucial roles of the environment in shaping HDG compositional patterns (Supplementary Table S9). Random forest analysis identified 34 hydrocarbon degradation marker gene annotations significantly associated with specific environments. (Fig. 5c, Supplementary Fig. S11 and Table S10). In the benthic samples, *cmdA*, *ebdA*, and *ahyA* exhibited the highest importance scores (Mean Decrease Gini values) of 7.85, 5.98, and 5.89, respectively. In the pelagic samples, *tomA1*, *almA* Group III, and *ladA* alpha showed



**Fig. 5** Hydrocarbon degradation gene (HDG) composition. **a** Relative abundance of the top HDGs in each sub-basin and environment. **b** Non-metric multidimensional scaling (NMDS) plot of HDG composition based on Jaccard dissimilarity. **c** Random forest classification of the relative abundance of HDGs across environments: (left) HDGs with Mean Decrease Gini > 2, presented in descending order of importance; see Supplementary Fig. S11 for the plot of all significantly differential genes; (right) Relative abundance of HDGs enriched in each environment. The bars represent the mean relative abundance, and error bars indicate the standard error

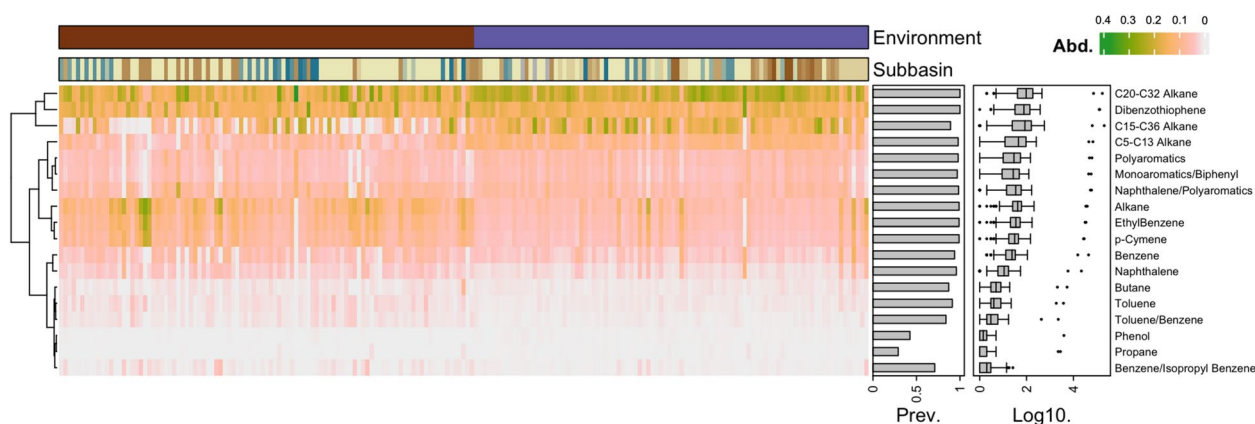
high influence with scores of 4.61, 4.37, and 3.51. These differences, supported by strong statistical significance ( $P_{adj} < 0.001$ ), emphasize niche specialization and its impact on shaping microbial dynamics and hydrocarbon degradation functions in each environment.

### Substrate preferences of potential microbial hydrocarbon degraders

The annotated hydrocarbon degradation genes were grouped into 18 categories of hydrocarbon substrates based on the CANT-HYD database [63]. Figure 6 and Supplementary Table S8 show the distribution of substrates associated with the HDGs identified in benthic and pelagic samples. Among the substrates, genes linked to C20-C32 alkanes display the highest abundance, contributing 18.67% of the total RPKM HDG abundance, with a significant presence in both benthic (6548) and pelagic (16,927) zones. This is followed by genes related to C15-C36 alkanes (14.86%) and dibenzothiophene degradation (14.24%), which are also predominantly found in pelagic environments. Genes involved in the degradation of short-chain alkanes (C5-C13) make up 8.80% of the total, while those associated with generic alkane degradation account for 7.09%. Other notable contributions come from genes related to naphthalene/polyaromatics (6.38%), ethylbenzene (6.04%), and polyaromatics (5.59%). Minor contributions are observed for compounds like benzene (3.63%), naphthalene (1.88%), and toluene (0.78%), with trace levels of genes associated with propane (0.07%) and phenol degradation (0.11%). We emphasize the prevalence of genes associated with the degradation of long-chain alkanes and certain aromatic compounds in Baltic Sea environments, with pelagic zones typically exhibiting higher gene abundances than benthic zones.

We also observed interesting patterns based on the relative abundance of the substrates degraded by the HDGs across the Baltic Sea. Although the HDGs associated with the degradation of C20-C32 alkanes were the most abundant substrates in each sub-basin, the Gulf of Finland stands out for having the highest relative abundance (24.2% of the HDGs identified in the sub-basin), suggesting elevated activity for degrading these substrates in this area, followed by the Gulf of Riga (23%) and the Bay of Mecklenburg (22%). The subbasins Skagerrak (8%) and Kattegat (8%) exhibited relatively higher relative abundances of polyaromatics compared to other areas (Supplementary Table S10).

Using a random forest analysis to assess the differential abundance of genes between the environments (Supplementary Fig. S12 and Table S10), genes involved in the degradation of the substrates p-Cymene, alkanes, and ethylbenzene exhibited the highest importance values of 12.21, 8.63, and 8.60, respectively, which were differentially abundant in benthic samples ( $p < 0.0001$ ). Other compounds, i.e., toluene/benzene, naphthalene, and benzene, also contributed to the classification, albeit to a lesser extent (Mean Decrease Gini values  $> 3$ ). In the pelagic environment, the genes associated with the degradation of long-chain (i.e., C15-C36 and C20-C32) and short-chain alkanes (i.e., C5-C13) were the most differentially abundant, with values of 5.04, 4.33, and 3.52 ( $p < 0.001$ ). Additionally, phenol and polyaromatics demonstrated notable importance, while monoaromatics/biphenyl and propane also contributed to the random forest classification model.



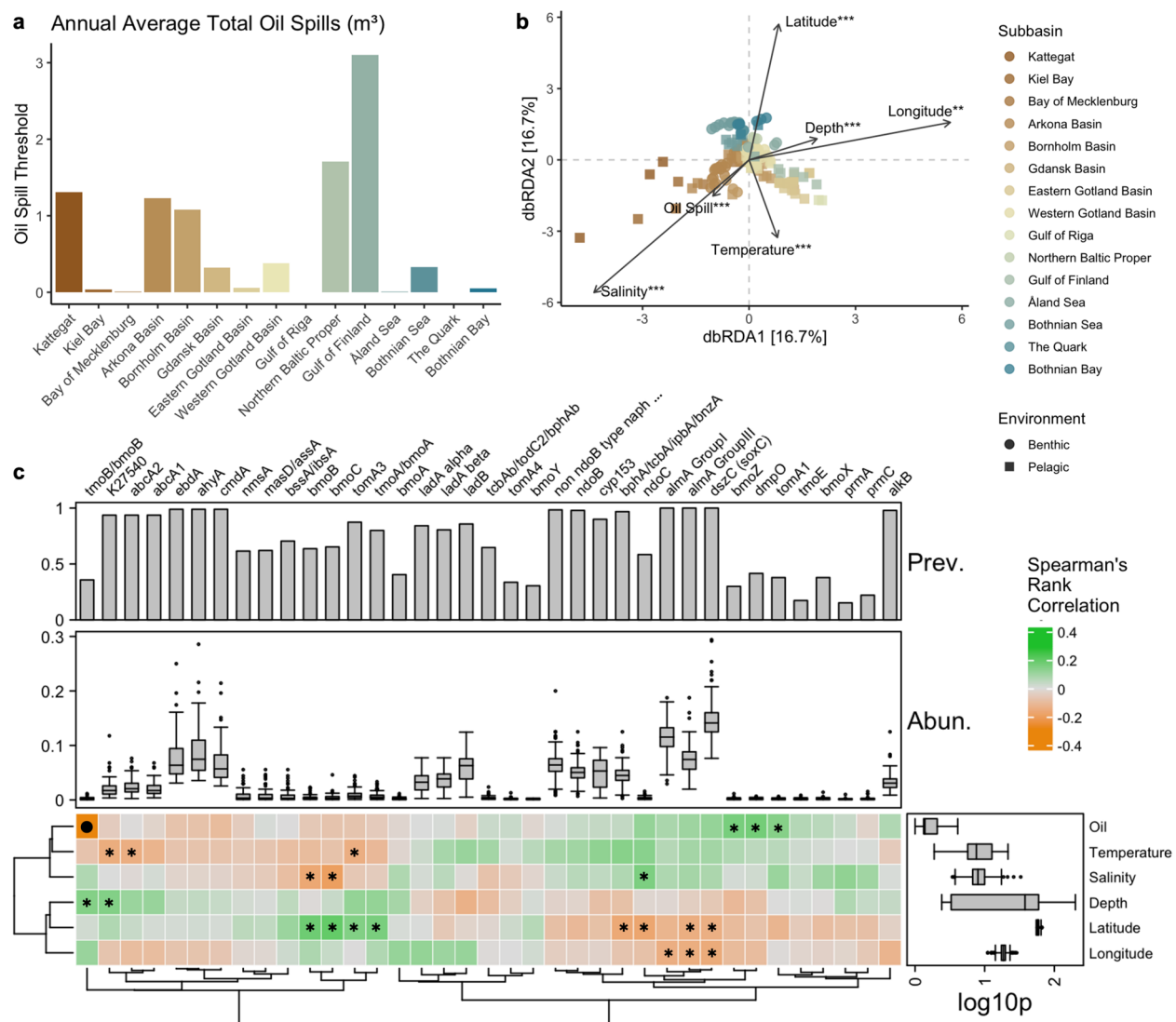
**Fig. 6** Potential substrates of the microbial hydrocarbon degraders. Heatmap of the relative abundance of hydrocarbon substrates associated with the degradation genes. Hierarchical clustering analysis was performed using Ward's method and Euclidean distance metric. Refer to Fig. 5 for the color legend of the Environment and Subbasin groupings

**Correlations among hydrocarbon degradation potential, environmental variables, and oil spill volume**

The annual average of total oil spills across the subbasins during the 2016–2021 assessment period varied significantly (Supplementary Table S1 and Fig. 7a). The Gulf of Finland recorded the highest average spill volume at 3.10 m<sup>3</sup>, followed by the Northern Baltic Proper at 1.71 m<sup>3</sup> and the Kattegat at 1.31 m<sup>3</sup>. Moderate spill levels were observed in subbasins such as the Arkona Basin (1.23 m<sup>3</sup>), Bornholm Basin (1.08 m<sup>3</sup>), and Western Gotland

Basin (0.38 m<sup>3</sup>). The Gulf of Riga and the Quark reported negligible or no recorded oil spills [15]. In addition to depth, salinity, temperature, latitude, and longitude (Supplementary Table S1), we evaluated the relationship between the HDG composition and annual average total volume of oil spills.

Distance-based redundancy analysis (dbRDA) revealed significant relationships between environmental variables and multivariate response patterns. Depth was the most influential factor, accounting for 80.27% of the variance



**Fig. 7** Environmental and oil spill association of the hydrocarbon degradation gene (HDG) profile. **a** The annual average total volume of oil spills (m<sup>3</sup>) for each subbasin (excluding Skagerrak) from 2016 to 2021, as reported by HELCOM (2023b). **b** Distance-based redundancy analysis (dbRDA) ordination plot of the HDG composition based on Jaccard distance and the environmental variables. "\*\*\*\*" indicates a significant  $p < 0.001$ , "\*\*\*\*"  $p < 0.01$ , and "\*\*\*\*"  $p < 0.05$ . **c** Spearman's rank correlation illustrating the relationship between the HDGs and environmental variables. Positive correlations are indicated in green, while negative correlations are shown in orange. The asterisk indicates  $p < 0.05$ , and the filled circle represents FDR-corrected  $p < 0.05$ . "Oil" refers to total oil spills (m<sup>3</sup>). Hierarchical clustering analysis was conducted using Ward's method and the Euclidean distance matrix

( $R^2=0.80$ ,  $p<0.001$ ). Other significant variables included the oil spill threshold ( $R^2=0.63$ ,  $p<0.001$ ), salinity ( $R^2=0.28$ ,  $p<0.001$ ), and latitude ( $R^2=0.18$ ,  $p<0.001$ ). In contrast, longitude and temperature, although significant ( $p<0.005$ ), explained smaller proportions of the variance ( $R^2=0.05$  and  $R^2=0.15$ ) (Fig. 7b and Supplementary Table S11).

Furthermore, Spearman's rank correlation analysis (Fig. 7b and Supplementary Table S12) offered statistically significant insights into the relationships between each environmental variable and the specific HDGs. For example, the correlation between depth and the genes *K27540* (estimate: 0.147,  $p$ -value  $<0.05$ ) and *tmoB/bmoB* (estimate: 0.147,  $p$ -value  $<0.05$ ) indicates weak positive associations, suggesting that the expression levels of these genes may increase slightly with depth. Similarly, latitude showed mixed relationships across various genes. Positive correlations, e.g., with *bmoC* (estimate: 0.199,  $p$ -value  $<0.05$ ), suggest that these gene may increase with latitude, whereas negative correlations, e.g., with *ndoC* (estimate:  $-0.174$ ,  $p$ -value  $<0.01$ ) and *almA Group III* (estimate:  $-0.147$ ,  $p$ -value  $<0.05$ ), imply reduced expression at higher latitudes. The *bmoZ* (estimate: 0.172,  $p$ -value  $<0.05$ ), *dmpO* (estimate: 0.017,  $p$ -value  $<0.05$ ), and *tomA1* (estimate: 0.014,  $p$ -value  $<0.05$ ) genes were positively correlated with the annual average total oil spills. This suggests that the expression of this hydrocarbon degradation gene increases with increasing oil volume. On the other hand, the relationship between oil spill volume and *tmoB/bmoB* is particularly notable, demonstrating a strong negative correlation (estimate:  $-0.291$ ,  $p$ -value  $<0.001$ , FDR: 0.011). While these correlations revealed several potentially meaningful interactions between environmental variables and HDGs, including the strong and significant link between oil spill and *tmoB/bmoB*, most of the other associations were statistically not significant after controlling for multiple tests (i.e., adjusted FDR), which warrants cautious interpretation of the weaker trends.

## Discussion

### Most Baltic Sea MAGs have the potential to degrade hydrocarbons

We identified all 37 key marker genes linked to anaerobic and aerobic degradation pathways for both aliphatic and aromatic hydrocarbons [63]. Eighty percent of the medium-to-high quality MAGs recovered had at least 10 HDG annotations and exhibited multiple copies within a single MAG, with an average HDG count of 34 RPKM (Supplementary Table S3). Additionally, 5.7% of the MAGs contained more than 100 RPKM of the genes related to hydrocarbon degradation (Fig. 1c). Liu et al. [37] also observed multiple copies of HDGs within a

single bacterial strain, suggesting that each copy may serve a unique function (e.g., diverse activation patterns). Moreover, we found a statistically significant, yet moderately positive, linear relationship between genome size and HDG count for all MAGs (Fig. 4c). Benthic MAGs also had higher HDG annotations than pelagic MAGs and showed a stronger association with HDG count compared to the pelagic samples (Supplementary Fig. 3b-c).

Pseudomonadota was the most prevalent phylum and had multiple HDG annotations, followed by Myxococcota A, Actinomycetota, and Desulfobacterota (Figs. 2 and 3a). These phyla were identified as broad-spectrum hydrocarbon-degrading bacteria from the global ocean metagenome [37]. In particular, Pseudomonadota comprises several species renowned for their ability to degrade hydrocarbons, including *Pseudomonas* spp. and *Acinetobacter* spp. [69, 70]. These bacterial species can degrade a wide range of hydrocarbons, e.g., alkanes, aromatic compounds, and polycyclic aromatic hydrocarbons, often using them as their sole carbon source [69]. The degradation process involves specialized enzymes, e.g., oxygenases, which activate and transform most harmful hydrocarbons (e.g., cyclic and aromatic hydrocarbons) into less toxic compounds (e.g., linear alkanes) [69, 71]. This makes Pseudomonadota crucial for bioremediation efforts in environments contaminated by petroleum and other hydrocarbon pollutants [72, 73]. In addition, Myxococcota A, Actinomycetota, and Desulfobacterota were previously identified as hydrocarbon degraders in petroleum-rich marine or brackish sediments [74–77].

The species with the most annotation, *UBA1847 sp022866645* (Pseudomonadota, Woeseiaceae), was found exclusively in benthic samples, followed by *D2472 sp002358345* (Pseudomonadota, D2472), and *MAG-120802 sp030730425* (Actinomycetota, Nanopelagicaceae) (Supplementary Fig. 8). On the other hand, *UBA9145 sp001438145* (Pseudomonadota, Pseudohongiellaceae) was only found in pelagic samples. Previous reports identified *UBA1847* sp. as one of the most abundant bacteria in coastal sediments of the Baltic Sea [78]. This species exhibits a wide range of functions, from facultative chemolithoautotrophy to obligate chemorganoheterotrophy [79]. The MAGs identified as *UBA1847* sp. had an average HDG annotation of  $39 \pm 7$  RPKM, suggesting that the most prevalent bacteria in the benthic ecosystem of the Baltic Sea have the potential to degrade various hydrocarbon substrates. We also recovered MAGs identified as *JAJDAM01* sp. (Myxococcota A, SMWR01) and *CALGIA01* sp. (Pseudomonadota, MPNO01) that had the topmost HDG annotations ( $>300$  RPKM).

Additionally, 66 of the 80 archaeal MAGs we recovered had HDG annotations of more than 10 RPKM (Supplementary Fig. 7). These hydrocarbon-degrading archaea, mostly *Nitrosopumilus* sp. (Thermoproteota, Nitrosopumilaceae), are found in both benthic and pelagic environments. These archaeal species are considered one of the most abundant and ubiquitous microorganisms, primarily found in open oceans and marine sediments [70]. The previous global metagenomic profiling of hydrocarbon-degrading microorganisms found no archaeal genomes [37], whereas assessments in the high Arctic reported three dominant archaeal phyla, i.e., Halobacterota, Euryarchaeota, and Thermoplasmatota. We also identified these phyla in our pelagic samples with low representation. Archaea have fewer genes for hydrocarbon degradation than bacteria [63], so the limited identification of archaeal hydrocarbon degraders might stem from the small pool of sequenced archaeal genomes or the significant phylogenetic divergence between archaeal hydrocarbon degradation genes and the mainly bacterial sequences that have been experimentally validated. The identified archaea and bacteria may be key contributors to hydrocarbon degradation in the benthic and pelagic ecosystems of the Baltic Sea, playing a crucial role in the natural bioremediation process in the area. Moreover, their prevalence and activity suggest an intrinsic mechanism within the benthic and pelagic ecosystems to mitigate pollution, which could have implications for environmental management strategies in the Baltic Sea and similar marine ecosystems.

#### HDG distribution showed distinct metabolic adaptations between pelagic and benthic communities

Among the degradation pathways, aerobic metabolism was the most prominent in the Baltic Sea (Fig. 1d,e), with HDGs associated with aerobic alkane metabolism being significantly more abundant in pelagic MAGs, reflecting the oxygen-rich conditions of pelagic environments, which favor aerobic processes [80]. On the other hand, HDGs linked to anaerobic alkane metabolism were more prevalent in benthic MAGs, aligning with the low-oxygen or anoxic conditions typically found in seabed sediments [81]. Similarly, Górgora et al. [36] evaluated the hydrocarbon degradation potential of Canadian high Arctic beaches and discovered that alkane metabolism is the prevalent form of hydrocarbon degradation in tidal beach ecosystems. For the aromatic degradation pathways, HDGs involved in anaerobic metabolism were higher in benthic compared to pelagic, while HDGs for aerobic metabolism exhibited similar levels. These findings highlight the ecological specialization of microbial communities in various marine habitats, which is driven by the availability of oxygen and other environmental factors.

Petroleum hydrocarbons are recalcitrant compounds that persist in the environment, influencing microbial communities through prolonged exposure [82, 83]. Hydrocarbons in dispersed oil can degrade in aerobic marine waters within days to months, whereas oil on shorelines is often more concentrated, nutrient-limited, and persists in the environment for a longer period [84]. Benthic habitats, characterized by more stable sediments, may experience slower degradation due to limited oxygen and nutrients [85, 86]. In contrast, pelagic environments, characterized by dynamic conditions and enhanced oxygenation, facilitate faster degradation [87]. Consequently, microbial degradation capacities are expected to vary between benthic and pelagic habitats due to differences in microbial communities and physicochemical conditions [88]. The pelagic ecosystem, characterized by its open water column, supports greater microbial and genetic diversity than the benthic ecosystem, which is influenced by sedimentary conditions [89]. The observed higher species and HDG diversity in the Baltic Sea's pelagic samples may be attributed to environmental factors, e.g., greater oxygen availability, dynamic nutrient cycling, and the presence of diverse organic matter sources [90]. In contrast, the benthic ecosystem often has limited oxygen levels and a more stable, yet nutrient-constrained, environment. These conditions necessitate specific metabolic adaptations among microbial communities, e.g., the ability to utilize alternative electron acceptors or degrade complex organic compounds [91].

Miettinen et al. [38] explored the hydrocarbon degradation capabilities of bacterial, archaeal, and fungal microbiomes in the Baltic Sea by comparing coastal sites with varying levels of long-term oil exposure. Through the identification of specific genes (i.e., *alkB* and PAH ring-hydroxylating dioxygenase), their study revealed differences in microbial composition and degradation potential between water and sediment samples, as determined by taxonomic profiling using amplicon sequencing. Additionally, previous studies have isolated alkane-degrading bacterial strains from water and sediment samples collected in the Baltic Sea and identified diverse *alkB* genes in the studied strains [42, 45]. Similar to our observations, the differences in hydrocarbon degradation pathways observed between these ecosystems further underscore the metabolic specialization of microorganisms in different habitats, with certain pathways being more prominent in oxygen-rich pelagic zones and others adapted to the anoxic and suboxic conditions of the benthic ecosystem across the Baltic Sea. This highlights the influence of distinct ecological niches and the possible metabolic adaptations of microbial inhabitants in the benthic and pelagic environments, alongside the intricate interplay between environmental gradients,

microbial diversity, and metabolic functions that shape the ecological roles of microorganisms. Moreover, the observed differences between benthic and pelagic microorganisms contribute to their survival and play a crucial role in biogeochemical processes, including nutrient cycling and the degradation of pollutants.

#### **Long-chain alkanes and dibenzothiophene are the preferred substrates of hydrocarbon degraders**

The genes associated with the degradation of long-chain alkanes (i.e., C20–C32) and dibenzothiophene were the most prevalent and exhibited the highest abundance (Fig. 6). Long-chain alkanes are challenging to degrade due to their large molecular size and hydrophobic nature. Both benthic and pelagic ecosystems contained alkane monooxygenases (*alma* Group I and III), which are reported to play crucial roles in the metabolic pathway of extremely hydrophobic long-chain n-alkanes (>C20) that are particularly difficult to degrade [92]. These flavin-binding monooxygenases utilize flavin cofactors to catalyze the oxidation of long-chain hydrocarbons. Additionally, we identified the long-chain alkane monooxygenase (*ladB*) in the Baltic Sea samples, which specializes in oxidizing long-chain alkanes. The dibenzothiophene monooxygenase (*dszC*) gene was the second most abundant hydrocarbon degradation-associated gene in the Baltic Sea samples. This gene plays a key role in the microbial desulfurization of dibenzothiophenes (DBT), a sulfur-containing compound commonly found in crude oil [93]. As one of the four enzymes in the 4S pathway, *dszC* catalyzes the initial step of converting DBT into DBT sulfone, setting the initial stage for subsequent enzymatic reactions that ultimately remove sulfur without breaking carbon–carbon bonds [94]. Its abundance suggests that the microbial communities of the Baltic Sea are well-suited to handle sulfur-rich hydrocarbons, potentially due to the levels of oil resulting from anthropogenic activities or natural seepage in the region [95]. The presence of these genes reflects the adaptive capabilities of Baltic Sea microorganisms in addressing recalcitrant hydrocarbons, converting them into more reactive and degradable intermediates. This step is crucial for the complete elimination of these hydrocarbons from the environment. Our findings suggest that both benthic sediments and the pelagic water column are equipped to handle hydrocarbon pollution, including oil spills, by leveraging these enzymatic pathways.

#### **Depth and oil spill volume were the most influential factors on HDG composition**

The chemical complexity of oil and its pollution, combined with environmental factors, i.e., depth, oxygen levels, temperature, nutrient availability, and other physical

and chemical influences, governs the assembly of microbial communities and the degradation processes they carry out in seawater and marine sediments following oil spills [96]. In particular, we report that depth plays a key role in shaping the composition of genes associated with hydrocarbon degradation in the Baltic Sea (Fig. 7b). Vigneron et al. [83] reported similar observations, profiling the distribution of microorganisms with hydrocarbon degradation potential at various depths in Lake A in the Canadian High Arctic. In deeper waters, reduced oxygen levels may favor anaerobic degradation pathways, while aerobic processes might dominate in shallower areas [89]. Additionally, sedimentation rates and the accumulation of hydrocarbons in benthic zones can vary with depth, influencing the availability of substrates for microbial degradation [91, 97].

Natural hydrocarbon degraders in the Baltic Sea are closely linked to the annual average oil spill documented in each subbasin (Fig. 7c). For example, the Gulf of Finland, which had the highest annual average oil spill volume record, showed relatively lower HDG diversity (Supplementary Fig. S5), having genes associated with the degradation of long-chain (C20–C32) hydrocarbons more prevalent. This suggests the potential enrichment of specific microbial hydrocarbon degraders in the presence of their preferred substrates. On the other hand, the Gulf of Riga and the Quark, where no significant oil spills were recorded, had more diverse HDG profiles. Similar observations were reported for petroleum hydrocarbon-polluted soil, where unpolluted soils had higher diversity and evenly distributed communities than contaminated samples [98]. However, this is not consistent with other subbasins. For example, the Bothnian Bay experienced a low annual oil spill volume and exhibited the highest benthic HDG diversity. Still, these observations indicate that the natural presence of hydrocarbon degraders in the Baltic Sea highlights the importance of understanding spatial variability in microbial responses to oil spills and the corresponding spill volumes. Moreover, salinity and latitude also play significant roles, suggesting that a combination of abiotic factors influences the distribution and activity of these microbial communities. Overall, these observations highlight the importance of depth and salinity as key factors influencing microbial responses to oil pollution. They also highlight the spatial heterogeneity of oil spills and their ecological implications.

#### **Implications for the bioremediation of petroleum hydrocarbons**

The volume of oil spills in the Baltic Sea is influenced by a combination of anthropogenic activities, environmental conditions, and governance challenges [15, 17, 99]. Key factors include the high volume of oil transportation

through the region, with major shipping routes and oil terminals contributing to accidental and operational spills [17]. For instance, the Gulf of Finland experiences higher spill levels due to its role as a critical transit area for oil exports [100]. Similarly, the Northern Baltic Proper and Kattegat are busy maritime zones, where dense traffic increases the likelihood of spills [101]. The natural presence of hydrocarbon-degrading microorganisms in the Baltic Sea, coupled with their capacity to adapt to varying oil spill thresholds, can be strategically utilized to estimate the persistence of petroleum hydrocarbons and inform oil spill management. Understanding the functional potential of microbial communities can be used to predict how long petroleum-based pollutants may persist in the environment and how they might be transported across different areas [31, 32]. This predictive ability is crucial for guiding cleanup operations, allocating resources more effectively, and prioritizing areas in most urgent need of intervention. The ability to monitor variations in hydrocarbon degradation across regions also supports more precise environmental monitoring. These insights could further inform policies and guidelines for managing oil spills, integrating microbial degradation as a key element of broader mitigation strategies.

Biostimulation techniques can enhance microbial degradation in subbasins with high spill volumes by optimizing environmental conditions like nutrient and oxygen levels [102]. Previous studies have shown that nutrient amendments to contaminated areas, e.g., the Labrador Sea [103], or from potentially at-risk areas, e.g., high-Arctic beaches [104], enhanced the hydrocarbon degradation activity of their native microbial communities. Additionally, dispersants can be used to enhance the natural biodegradation of petroleum hydrocarbons by breaking down oil particles into smaller droplets, increasing the surface area available for microbial activity [105]. Bioaugmentation, the introduction of robust hydrocarbon-degrading strains in contaminated areas, can also be utilized for cleanup efforts [106]. In less-exposed regions, the resilience of native microorganisms ensures that they can still respond effectively to occasional spills.

## Conclusions

We recovered metagenome-assembled genomes (MAGs) from Baltic Sea metagenomic datasets that showed significant genetic potential for hydrocarbon degradation, with 80% containing more than 10 RPKM of hydrocarbon degradation genes (HDG). We also identified known broad-spectrum hydrocarbon-degrading bacteria, i.e., *Pseudomonadota*, *Myxococcota*, *Actinomycetota*, and *Desulfobacterota*, each represented by multiple HDG annotations. Genes involved in the degradation of long-chain alkanes (C20–C32) and dibenzothiophene were the

most abundant and widespread among Baltic Sea microbial communities. Both benthic and pelagic environments harbored key monooxygenases, including alkane monooxygenases (i.e., *alma* Group I and III) and *ladB*, which facilitate the oxidation of recalcitrant, hydrophobic hydrocarbons. Notably, the high prevalence of the *dszC* gene underscores the microbial community's potential to initiate desulfurization of sulfur-rich compounds (i.e., dibenzothiophene), highlighting an inherent capacity to mediate complex hydrocarbon transformations and suggesting a regionally adaptive enzymatic response to petroleum hydrocarbon pollutants.

Our study also highlights the significant differences in degradation potential between benthic and pelagic environments, as well as among various subbasins within the Baltic Sea. Understanding how environmental factors, i.e., depth and salinity, influence microbial activity enables the development of tailored, site-specific bioremediation plans. By harnessing these microbial communities, oil spill management can become more efficient and sustainable, minimizing environmental damage while supporting ecosystem recovery. Our findings not only enhance our understanding of hydrocarbon-degrading microbial communities in the benthic and pelagic ecosystems of the Baltic Sea but also offer insights into potential microbial-based strategies for the environmental remediation of petroleum hydrocarbons in marine environments. With this, we recommend further assessment and validation of the functional capacities of native microbial communities in the Baltic Sea ecosystem using culture-dependent methods and multi-omic approaches to provide a comprehensive understanding of the specific mechanisms of natural hydrocarbon attenuation in marine ecosystems.

## Abbreviations

|        |                                                     |
|--------|-----------------------------------------------------|
| dbRDA  | Distance-based redundancy analysis                  |
| HDG    | Hydrocarbon degradation gene                        |
| HELCOM | The Baltic Marine Environment Protection Commission |
| HMM    | Hidden Markov model                                 |
| HOLAS  | HELCOM holistic assessments                         |
| NMDS   | Non-metric multidimensional scale                   |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-025-02211-w>.

Additional file 1: Supplementary Fig. S1 Map of the sampling locations of the 203 Baltic Sea metagenomes. Colors indicate the distribution of each sample based on the Baltic Sea sub-basin category of HELCOM (2022). Supplementary Fig. S2 Histograms showing the distribution of MAG characteristics. (a) Genome size, (b) GC Content, (c) N50 (contigs), and (d) predicted gene counts. Colors represent the environment where the MAGs were located, i.e., pelagic (purple) and benthic (brown). The dashed line indicates mean values. Supplementary Fig. S3 Scatter plots of the relationship between HDG Count and Genome Size (bp) for (a) all MAGs, (b) benthic MAGs, and (c) pelagic MAGs. Supplementary Fig. S4 Frequency of HDGs annotated per Baltic Sea sub-basin. Supplementary Fig. S5 Alpha

diversity estimates of the hydrocarbon degradation genes (HDGs) per sub-basin for (a) benthic and (b) pelagic samples. Supplementary Fig. S6 Upset plot of the shared and unique HMM annotations per Baltic Sea sub-basin. Supplementary Fig. S7 Maximum-likelihood phylogenomic tree of the archaeal MAGs. Supplementary Fig. S8 Relative abundance of the top 10 most abundant species based on the total number of medium-quality MAGs visualized (a) per environment and (b) per Baltic Sea sub-basin. Supplementary Fig. S9 Upset plot of the shared and unique species per Baltic Sea sub-basin. Supplementary Fig. S10 Bubble plot of the classified microbial taxa at the phylum-level and the occurrence of hydrocarbon degradation genes (HDGs) in each taxa. Supplementary Fig. S11 Random-forest classification of the relative abundance HDGs across the environments: (left) HDGs presented in descending order of importance, and the (right) relative abundance of HDGs enriched in each environment. Supplementary Fig. S12 Random-forest classification of the relative abundance of HDG substrates across the environments: (left) Substrates presented in descending order of importance, and the (right) relative abundance of HDG substrates enriched in each environment

Additional file 2: Supplementary Table S1 Metagenomics sample information and environmental factors. Metagenomics sequencing data were obtained from the European Nucleotide Archive (ENA) database: accession numbers PRJEB41834 (Broman et al., 2022; Rodríguez-Gijón et al., 2023), PRJEB22997 (Alneberg et al., 2018), and PRJEB34883 (Alneberg et al., 2020). The environmental parameters, i.e., depth (m), salinity (PSU), and temperature (°C), compiled by Rodríguez-Gijón et al. (2023) were used in this study. Table S2 Metagenomics data processing information: Sequence codes, read processing counts (from quality-filtering to contig assembly), prodigal annotations and binning stat values. The Calgary approach to ANnotating HYDrocarbon degradation genes (CANT-HYD; Khot et al., 2022) counts highlighted in blue. Table S3 Metagenome-assembled genome (MAG) stats and taxonomic classification based on using GTDB-Tk v2.4.0 with R220 (Chaumeil et al., 2022). Table S4 The Calgary approach to ANnotating HYDrocarbon degrading enzymes database (CANT-HYD) annotations (Khot et al. 2022). Table S5 Count and annotation table of the annotated hydrocarbon degradation genes (HDGs). Reads per kilobase million (RPKM) absolute counts. Table S6. The annual average of total oil spills during the assessment period 2016–2021 (m<sup>3</sup>) values (Supplementary Table S6) from HELCOM (2023b). Table S7 Average metagenome-assembled genome (MAG) quality stats per environment. Table S8 Total metagenome-assembled genome (MAG) counts per environment and relative abundance based on species-level classification. Table S9 Statistical results of the permutation test for homogeneity of multivariate dispersions and permutation test for adonis under a reduced model of the hydrocarbon degradation genes (HDG) dataset. Table S10 Differential abundance test by Random Forests for the HDGs per environment and per subbasin. Table S11 Statistical results of the permutation test for distance-based redundancy analysis (dbRDA) analyses in R of the ARG composition and environmental factors. Table S12 Spearman rank correlations between environmental variables and the RPKM abundance of the hydrocarbon degradation genes (HDG)

## Acknowledgements

The authors acknowledge the National Academic Infrastructure for Super-computing in Sweden for granting us access and storage to the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) computational infrastructure.

## Authors' contributions

JMS, BD, EB, and MP conceptualized the study. BD compiled the HELCOM literature and data. EB and FN provided support on the bioinformatics framework and analysis. JMS performed data analysis and wrote the first draft of the manuscript. All authors contributed to the writing and editing of the manuscript.

## Funding

Open access funding provided by Stockholm University. Open-access funding is provided by Stockholm University. JMS is supported by the Stockholm University Center for Circular and Sustainable Systems (SUCCeSS) (Project No. 30002687) postdoc funding. The computations were enabled by resources

in projects NAISS 2023/22–846 and 2023/23–424 provided by the National Academic Infrastructure for Super-computing in Sweden at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

## Data Availability

All the raw sequence data are obtained from the European Nucleotide Archive (ENA) database with project accession numbers PRJEB41834 (Broman et al., 2022; Rodríguez-Gijón et al., 2023), PRJEB22997 (Alneberg et al., 2018), and PRJEB34883 (Alneberg et al., 2020). The specific accession numbers for each metagenome are listed in Table S1. Additional data from the analyses presented in this paper are available in the Supplementary Figures and Tables. The corresponding visualization and downstream analysis codes are publicly accessible at <https://github.com/jserrana/baltic-hdg>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Stockholm University Center for Circular and Sustainable Systems (SUCCeSS), Stockholm University, 106 91 Stockholm, Sweden. <sup>2</sup>Department of Environmental Science (ACES), Stockholm University, 106 91 Stockholm, Sweden.

<sup>3</sup>Baltic Sea Centre, Stockholm University, Stockholm, Sweden. <sup>4</sup>Department of Ecology, Environment, and Plant Sciences (DEEP), Stockholm University, 106 91 Stockholm, Sweden.

Received: 8 May 2025 Accepted: 23 July 2025

Published online: 15 October 2025

## References

- Uddin S, Fowler SW, Saeed T, Jupp B, Faizuddin M. Petroleum hydrocarbon pollution in sediments from the Gulf and Omani waters: status and review. *Mar Pollut Bull.* 2021;173: 112913.
- Liu F, Huang Q, Du Y, Li S, Cai M, Huang X, Zheng Fengying, Lin L. The interference of marine accidental and persistent petroleum hydrocarbons pollution on primary biomass and trace elements sink. *Sci Total Environ.* 2023;883: 163812.
- Kuppusamy, S., Maddela, N. R., Megharaj, M., Venkateswarlu, K., Kuppusamy, S., Maddela, N. R., ... & Venkateswarlu, K. (2020). Ecological impacts of total petroleum hydrocarbons. *Total Petroleum Hydrocarbons: Environmental Fate, Toxicity, and Remediation*, 95–138.
- McGenity TJ, Folwell BD, McKew BA, Sanni GO. Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat Biosyst.* 2012;8:1–19.
- Brussaard, C. P., Peperzak, L., Beggah, S., Wick, L. Y., Wuerz, B., Weber, J., ... & Van Der Meer, J. R. . Immediate ecotoxicological effects of short-lived oil spills on marine biota. *Nat Commun.* 2016;7(1), 11206.
- Berenshtein I, Paris CB, Perlín N, Alloy MM, Joye SB, Murawski S. Invisible oil beyond the Deepwater Horizon satellite footprint. *Sci Adv.* 2020;6(7): eaaw8863.
- Love CR, Arrington EC, Gosselin KM, Reddy CM, Van Mooy BA, Nelson RK, Valentine DL. Microbial production and consumption of hydrocarbons in the global ocean. *Nat Microbiol.* 2021;6(4):489–98.
- McGenity TJ, McKew BA, Lea-Smith DJ. Cryptic microbial hydrocarbon cycling. *Nat Microbiol.* 2021;6(4):419–20.
- Volkman JK, Holdsworth DG, Neill GP, Bavor HJ Jr. Identification of natural, anthropogenic and petroleum hydrocarbons in aquatic sediments. *Sci Total Environ.* 1992;112(2–3):203–19.

10. NRC: National Research Council, Transportation Research Board, Division on Earth, Life Studies, Marine Board, Ocean Studies Board, ... & Fates. Oil in the sea III: inputs, fates, and effects. 2003
11. Dong Y, Liu Y, Hu C, MacDonald IR, Lu Y. Chronic oiling in global oceans. *Science*. 2022;376(6599):1300–4.
12. HELCOM. HELCOM Assessment on Maritime Activities in the Baltic Sea 2018. Baltic Sea Environment Proceedings No. 152. 2018
13. Kostianoy, A. G. Oil pollution: the Baltic sea. In *Managing Water Resources and Hydrological Systems 2020*; (pp. 349–368). CRC Press.
14. Sonesten, L., Undeman, E., Svendsen, L. M., Frank-Kamenetsky, D., & Haapaniemi, J. Inputs of hazardous substances to the Baltic Sea. *Baltic Sea Environment Proceedings* 196. 2021
15. Hassler B. Accidental versus operational oil spills from shipping in the Baltic Sea: risk governance and management strategies. *Ambio*. 2011;40:170–8.
16. HELCOM. HELCOM Thematic assessment of hazardous substances, marine litter, underwater noise and non-indigenous species 2016–2021. *Baltic Sea Environment Proceedings* 190. (2023a) [https://helcom.fi/post\\_type\\_publ/holas3\\_haz/](https://helcom.fi/post_type_publ/holas3_haz/)
17. HELCOM. Operational oil spills from ships. HELCOM core indicator report. Online. July 20, 2024, (2023b). <https://indicators.helcom.fi/indicator/oil-spills/>.
18. Venesjärvi R, Jolma A, Helle I. Sensitivity index for conservation priority ranking in the oil spill response: a case study for the coastal and marine species and habitat types in the Baltic Sea. *Ecotoxicol Environ Saf*. 2023;257: 114936.
19. Szpiech, A., Bojke, A., Littwin, M., Dzendrowska, A., Duljas, E., Flasińska, A., ... & Hac, B. Baltic Sea shipwrecks as a source of hazardous pollution. *PeerJ Anal Chem*;2024; 6, e31.
20. Raymond C, Gorokhova E, Karlson AM. Polycyclic aromatic hydrocarbons have adverse effects on benthic communities in the Baltic Sea: implications for environmental status assessment. *Front Environ Sci*. 2021;9: 624658.
21. Ivshina IB, Kuyukina MS, Krivoruchko AV, Elkin AA, Makarov SO, Cunningham CJ, Peshkur Tatyana A., Atlas Ronald M., Philip JC. Oil spill problems and sustainable response strategies through new technologies. *Environ Sci Process Impacts*. 2015;17(7):1201–19.
22. Beyer J, Trannum HC, Bakke T, Hodson PV, Collier TK. Environmental effects of the Deepwater Horizon oil spill: a review. *Mar Pollut Bull*. 2016;110(1):28–51.
23. Braw, E. Russia's growing dark fleet: Risks for the global maritime order. Atlantic Council. 2024. <https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/russias-growing-dark-fleet-risks-for-the-global-maritime-order/>
24. Jack, V., Gambarini C, Mathiesen K, Guillot L & Cokelaere H Revealed: Putin's sanctions-busting shadow fleet is spilling oil all over the world. Politico. 2024 <https://www.politico.eu/article/revealed-putins-sanctions-busting-shadow-fleet-is-spilling-oil-all-over-the-world/>
25. More, R., & Heine, F. Germany seizes tanker belonging to Russian 'shadow fleet,' Spiegel reports. Reuters. (2025, March 21).
26. Head IM, Jones DM, Röling WF. Marine microorganisms make a meal of oil. *Nat Rev Microbiol*. 2006;4(3):173–82.
27. Chunyan X, Qaria MA, Qi X, Daochen Z. The role of microorganisms in petroleum degradation: current development and prospects. *Sci Total Environ*. 2023;865: 161112.
28. Lin L, Ning J, Li C, Zhang L, Gao Y, Bo W, Shi K. Systematic degradation process of petroleum hydrocarbons by an integrated bacterial consortium under bottom seawater and surface seawater environments. *Environ Technol*. 2025;46(5):752–60.
29. Xue J, Yu Y, Bai Y, Wang L, Wu Y. Marine oil-degrading microorganisms and biodegradation process of petroleum hydrocarbon in marine environments: a review. *Curr Microbiol*. 2015;71(2):220–8.
30. Zhou Y, Wang Y, Yao S, Zhao X, Kong Q, Cui L, Zhang H. Driving mechanisms for the adaptation and degradation of petroleum hydrocarbons by native microbiota from seas prone to oil spills. *J Hazard Mater*. 2024;476: 135060.
31. Hazaimah MD, Ahmed ES. Bioremediation perspectives and progress in petroleum pollution in the marine environment: a review. *Environ Sci Pollut Res*. 2021;28(39):54238–59.
32. Howe KL, Zaugg J, Mason OU. Novel, active, and uncultured hydrocarbon-degrading microbes in the ocean. *Appl Environ Microbiol*. 2024;90(9):e01224-e1324.
33. Atlas, R. M., & Hazen, T. C. Oil biodegradation and bioremediation: a tale of the two worst spills in US history. 2011
34. Bacosa HP, Erdner DL, Rosenheim BE, Shetty P, Seitz KW, Baker BJ, Liu Z. Hydrocarbon degradation and response of seafloor sediment bacterial community in the northern Gulf of Mexico to light Louisiana sweet crude oil. *ISME J*. 2018;12(10):2532–43.
35. Liu, J., Zheng, Y., Lin, H., Wang, X., Li, M., Liu, Y., ... & Zhang, X. H. (2019). Proliferation of hydrocarbon-degrading microbes at the bottom of the Mariana Trench. *Microbiome*, 7, 1–13.
36. Góngora E, Lirette AO, Freyria NJ, Greer CW, Whyte LG. Metagenomic survey reveals hydrocarbon biodegradation potential of Canadian high Arctic beaches. *Environ Microbiome*. 2024;19(1):72.
37. Liu Q, Peng Y, Liao J, Liu X, Peng J, Wang JH, Shao Z. Broad-spectrum hydrocarbon-degrading microbes in the global ocean metagenomes. *Sci Total Environ*. 2024;926: 171746.
38. Miettinen H, Bomberg M, Nyyssönen M, Reunamo A, Jørgensen KS, Vikman M. Oil degradation potential of microbial communities in water and sediment of Baltic Sea coastal area. *PLoS One*. 2019;14(7): e0218834.
39. Picazo-Espinosa, R., Overlingé, D., Mineiké, E., Paulauskiene, T., Uebe, J., Bôto, M. L., ... & Kataržytė, M. Biodegradation Potential of Microorganisms in Hydrocarbon-Polluted Baltic Sea Sediments. SSRN, <https://doi.org/10.2139/ssrn.5080923>
40. Reunamo A, Riemann L, Leskinen P, Jørgensen KS. Dominant petroleum hydrocarbon-degrading bacteria in the Archipelago Sea in South-West Finland (Baltic Sea) belong to different taxonomic groups than hydrocarbon degraders in the oceans. *Mar Pollut Bull*. 2013;72(1):174–80.
41. Reunamo A, Yli-Hemminki P, Nuutinen J, Lehtoranta J, Jørgensen KS. Degradation of crude oil and PAHs in iron–manganese concretions and sediment from the Northern Baltic Sea. *Geomicrobiol J*. 2017;34(5):385–99.
42. Izmalkova TY, Gafarov AB, Sazonova OI, Sokolov SL, Kosheleva IA, Boronin AM. Diversity of oil-degrading microorganisms in the Gulf of Finland (Baltic Sea) in spring and in summer. *Microbiology*. 2018;87:261–71.
43. Eronen-Rasimus EL, Nakki PP, Kaartokallio HP. Degradation rates and bacterial community compositions vary among commonly used bioplastic materials in a brackish marine environment. *Environ Sci Technol*. 2022;56(22):15760–9.
44. Viggor S, Juhanson J, Jõesaar M, Mitt M, Truu J, Vedler E, Heinaru A. Dynamic changes in the structure of microbial communities in Baltic Sea coastal seawater microcosms modified by crude oil, shale oil or diesel fuel. *Microbiol Res*. 2013;168(7):415–27.
45. Broman E, Izabel-Shen D, Rodríguez-Gijón A, Bonaglia S, García SL, Nascimento FJ. Microbial functional genes are driven by gradients in sediment stoichiometry, oxygen, and salinity across the Baltic benthic ecosystem. *Microbiome*. 2022;10(1): 126.
46. Rodríguez-Gijón A, Buck M, Andersson AF, Izabel-Shen D, Nascimento FJ, García SL. Linking prokaryotic genome size variation to metabolic potential and environment. *ISME Commun*. 2023;3(1):25.
47. Alneberg J, Bennke C, Beier S, Bunse C, Quince C, Ininbergs K, Riemann Lasse, Ekman Martin, Jürgens Klaus, Labrenz Matthias, Pinhassi Jarone, Andersson AF. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol*. 2020;3(1): 119.
48. Alneberg J, Sundh J, Bennke C, Beier S, Lundin D, Hugerth LW, Pinhassi J, Kisand V, Riemann L, Jürgens K, Labrenz M, Andersson AF. BARM and BalticMicroDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. *Sci Data*. 2018;5: 180146.
49. HELCOM. HELCOM Monitoring and Assessment Strategy: Attachment 4 - HELCOM sub-divisions of the Baltic Sea, 2022;23–39.
50. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics*. 2018;34(17):i884–90.
51. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol*. 2019;20:1–13.
52. Bushnell, B. BBTools: A suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. 2018. URL <http://www.sourceforigenet/projects/bbmap>.

53. Li D, Liu CM, Luo R, Sadakane K, Lam TW. Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics*. 2015;31(10):1674–6.
54. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11:1–11.
55. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–100.
56. Liao Y, Smyth GK, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
57. Pan S, Zhao XM, Coelho LP. Semibin2: self-supervised contrastive learning leads to better MAGs for short-and long-read sequencing. *Bioinformatics*. 2023;39(Supplement\_1):i21–9.
58. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043–55.
59. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*. 2022;38(23):5315–6.
60. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
61. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4.
62. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
63. Khot V, Zorz J, Gittins DA, Chakraborty A, Bell E, Bautista MA, Paquette Alexandre J, Hawley Alyse K, Novotnik Breda, Hubert Casey R. J., Strous Marc, Bhatnagar S. CANT-HYD: a curated database of phylogeny-derived hidden Markov models for annotation of marker genes involved in hydrocarbon degradation. *Front Microbiol*. 2022;12: 764058.
64. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–8.
65. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2025. <https://www.R-project.org/>.
66. Liu C, Cui Y, Li X, Yao M. microeco: an R package for data mining in microbial community ecology. *FEMS Microbiol Ecol*. 2021;97(2): faa255.
67. An J, Liu C, Wang Q, Yao M, Rui J, Zhang S, Li X. Soil bacterial community structure in Chinese wetlands. *Geoderma*. 2019;337:290–9.
68. Barnett DJ, Arts IC, Penders J. microViz: an R package for microbiome data visualization and statistics. *JOSS*. 2021;6(63): 3201.
69. Camacho J, Mesén-Porras E, Rojas-Gätjens D, Pérez-Pantoja D, Puente-Sánchez F, Chavarría M. Draft genome sequence of three hydrocarbon-degrading *Pseudomonadota* strains isolated from an abandoned century-old oil exploration well. *MRA*. 2024;13(2):e01076–e1123.
70. Ramos-Mendoza IS, Embarcadero-Jiménez S, Barrios-Navarro AF, Ledezma-González DC, Lago-Lestón A, Valenzuela-Suárez BJ, Ramírez-Álvarez Nancy, Silva-Jiménez H. Prokaryotic community structure and predicted metabolism associated with hydrocarbon degradation in marine sediments from the Northwest Coast of Baja California, Mexico. *J Soils Sediments*. 2024;24(8):3148–66.
71. Rehman K, Imran A, Amin I, Afzal M. Inoculation with bacteria in floating treatment wetlands positively modulates the phytoremediation of oil field wastewater. *J Hazard Mater*. 2018;349:242–51.
72. Walton JL, Buchan A. Evidence for novel polycyclic aromatic hydrocarbon degradation pathways in culturable marine isolates. *Microbiol Spectr*. 2024;12(1):e03409–e3423.
73. Deng Y, Sun W, Li Y, Wei J, Rao R, Cao Q, Wang Sichang, Zhang Qunzheng, Ke C. Innovative microbial activators for enhanced bioremediation of oil-contaminated soils: mechanistic insights. *World J Microbiol Biotechnol*. 2025;41(2): 47.
74. Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM, Kuker Jerome J, Wawrik B. Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol*. 2010;44(19):7287–94.
75. Dell'Anno F, Rastelli E, Tangherlini M, Corinaldesi C, Sansone C, Brunet C, Balzano Sergio, Ianora Adrianna, Musco Luigi, Montereali Maria Rita, Dell'Anno A. Highly contaminated marine sediments can host rare bacterial taxa potentially useful for bioremediation. *Front Microbiol*. 2021;12: 584850.
76. Davidova, I. A., Marks, C. R., & Suflita, J. M. Anaerobic hydrocarbon-degrading Deltaproteobacteria. *Taxonomy, genomics and ecophysiology of hydrocarbon-degrading microbes*, 2019;207–243.
77. Pavlova ON, Bukin SV, Izosimova ON, Chernitsina SM, Ivanov VG, Khabuev AV, Pogodaeva TV, Elovskaya IS, Gorshkov AG, Zemskaya TI. Anaerobic oxidation of oil by microbial communities of bottom sediments of a natural oil seepage site (Bolshaya Zelenovskaya, Middle Baikal). *Microbiology*. 2024;93(5):563–75.
78. Seidel L, Broman E, Turner S, Ståhle M, Dopson M. Interplay between eutrophication and climate warming on bacterial communities in coastal sediments differs depending on water depth and oxygen history. *Scie Rep*. 2021;11(1):23384.
79. Mußmann M, Pjevac P, Krüger K, Dyksa S. Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. *ISME J*. 2017;11(5):1276–81.
80. Vázquez Rosas Landa, M., De Anda, V., Rohwer, R. R., Angelova, A., Waldram, G., Gutierrez, T., & Baker, B. J. Exploring novel alkane-degradation pathways in uncultured bacteria from the North Atlantic Ocean. *Msystems*, 2023; 8(5), e00619–23.
81. Stagars MH, Ruff SE, Amann R, Knittel K. High diversity of anaerobic alkane-degrading microbial communities in marine seep sediments based on (1-methylalkyl) succinate synthase genes. *Front Microbiol*. 2016;6:1511.
82. Gao H, Wu M, Liu H, Xu Y, Liu Z. Effect of petroleum hydrocarbon pollution levels on the soil microecosystem and ecological function. *Environ Pollut*. 2022;293: 118511.
83. Vigneron A, Guyoneaud R, Goñi-Urriza M. Genome-centric metatranscriptomics reveals multiple co-occurring routes for hydrocarbon degradation in chronically contaminated marine microbial mats. *Environ Sci Technol*. 2024;58(3):1551–62.
84. Hazen TC, Prince RC, Mahmoudi N. Marine oil biodegradation. *Environ Sci Technol*. 2016;50:2121–9. <https://doi.org/10.1021/acs.est.5b03333>.
85. Widdel F, Knittel K, & Galushko, A. Anaerobic hydrocarbon-degrading microorganisms: an overview. *Handbook of hydrocarbon and lipid microbiology*, 1998–2022. 2010.
86. Das N, Chandran P. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int*. 2011;2011(1): 941810.
87. Vigneron A, Cruaud P, Lovejoy C, Vincent WF. Genomic insights into cryptic cycles of microbial hydrocarbon production and degradation in contiguous freshwater and marine microbiomes. *Microbiome*. 2023;11(1):104.
88. Abed RM, Al-Hinai M, Al-Balushi Y, Haider L, Muthukrishnan T, Rinner U. Degradation of starch-based bioplastic bags in the pelagic and benthic zones of the Gulf of Oman. *Mar Pollut Bull*. 2023;195: 115496.
89. Wang M, Ma Y, Feng C, Cai L, Li W. Diversity of pelagic and benthic bacterial assemblages in the western Pacific Ocean. *Front Microbiol*. 2020;11:1730.
90. Zhao S, Hu X, Li H, Zhang H, Lu J, Li Y, Chen Zhaohui, Bao M. Diversity and structure of pelagic microbial community in Kuroshio Extension. *Mar Environ Res*. 2024;201: 106697.
91. Wang Z, Wang X, Huang Y, Hou B. Metagenomic insights into nutrient and hypoxic microbial communities at the macrofouling/steel interface leading to severe MIC. *npj Mater Degrad*. 2023;7(1): 41.
92. Chen, S., Cao, L., Lv, T., Liu, J., Gao, G., Li, M., ... & Ma, T. . Regulation mechanism of the long-chain n-alkane monooxygenase gene *almA* in *Acinetobacter venetianus* RAG-1. *Appl Environ Microbiol*, 2025;91(1), e02050–24.
93. Cao Y, Zhang B, Dong G, Chen B. Blue economy: a new era of petroleum microbiology in a changing climate. *NSO*. 2024;3(1):20230024.
94. Thor P, Granberg ME, Winnes H, Magnusson K. Severe toxic effects on pelagic copepods from maritime exhaust gas scrubber effluents. *Environ Sci Technol*. 2021;55(9):5826–35.
95. Jutterström S, Moldan F, Moldanová J, Karl M, Matthias V, Posch M. The impact of nitrogen and sulfur emissions from shipping

- on the exceedance of critical loads in the Baltic Sea region. *ACP*. 2021;21(20):15827–45.
96. Bargiela, R., Yakimov, M. M., Golyshin, P. N., & Ferrer, M. (2014). Distribution of hydrocarbon degradation pathways in the sea. *Sea*, 3(2003).
  97. Main CE, Ruhl HA, Jones DOB, Yool A, Thornton B, Mayor DJ. Hydrocarbon contamination affects deep-sea benthic oxygen uptake and microbial community composition. *Deep Sea Res Part I*. 2015;100:79–87.
  98. Bidja Abena, M. T., Chen, G., Chen, Z., Zheng, X., Li, S., Li, T., & Zhong, W. . Microbial diversity changes and enrichment of potential petroleum hydrocarbon degraders in crude oil-, diesel-, and gasoline-contaminated soil. *3 Biotech*, 2020;10, 1–15.
  99. Hassler, B. Oil spills from shipping: a case study of the governance of accidental hazards and intentional pollution in the Baltic Sea. *Environmental governance of the Baltic Sea*, 2016;125–146.
  100. Pärt S, Kankaanpää H, Björkqvist JV, Uiboupin R. Oil spill detection using fluorometric sensors: laboratory validation and implementation to a ferrybox and a moored smartbuoy. *Front Mar Sci*. 2021;8: 778136.
  101. Jalkanen JP, Johansson L, Wilewska-Bien M, Granhag L, Ytreberg E, Eriksson KM, Yngsøell Daniel, Hassellöv Ida-Maja, Magnusson Kerstin, Raudsepp Urmas, Maljutenko Ilja, Winnes Hulda, Moldanova J. Modelling of discharges from Baltic Sea shipping. *Ocean Sci*. 2021;17(3):699–728.
  102. Pete AJ, Bharti B, Benton MG. Nano-enhanced bioremediation for oil spills: a review. *ACS ES&T Eng*. 2021;1(6):928–46.
  103. Murphy SM, Bautista MA, Cramm MA, Hubert CR. Diesel and crude oil biodegradation by cold-adapted microbial communities in the Labrador Sea. *Appl Environ Microbiol*. 2021;87(20):e00800–e821.
  104. Chen YJ, Altshuler I, Freyria NJ, Lirette A, Góngora E, Greer CW, Whyte LG. Arctic's hidden hydrocarbon degradation microbes: investigating the effects of hydrocarbon contamination, biostimulation, and a surface washing agent on microbial communities and hydrocarbon biodegradation pathways in high-Arctic beaches. *Environ Microbiome*. 2024;19(1):81.
  105. Zhu Z, Merlin F, Yang M, Lee K, Chen B, Liu B, Cao Yiqi, Song Xing, Ye Xudong, Li Qingqi K., Greer Charles W., Boufadel Michel C., Isaacman Lisa, Zhang B. Recent advances in chemical and biological degradation of spilled oil: a review of dispersants application in the marine environment. *J Hazard Mater*. 2022;436: 129260.
  106. Gao Y, Cai M, Shi K, Sun R, Liu S, Li Q, Wang Xiaoyan, Hua Wenxin, Qiao Yanlu, Xue Jianliang, Xiao X. Bioaugmentation enhance the bioremediation of marine crude oil pollution: microbial communities and metabolic pathways. *Water Sci Technol*. 2023;87(1):228–38.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.