

RESEARCH

Open Access



# Genome-resolved analyses show an extensive diversification in key aerobic hydrocarbon-degrading enzymes across bacteria and archaea

Maryam Rezaei Somee<sup>1</sup>, Mohammad Ali Amoozegar<sup>1</sup>, Seyed Mohammad Mehdi Dastgheib<sup>2</sup>, Mahmoud Shavandi<sup>2</sup>, Leila Ghanbari Maman<sup>3</sup>, Stefan Bertilsson<sup>4</sup> and Maliheh Mehrshad<sup>4\*</sup>

## Abstract

**Background:** Hydrocarbons (HCs) are organic compounds composed solely of carbon and hydrogen that are mainly accumulated in oil reservoirs. As the introduction of all classes of hydrocarbons including crude oil and oil products into the environment has increased significantly, oil pollution has become a global ecological problem. However, our perception of pathways for biotic degradation of major HCs and key enzymes in these bioconversion processes has mainly been based on cultured microbes and is biased by uneven taxonomic representation. Here we used Annotree to provide a gene-centric view of the aerobic degradation ability of aliphatic and aromatic HCs in 23,446 genomes from 123 bacterial and 14 archaeal phyla.

**Results:** Apart from the widespread genetic potential for HC degradation in Proteobacteria, Actinobacteriota, Bacteroidota, and Firmicutes, genomes from an additional 18 bacterial and 3 archaeal phyla also hosted key HC degrading enzymes. Among these, such degradation potential has not been previously reported for representatives in the phyla UBA8248, Tectomicrobia, SAR324, and Eremiobacterota. Genomes containing whole pathways for complete degradation of HCs were only detected in Proteobacteria and Actinobacteriota. Except for several members of Crenarchaeota, Halobacterota, and Nanoarchaeota that have *tmoA*, *ladA*, and *alkB/M* key genes, respectively, representatives of archaeal genomes made a small contribution to HC degradation. None of the screened archaeal genomes coded for complete HC degradation pathways studied here; however, they contribute significantly to peripheral routes of HC degradation with bacteria.

**Conclusion:** Phylogeny reconstruction showed that the reservoir of key aerobic hydrocarbon-degrading enzymes in Bacteria and Archaea undergoes extensive diversification via gene duplication and horizontal gene transfer. This diversification could potentially enable microbes to rapidly adapt to novel and manufactured HCs that reach the environment.

**Keywords:** Aerobic hydrocarbon degradation, Alkane mono-oxygenase, Ring-hydroxylating oxygenase, Archaea, Bacteria

\*Correspondence: maliheh.mehrshad@slu.se

<sup>4</sup> Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), Box 7050, 75007 Uppsala, Sweden  
Full list of author information is available at the end of the article

## Background

According to the biogenic (organic) theory, petroleum hydrocarbons originate from the ancient remains of detrital matter buried and diagenetically modified in



marine sediments. This organic matter is then gradually converted to petroleum compounds enriched in aromatic and aliphatic hydrocarbons (HCs) via the sequential activity of aerobic and anaerobic microorganisms [1–3]. In addition to their role in the formation of oil HCs, microbes play a crucial role in the biological integration of these HCs into the actively cycled carbon pool [4]. Microbial HC degradation occurs through a cascade of enzymatic reactions in three main steps: (i) activation or attacking the HC-bond, (ii) producing signature intermediate compounds, and (iii) conversion of signature degradation intermediates to central cell metabolites, followed by their mineralization to CO<sub>2</sub>. Microorganisms must overcome and break the stability and energy in carbon-hydrogen bonds in order to degrade HCs. Since HCs are structurally diverse, a plethora of enzymes are involved in their activation and degradation, and consequently, the energy that needs to be invested in the initial degradation step varies. Vast number of microorganisms can degrade different HCs according to their enzymatic repertoire and available energy [5]. Microorganisms have evolved to degrade different HCs under both aerobic and anaerobic conditions. However, biodegradation typically occurs much faster under aerobic conditions, in part due to the availability of thermodynamically favorable electron acceptors that leads to higher energy yield [6], but also because of the action of some HC-degrading enzymes requires oxygen as substrate or cofactor. Similar to all biological pathways, rate-limiting key enzymes drive the main steps of HC degradation.

Under aerobic conditions, oxygenase enzymes initiate the degradation of different aliphatic or aromatic compounds by adding one (mono-oxygenase) or two (di-oxygenase) oxygen molecules. Saturated aliphatic compounds such as alkane and cycloalkane (studied here) are converted to their corresponding carboxylic acid in this process. Catechol/gentisate derivatives are intermediate compounds during aerobic degradation of mono- and polycyclic aromatic HCs. They are then de-aromatized via subsequent meta/ortho cleavage. Intermediate compounds produced during the degradation of aliphatic and aromatic HCs converge to the  $\beta$ -oxidation and tricarboxylic acid (TCA) cycle [7]. While enzymes involved in the downstream part of the degradation process are widespread across living cells shared by many metabolic pathways, the mono/di-oxygenase enzymes catalyzing the first hydroxylation of aliphatic/aromatic compounds are crucial for the initial step in the HC degradation process and likely rate-limiting. Accordingly, microorganisms carrying the enzymes for such

initial degradation will be rate-controlling drivers of HC degradation.

The capacity of microbial isolates to metabolically degrade oil HCs have been frequently studied [8–11]. However, our knowledge has been mainly limited to cultivated microorganisms until recently. The present study provides a systemic and genome-resolved view of hydrocarbon degradation potential in the growing database of archaeal and bacterial genomes. To provide this extensive view, we compiled a database of enzymes involved in the aerobic degradation pathway of aliphatic (short-chain and long-chain n-alkanes) and aromatic HCs (toluene, phenol, xylene, benzene, biphenyl, and naphthalene). We then explored the distribution of these enzymes in 24,692 publicly available archaeal ( $n = 1246$ ) and bacterial ( $n = 23,446$ ) genomes via AnnoTree [12] and manually confirmed all annotations. We focused on the microbial genomes containing enzymes for complete/near complete degradation of specific HCs and suggested that lineages with the great genetic potential to degrade a broad range of HC compounds can be exploited for bioremediation purposes. We also reconstructed the phylogenetic relationships of the recovered key HC degradation enzymes to investigate their evolution and explore the potential role of horizontal gene transfer. Several microorganisms contain multiple copies of key HC degrading genes across their genome. We thus explored whether these copies are likely to have been acquired through HGT or if they are likely to be paralogs.

Having a genome-resolved view, we also surveyed these genomes with regard to their ecological strategies by leveraging information on their GC content and genome size to see whether all genomes containing key genes for HC degradation adopt similar growth strategy in term of canonical r and K strategists.

## Results and discussion

### HC degradation across domain bacteria

alkB/M and almA/ladA genes are alkane mono-oxygenases that initiate the degradation of short (C<sub>5</sub>–C<sub>15</sub>) and long-chain alkanes (>C<sub>15</sub>), respectively. The alkB/M is rubredoxin-dependent, while almA and ladA are flavin-dependent mono-oxygenases.

The genes pheA (phenol 2-monooxygenase), xylM (xylene monooxygenase), xylX (toluate/benzoate 1,2-dioxygenase subunit alpha), todC1 (benzene 1,2-dioxygenase subunit alpha), and tmoA (toluene-4-monooxygenase system, hydroxylase component subunit alpha) for monocyclic, and bphA1 (biphenyl 2,3-dioxygenase subunit alpha), ndoB (naphthalene 1,2-dioxygenase system, large oxygenase component) for polycyclic compounds code for catalytic domains of ring hydroxylating oxygenases (RHOs) that add -OH group(s) to compounds

undergoing degradation (Supplementary Figure S1). We explored the distribution of these genes and associated degradation pathways in a total of 23,446 representatives out of 143,512 bacterial genomes available in release 89 of the GTDB database that has been annotated via Annotree [12]. These annotated genomes are dominated by representatives of phyla Proteobacteria (32.5%), Actinobacteriota (13.3%), Bacteroidota (12.13%), and Firmicutes (8.01%) (Supplementary Figure S2 and Supplementary Table S4). Among the 123 represented bacterial phyla, 58 phyla had  $\leq$  five genomes available per phylum and combined only represented 0.57% of the explored genomes. To avoid misinterpretations due to this uneven taxonomic distribution of representative genomes, we explored the contribution of members of each phylum in the HC degradation process by showing what proportion of microbes containing each HC degrading enzyme exist in each phylum (panel A of Supplementary Figures S3-S6). We also analyze the percentage of members of each phylum containing each HC degrading enzyme to ensure that we consider the contributions of underrepresented phyla in the HC degradation (panel B of Supplementary Figures S3-S6).

As expected, representatives of the phylum Proteobacteria (Pseudomonadales and Burkholderiales orders) presented the highest abundance of aliphatic and aromatic HC degrading enzymes, followed by Actinobacteriota and Bacteroidota for aliphatic and Actinobacteriota and Firmicutes for aromatic HC degrading enzymes (Supplementary Figures S3 and S4, panel A).

Underrepresented phyla remain mainly uncultured and are notably underexplored for metabolic potential (58 of 123 phyla,  $n = 131$  genomes). Our analyses revealed that representatives of these taxa contain HC degrading enzymes involved in both the initiation and downstream steps of HC degradation processes. For example, phyla Tectomicrobia (*Entotheonella*), Binatota, Firmicutes\_K, and Firmicutes\_E contained mono-aromatic HC degradation enzymes (Fig. 2). In addition to these phyla, we annotated enzymes involved in the degradation of aliphatic HC in representatives of phyla SAR324, Eremiobacterota (Baltobacteriales), Bdellovibrionota\_B, and Chloroflexota\_B (Fig. 1).

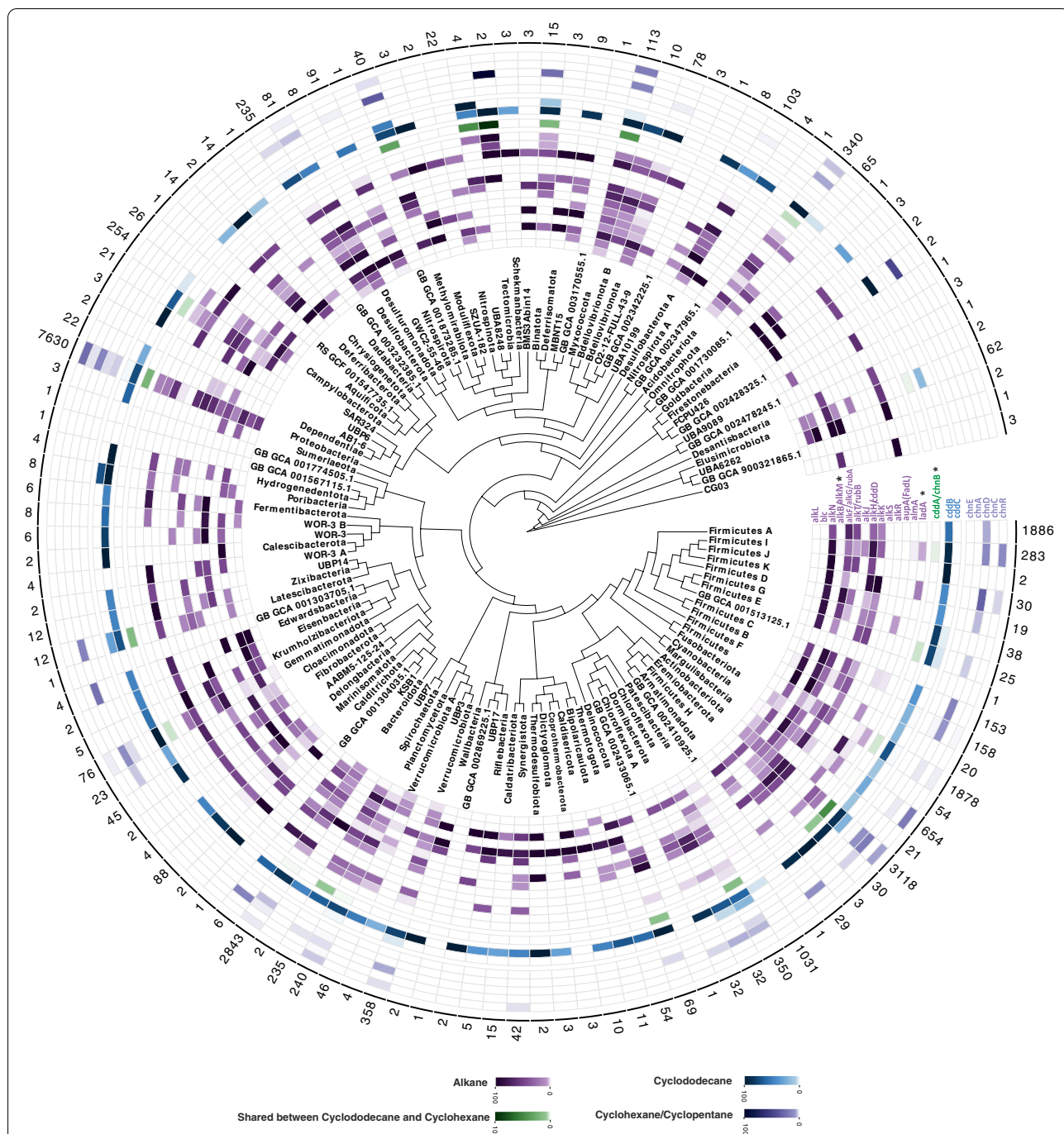
Other enzymes in the degradation pathways beyond the key genes for the initial degradation (Supplementary Table S1) are typically involved in several degradation pathways and are broadly distributed accordingly. As an example, the process of converting catechol to non-aromatic compounds with further conversion to intermediates of the TCA cycle (e.g., acetaldehyde and pyruvate) (Supplementary Figure S1) is shared among degradation pathways of xylene, naphthalene/phenanthrene, and phenol (blue color in Fig. 2). The name of genes involved in

mentioned part are xylE/dmpB/nahH (catechol oxygenase), xylF/dmpD/nahN (2-hydroxymuconate semialdehyde hydrolase), xylG/dmpC/nahI (2-hydroxymuconic semialdehyde dehydrogenase), xylH/nahJ (2-hydroxymuconate tautomerase), xylI/dmpH/nahK (4-oxalocrotonate decarboxylase), xylJ/nahL (2-hydroxypent-2,4-dienoate hydratase), xylK/bphI/nahM (4-hydroxy-2-oxovalerate aldolase), xylR/pdeR (transcriptional regulatory protein). These ring-cleavage enzymes are also involved in the degradation of aromatic amino acids. Our analysis showed that representatives of phyla Firmicutes (mainly from the orders Bacillales and Staphylococcales), Firmicutes\_I, Firmicutes\_K, Firmicutes\_E, Firmicutes\_G, Firmicutes\_H, Eremiobacterota, Deinococcota, Chloroflexota, Campylobacterota, Myxococcota and Bdellovibrionota play a significant role in this part of HC degradation process (blue color in Fig. 2).

#### Distribution of key genes involved in the degradation of alkanes

At lower taxonomic rank, the alkB/M and ladA genes were differently distributed across members of phyla Gammaproteobacteria, Alphaproteobacteria, and Actinobacteriota, hinting at their capacity for degrading hydrocarbons of variable chain length. Altogether 2089 genomes in orders Mycobacteriales (23.95%), Rhodobacterales (20.46%), Pseudomonadales (17.13%), Flavobacteriales (8.3%), Burkholderiales (6.16%), *Cytophagales* (3.66%), Propionibacteriales (2.47%), Rhizobiales (1.89%), and Chitinophagales (1.81%) contained alkB/M genes, while ladA was present in 2154 genomes from Pseudomonadales (21.05%), Rhizobiales (16.27%), Burkholderiales (14.44%), Actinomycetales (13.44%), Mycobacteriales (13.05%), Bacillales (4.74%), Enterobacteriales (3.7%), Acetobacteriales (2.31%), Streptomycetales (1.91%). There were also several representative genomes with ladA gene which have not been previously reported and belonged to Tectomicrobia, UBA8248, and SAR324 phyla (Figs. 3 and 4, panel B, Supplementary Table S6).

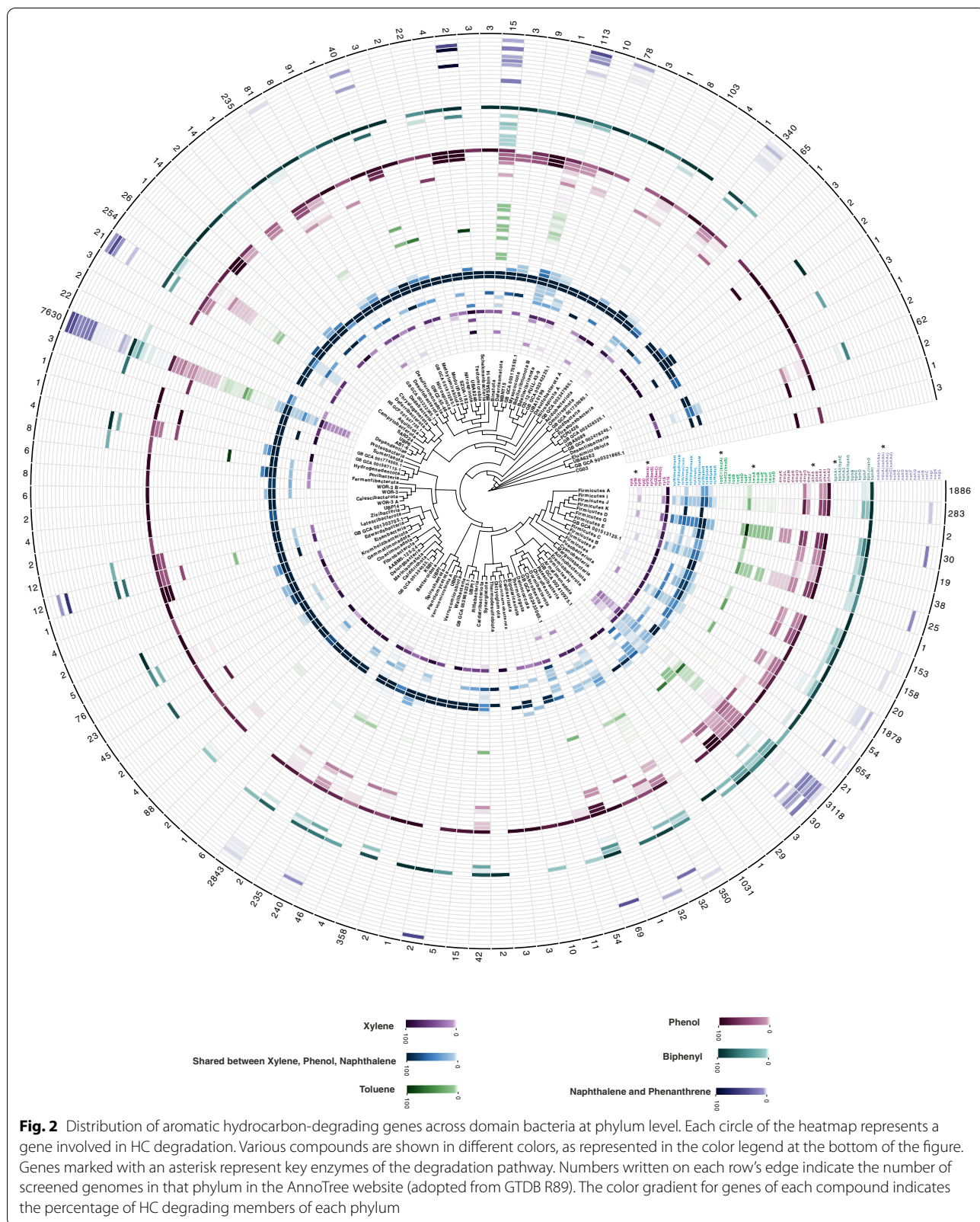
An indirect role of Cyanobacteria in HC degradation, especially in microbial mats, has been previously reported. These primary producers often have the nitrogen-fixing ability and can fuel and promote aerobic and anaerobic sulfate/nitrate-reducing HC degrading microorganisms in microbial mats [13]. There are also reports of a minor role of some Cyanobacteria members like *Phormidium*, *Nostoc*, *Aphanothece*, *Synechocystis*, *Anabaena*, *Oscillatoria*, *Plectonema*, and *Aphanocapsa* in direct HC degradation [14, 15]. In this study, we detected the presence of long-chain alkane degrading genes, ladA, in different members of Cyanobacteria with 0.31 and 12.54% of genomes in this phylum containing ladA (in *Elainella saxicola*, *Nodosilinea sp000763385*) and almA

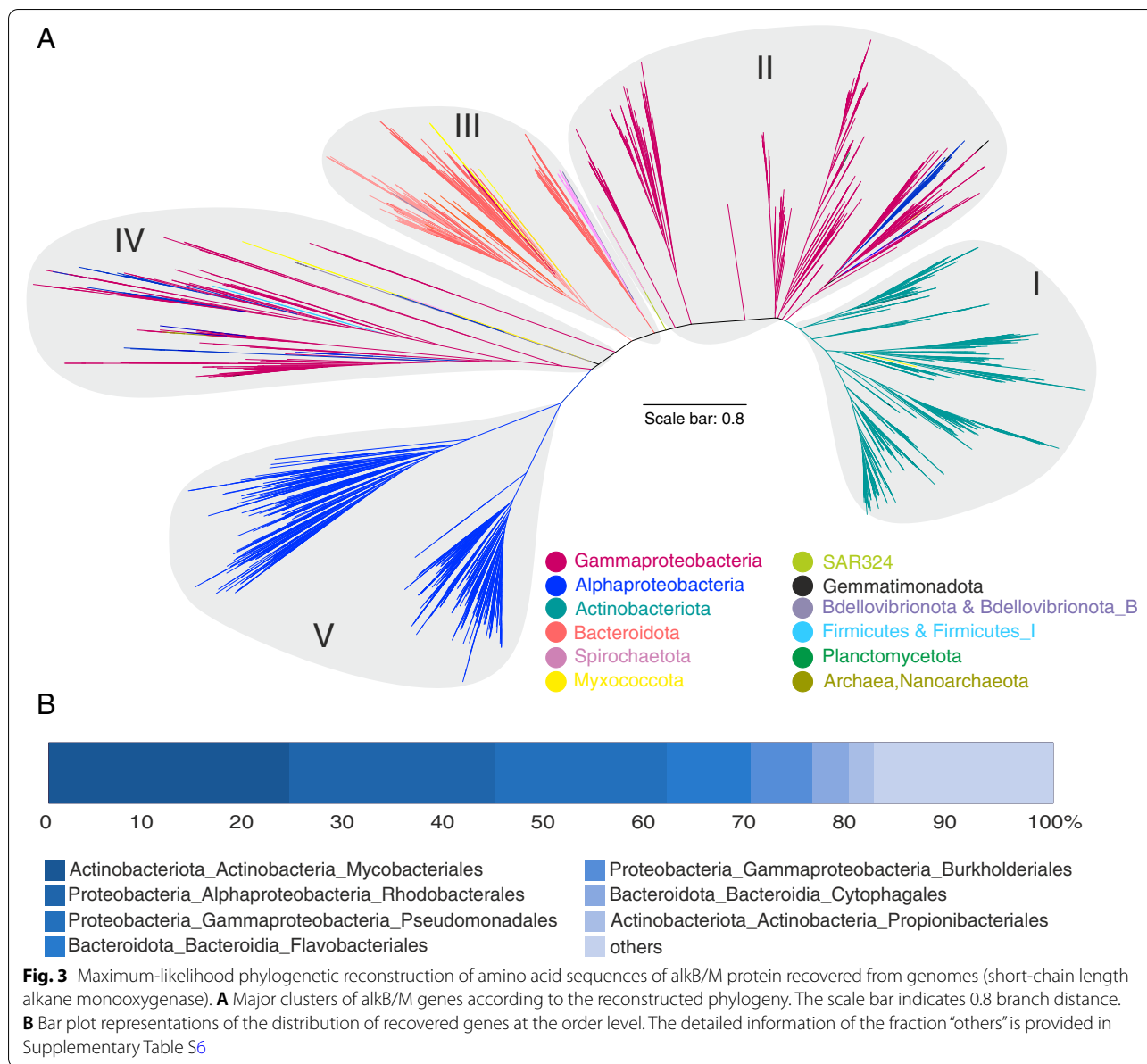


**Fig. 1** Distribution of aliphatic hydrocarbon-degrading genes across domain bacteria at phylum level. Each circle of the heatmap represents a gene involved in HC degradation. Various compounds are shown in different colors, as represented in the color legend at the bottom of the figure. Genes marked with an asterisk represent key enzymes of the degradation pathway. Numbers written on each row's edge indicate the number of screened genomes in that phylum in the Annotree website (adopted from GTDB R89). The color gradient for genes of each compound indicates the percentage of HC degrading members of each phylum

genes (in Synechococcales, Cyanobacteriales, Elainellales, Phormidemiales, Thermosynechococcales, Gloeobacteriales, Obscuribacteriales), respectively.

Phylogenetic reconstruction of recovered alkB/M and ladA genes grouped them into five and nine main clades, respectively (Figs. 3 and 4, panel A). The branching

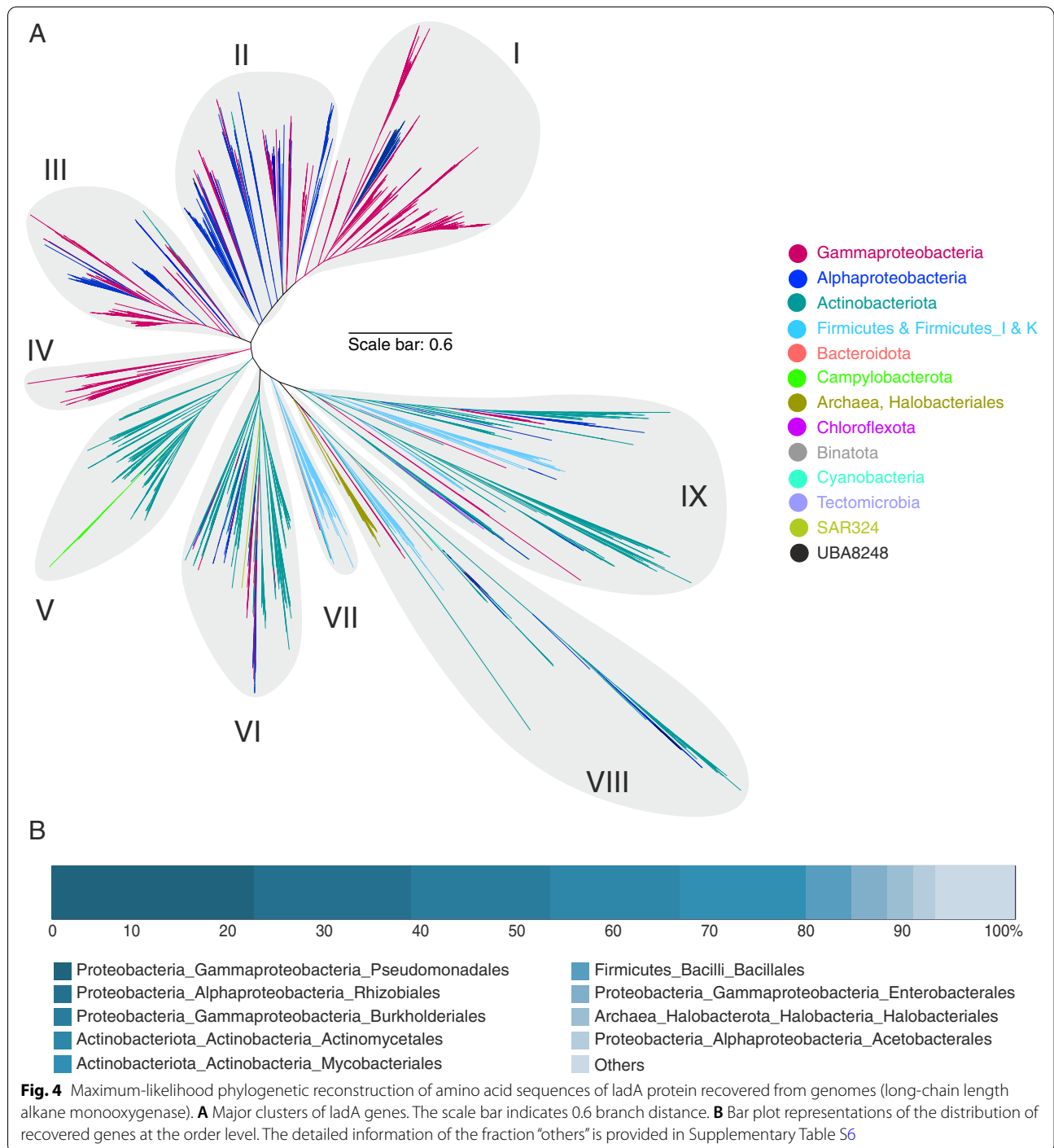




pattern of these clades partially followed the taxonomic signal of the genomes they were retrieved from, specifically for the most dominant phyla. However, as it is evident in Figs. 3 and 4 some branches contained alkB/M and ladA genes originating from genomes belonging to distantly related phyla. The placement of phylogenetically diverse groups in one branch hint at horizontal transfer of these genes between microbial taxa [16]. Additionally, apart from the chromosomal type, both alkB/M and ladA genes have previously been reported to be also located on plasmids (OCT and pLW1071), corroborating their potential for horizontal transfer. For instance, there are reports on the intraspecies transfer of alkB/M among *Pseudomonas* members [17]. Placement of ladA gene

originating from genomes affiliated to rare microbial taxa among clusters V-IX of the ladA phylogeny put forward the possibility of a prominent role for Actinobacteriota and Firmicutes members in expanding the distribution of this gene among rare taxa as representatives of Actinobacteriota and Firmicutes taxa frequently contain these genes (Fig. 4).

We also detected several genomes with multiple copies of the alkB gene that were not necessarily branching together in the reconstructed alkB phylogeny, hinting at the probability of either gene duplication, paralogue occurrence, or HGT. Examples of these genomes with more than 6 copies of alkB/M are *Polycyclovorans algicola* (10), *Nevskia ramose* (7), *Zhongshania*



*aliphaticivorans* (7), *Solimonas aquatic* (7), *Immundisolibacter cernigliae* (6), and *Rhodococcus qingshengii* (6). Multiple copies have also been detected in representatives of the genera *Nocardia*, *Rhodococcus*, and *Alcanivorax* (the full list of genomes with multiple copies of *alkB* gene is available in Supplementary Table S6).

Furthermore, the *ladA* gene was also detected in *Mycolicibacterium dioxanotrophicus*, *Cryobacterium\_A sp003065485*, *Kineococcus rhizosphaerae*, *Microbacterium sp003248605*, *Paenibacillus\_S sp001956045*, *Pararhizobium polonicum*, *Mycolicibacterium septicum*, and *Microbacterium sp000799385* with six copies in each

genome. Several examples were also present in genera *Pseudomonas*, *Bradyrhizobium*, *Rhizobium*, and *Paraburkholderia*, which had more than one copy (904 genomes) (Supplementary Table S6).

The presence of multiple copies of alkane hydroxylase genes has been hypothesized to enable cells to use an expanded range of n-alkanes or to adapt to different environmental conditions. However, the exact evolutionary rationale has not yet been established [18, 19]. To evaluate this hypothesis, we compared different sequences of each gene in an individual genome (mentioned above for *ladA* and *alkB*) using BLAST (Supplementary Table S7). The results showed that the identity of multiple gene copies in a single genome was in the range of 30 to 70 percent, while they are still predicted to have the same function. This wide range of BLAST identity between these gene copies suggests that these genes potentially originated from different sources and were transferred horizontally. We further explore in more detail the case for 21 *xylX* gene copies in *Immundisolibacter cernigliae* below.

#### Distribution of key genes (ring-hydroxylating oxygenases (RHOs)) involved in the degradation of aromatic HCs

Genomes containing RHOs (2761 genomes, 16 phyla) present an overall lower phylogenetic diversity than alkane mono-oxygenases (4669 genomes, 21 phyla for both *alkB/M* and *ladA*). In general, *AlkB/M* and *LadA* enzymes consist of *FA\_desaturase* (PF00487) and *Bac\_luciferase-like mono-oxygenase* (PF00296) domains, respectively (Supplementary Table S5). They act non-specifically on a wide range of alkanes of different chain lengths. Therefore, they are likely to be more widespread in genomes, especially because alkane compounds do not exclusively originate from petroleum. For instance, in pristine marine ecosystems, primary producers such as Cyanobacteria can release long chain-length aliphatic compounds (e.g., pentadecane, heptadecane). Alkane-producing Cyanobacteria include prominent and globally abundant genera such as *Prochlorococcus* and *Synechococcus* [20, 21]. Therefore, marine microorganisms are broadly exposed to aliphatic compounds with different chain lengths, even in environments without oil spills or industrial influence. This can explain why marine ecosystems host a plethora of hydrocarbonoclastic bacteria [22, 23]. This would probably be the reason why the number of genomes with *alkB/M* is higher than RHO-bearing genomes.

Enzymes *XylX*, *NdoB*, *BphA1*, and *TodC1* are composed of two pfam domains, PF00355 (Rieske center) and PF00848 (*Ring\_hydroxyl\_A*). These common domains impact the branching in the phylogenetic tree and lead

to the neighboring branching of these mentioned genes (Fig. 5).

RHO enzymes are predominantly present in Burkholderiales, Pseudomonadales, Sphingomonadales, Caulobacteriales, and Nevskiales orders of the phylum Proteobacteria (35 different Proteobacterial orders) (Fig. 5, B part). However, a significant number of *pheA* (phenol 2-monooxygenase) gene and, to a lesser degree, *xylX* and *tmoA* genes were also present in Actinobacteriota phylum (9 different Actinobacteriota orders) (Fig. 5, B part).

Sphingomonadales are prominent bacteria in the rhizosphere and are also abundant in littoral zones of inland waters. Accordingly, we suggest that these bacteria may have evolved a capacity to degrade different aromatic compounds in response to the high concentrations of aromatic secondary metabolites typically seen in the plant rhizosphere. Additionally, Sphingomonadales are known for their large plasmids with intraspecies transmission [24].

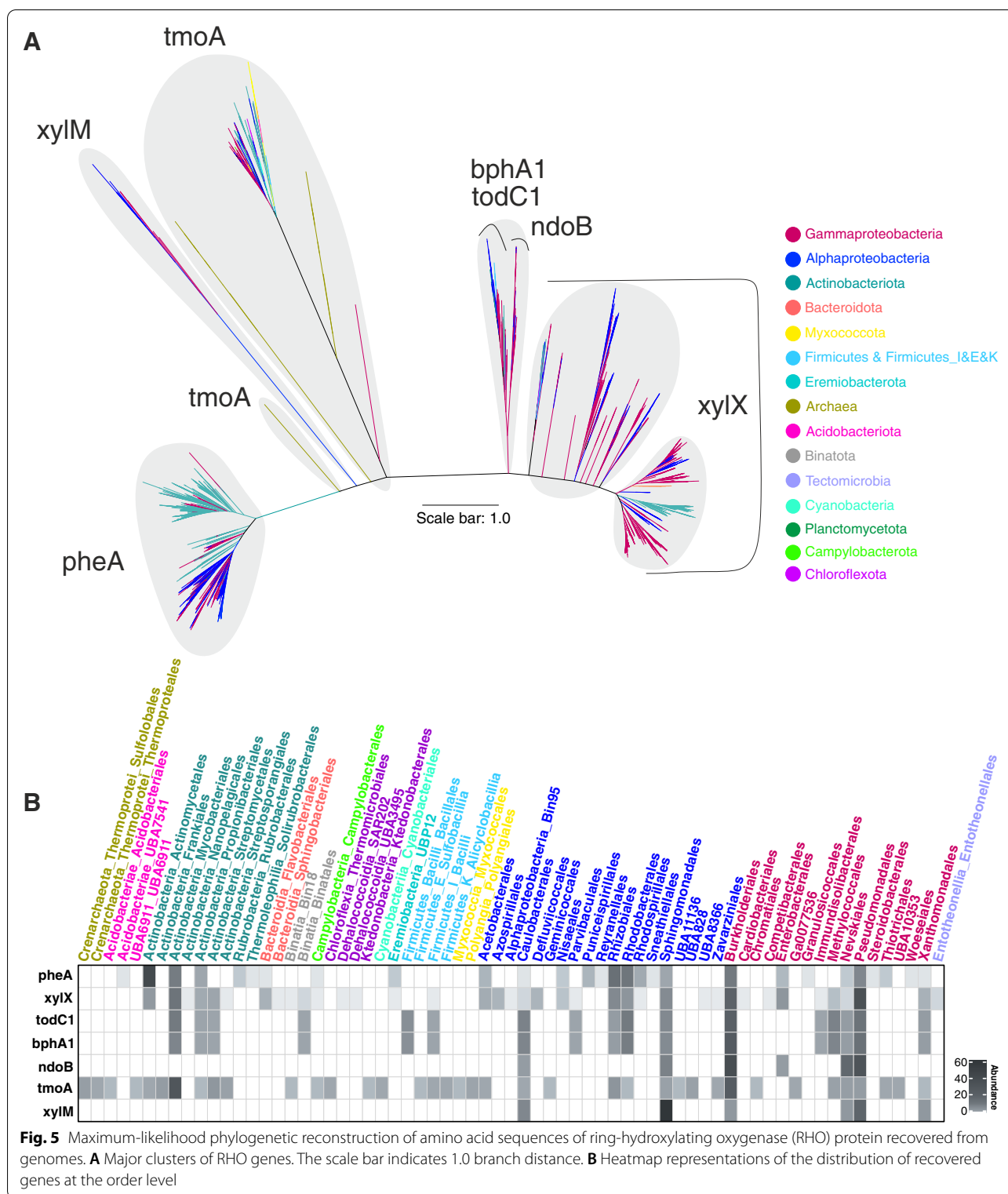
Representatives of Burkholderiales were one of the other prominent members with RHO genes. This potential was mainly attributed to *Paraburkholderia*, *Caballeronia*, *Burkholderia*, *Cupriavidus*, *Ralstonia* and *Massilia* genera. All of the mentioned genera in our study had a genome size between 6.5–7 Mb and have been reported to have multi-chromosomes (multipartite) [25]. According to the reports, the second chromosome has a prominent role in niche adaptation in *Burkholderiaceae* family [25]. Their HC degradation capability might be related to this feature.

Among Pseudomonadales, representatives of *Pseudomonas*, *Acinetobacter*, *Halomonas*, *Marinobacter*, *Marinobacterium* and *Psychrobacter* genera were detected to contain RHO genes in their genome. There are multiple reports of representatives of the *Pseudomonas* genus to degrade various HCs, where they mainly organize their degrading genes on conjugal plasmids [26, 27].

*Acinetobacter* representatives also had RHO genes. *Acinetobacter* spp. can produce bioemulsifier/biosurfactants and are naturally transformable species that might be relevant for their high degradation potential [28, 29]. *Marinobacter* and *Halomonas* genera were frequently reported as hydrocarbonoclastic bacteria which were isolated from oil-polluted marine water and sediment samples. Representatives of the *Halomonas* genus are also capable of EPS/biosurfactant production that accelerates the degradation of HC compounds, especially under saline conditions [30–32].

In our study the main RHO-bearing genome in the Caulobacteriales order belonged to *Hyphomonas* genus. Representatives of this genus have been isolated from marine aromatic hydrocarbon polluted sites with the capability





**Fig. 5** Maximum-likelihood phylogenetic reconstruction of amino acid sequences of ring-hydroxylating oxygenase (RHO) protein recovered from genomes. **A** Major clusters of RHO genes. The scale bar indicates 1.0 branch distance. **B** Heatmap representations of the distribution of recovered genes at the order level

to degrade HCs [33, 34]. Nevskiales aromatic degradation potential was mainly detected in Nevskia, Polycyclovorans, and Solimonas genera. A Nevskia representative

was reported to have biosurfactant production ability and Polycyclovorans representatives were shown to prefer HC as a nutrient compared to glucose [35, 36].

Among all investigated RHO genes, the highest phylogenetic diversity was observed in *tmoA* (208 genomes in 12 phyla and 38 orders) and *xylX* (1486 genomes in 9 phyla and 38 orders) genes (Fig. 5, B part). In the case of *tmoA* gene, it might be due to the wide range of HC compounds susceptible to this enzyme (e.g., benzenes, some PAHs, and alkenes) [37, 38]. Therefore, in our survey here we see that a diverse set of genera harbor *tmoA* gene and could potentially degrade different types of HCs (Fig. 5, B part).

Underrepresented microbial groups with a limited number of RHO genes also featured *tmoA*, *xylX*, and *pheA* genes. Myxococota, Acidobacteriota, Chloroflexota, Firmicutes\_I,E,K, and Cyanobacteria with *tmoA* gene were clustered separately, reflecting their distinct protein sequence and the lower possibility of HGT among these groups. For *xylX*, Eremiobacterota affiliated genes were placed together with genes from Gammaproteobacteria, and Tectomicrobia, Binatota, Chloroflexota, and Firmicutes\_I were placed in separate branches near Actinobacteriota. In addition, Acidobacteriota, Eremiobacterota, and Campylobacteria with *pheA* gene were nested within Alphaproteobacteria members. The phylogeny of RHO genes was also more consistent with taxonomy than the phylogeny of *alkB/M* and *ladA*.

Bionatota, a recently described phylum shown to be efficient in HC degradation, harbored *todC1*, *bphA1* (in Binatales order), and *xylX* (Bin18, Binatales) genes from RHOs and *ladA* (in Bin18) from alkane hydroxylases. Representatives of this phyla have been reported to play a role in methane and alkane metabolism [39]. However, we also noted the further potential of Binatales and Bin18 orders of this phylum in aromatic HC degradation.

RHOs can be located either on the chromosome or plasmid, depending on the organism. For instance, *todC1*, *bphA1*, and *tmoA* genes were reported to be on the chromosome [40], while in another study, they were detected on a plasmid [41]. Other RHOs, including *xylX*, *xylM*, *pheA*, and *ndoB* have mainly been reported to be hosted by plasmids [40, 38]. One of the main examples is TOL plasmid (carry genes for xylene and toluene degradation) which has been reported to be transferable between *Pseudomonas* spp. *Rhizobium* spp. and *Erwinia* spp. [42, 43].

Multiple copies of RHO genes in one genome were detected for *xylX* and *pheA*. *Immundisolibacter cernigliae* surprisingly contained 21 variants of *xylX*. This genome also had six copies of *alkB/M* and was isolated from a PAH-contaminated site [44]. The high HC degradation potential of other members of this genus has also been reported in the marine ecosystem [45, 46]. *Rugosibacter aromaticivorans* (containing 5, 2 and 2 copies of *xylX*, *ndoB*, and *tmoA* genes, respectively),

*Pseudoxanthomonas\_A spadix\_B* (with 4, 2 and 2 copies of *xylX*, *todC1* and *bphA1* genes, respectively), *Thauera sp002354895* (4), *Pigmentiphaga sp002188635* (4) are other examples of genomes that have multiple copies of the *xylX* gene. Although *xylX* gene was detected in Actinobacteriota, multiple copies in a genome were seen only among the *Proteobacteria* phylum.

The pairwise BLAST identity among 21 different variants of the *xylX* gene present in the *Immundisolibacter cernigliae* genome ranged between 35 to 81 percent. Among these 21 *xylX* copies, three sequences (i.e., *xylX* numbers 18, 19, and 22 in Supplementary Figure S7) showed higher BLAST identity percentages with the *xylX* gene of the *Rugosibacter* genus compared to other *xylX* copies present in the *Immundisolibacter cernigliae* genome itself (Supplementary Table S7 and Supplementary Figure S7). Several *xylX* copies of *I. cernigliae* (i.e., *xylX* numbers 10, 11, 13, and 15 in Supplementary Figure S7) had more edges than others in the network, and their interactions (Supplementary Figure S7, highlighted in red) represent their similarity with *xylX* copies of *Caballeronia*, *Sphingobium*, and *Pseudoxanthomonas*, *Pseudomonas*, and *Thauera* genera. In addition, *xylX* number 5 and 7 of *Immundisolibacter* had almost similar blast identities with the *xylX* gene of *Pigmentiphaga* genus and other *xylX* copies in *I. cernigliae* itself. These results suggest that copies of the *xylX* gene in *I. cernigliae* potentially originate from horizontal transfer from other microbial groups.

On the other hand, *Glutamicibacter mysorens* (4), *Enteractinococcus helveticum* (4), and many other genomes from the *Castellaniella*, *Kocuria*, and *Halomonas* genera, had several *pheA* copies in their individual genomes. To a lesser degree, *tmoA* gene was present in multiple copies in *Pseudonocardia dioxanivorans* (4), *Rhodococcus sp003130705* (3), *Amycolatopsis rubida* (3) and *Zavarzinia compransoris\_A* (3) genera.

While *bphA1* and *todC1* have different KO identifiers (Supplementary Table S1), our manual checks showed that they had the same conserved domain based on NCBI CD-Search [47]. We kept both annotations for cases where one gene was annotated with both KO identifiers. Previous studies also report similar homology and substrate specificity between *todC1* and *bphA1* [41].

*XylM*, as one of the enzymes mediating the initial steps in toluene/xylene degradation, showed the lowest abundance and phylogenetic diversity (27 genomes in 1 phylum and 6 orders). Toluene/benzene can generally be degraded through different routes and three of the most prevalent approaches were studied here. *XylX*, *TodC1*, and *TmoA* are the initial oxygenase enzymes of these three pathways. They are diverse in starting the degradation and composed of different domains, while

downstream degradation converges to catechol derivatives as intermediates. *xylM* can also initiate toluene degradation in addition to xylene. In this pathway *xylX* then converts produced benzoate to catechol. Therefore, while we see a lower diversity and abundance of genomes harboring *xylM*, we want to highlight the possibility of the presence of alternative degradation enzymes in different microorganisms that can degrade the same compound and initiate the degradation process.

As the number of rings in aromatic compounds increases, the number and diversity of microbial groups capable of degrading them decreases, and microbial groups with *ndoB* (naphthalene 1,2-dioxygenase) accordingly showed the lowest abundance after *xylM* gene. The genomes hosting *ndoB* had limited phylogenetic diversity (35 genomes in 1 phylum and 6 orders) and were found mainly in representatives of Alphaproteobacteria (Sphingomonadales (17) and Caulobacteriales (2)) and Gammaproteobacteria (Pseudomonadales (5), Burkholderiales (1), Nevskiales (1)).

#### Ecological strategy of HC degrading bacteria

Microorganisms are broadly divided into two main functional growth categories, i.e., oligotrophic/slow-growing/K-strategist or copiotrophic/fast-growing/r-strategist. These ecological strategies are associated with the genome size that, in turn, directly correlates with the GC content [48]. To get further insights into the ecological strategies of organisms that feature HC degrading genes, we compared the distribution of GC content and estimated genome size. This analysis revealed that HC degrading genes were present in genomes with a broad genome size range (1.34 to 16.9 Mb) and GC content (26.9 to 76.6%) (Supplementary Figure S8, data available in Supplementary Table S8). Genomes containing HC degrading enzymes with GC percent equal to or lower than 30% mainly had *alkB* gene and taxonomically belonged to Flavobacteriales order (genome sizes in the range of 1.4 to 4.2 Mb). The largest genome included in this study was *Minicystis rosea* from the phylum Myxococcota (genomes size of 16.9 Mb), which also contained *alkB*. In this phylogenetic reconstruction, the *alkB* gene of *Minicystis rosea* clustered together with homologs from representatives of phylum Gammaproteobacteria (*Immundisolibacter* and *Cycloclasticus* genera) (Fig. 3). The large genome size of *Minicystis rosea*, together with the phylogenetic placement of its *alkB* gene together with phylum Gammaproteobacteria representatives, hint at the horizontal transfer of this gene to the *Minicystis rosea* genome. These analyses indicate that HC degradation ability is present in both K-strategist and r-strategists microorganisms. Earlier studies have shown

that r-strategist serves as the principal HC degraders after oil spills and at other point sources of pollution in marine environments [49–51]. Indeed, most obligate hydrocarbonoclastic bacteria are r-strategists (Proteobacteria domain) and are mainly reported to be isolated from marine samples [52]. The r-strategists are adapted to live in oligotrophic environments with transient nutrient inputs and rapid consumption of substrates upon episodic inputs by means of fast growth and population expansion [53]. In contrast, studies on oil-polluted soil samples suggest a predominance of K-strategists, especially in the harsh conditions (High concentration of HC, soil dryness, etc.) commonly seen in many such soil environments [54–56]. Hosting multiple copies of genes coding for HC degrading enzymes seems to be a shared feature in both r- and K-strategists with small and large genome sizes alike and appears to be a universal evolutionary strategy for HC degradation.

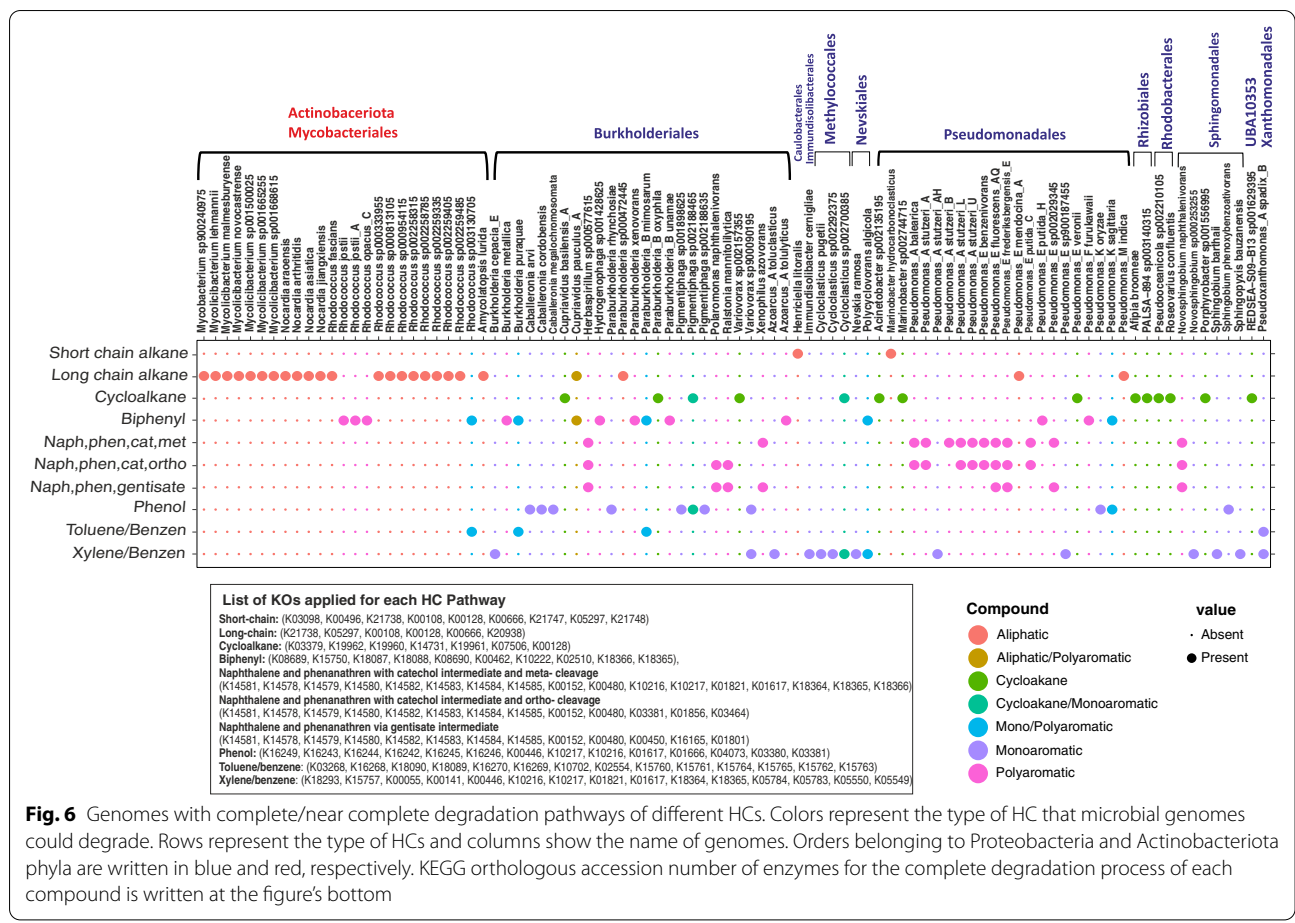
#### Genome-level analysis of HC degradation

Microorganisms are known to use division of labor or mutualistic interactions to perform HC degradation in the environment [57, 58]. However, 92 genomes (less than 0.5%) of 23,446 investigated bacterial genomes do in fact, contain all the enzymes required to degrade at least one HC compound completely. These 92 genomes all belong to Actinobacteriota ( $n=25$ ) and Proteobacteria ( $n=67$ ) (Fig. 6).

Microorganisms have evolved two pathways for naphthalene degradation that involve the production of either catechol or gentisate as aromatic degradation intermediate (Supplementary Figure S1). Catechol can in turn, be further degraded via meta- or -ortho cleavage. Several microorganisms, including *Novosphingobium naphthalenivorans*, *Pseudomonas\_E fluorescens\_AQ*, *Pseudomonas\_E frederiksbergensis\_E*, and *Herbaspirillum sp000577615*, feature both of the mentioned pathways and even have the genes to potentially perform ortho and meta cleavage simultaneously (Fig. 6).

Moreover, *Cupriavidus pauculus\_A* (long-chain alkanes and also biphenyl), *Cycloclasticus sp002700385* and *Paraburkholderia\_B oxyphila* (cycloalkane and xylene/benzene), *Pigmentiphaga sp002188465* (cycloalkane and phenol), *Rhodococcus sp003130705*, *Burkholderia puraquae*, and *Paraburkholderia\_B mimosarum* (toluene and biphenyl) contain the genetic potential to degrade more than one HC compound autonomously (Fig. 6).

Members of Burkholderiales had the genetic potential to degrade even more diverse compounds individually, while Actinobacteriota representatives mainly had the potential to contribute to the degradation of aliphatic compounds. This ability was also apparent in Figs. 1, 3,



**Fig. 6** Genomes with complete/near complete degradation pathways of different HCs. Colors represent the type of HC that microbial genomes could degrade. Rows represent the type of HCs and columns show the name of genomes. Orders belonging to Proteobacteria and Actinobacteriota phyla are written in blue and red, respectively. KEGG orthologous accession number of enzymes for the complete degradation process of each compound is written at the figure's bottom

and 4. The potential for autonomous HC degradation wasn't detected in genomes of more rare bacterial phyla. Moreover, none of the archaeal genomes investigated in this study contained all genes for the complete degradation of HCs.

**HC degradation across domain archaea**

Generally, HC degradation ability seems to be less prevalent among archaea as compared to bacteria. The phylum Halobacterota had the highest proportion of enzymes involved in the degradation of both aliphatic ( $n=14$  enzymes of aliphatic degradation pathway) and aromatic ( $n=25$  enzymes of aromatic degradation pathway) compounds among the studied archaea (Supplementary Figure S9). The alkB enzyme, responsible for short-chain alkane degradation, was detected in two copies in a single member of the phylum Nanoarchaeota (*ARS21 sp002686215*). This gene was clustered together with alkB identified in Gammaproteobacteria representatives (*GCA-002705445* order) (Fig. 3). Genes needed to initiate degradation of long-chain alkanes and cyclododecane/cyclohexane as well as cyclopentane degradation via ladA and cddA/

chnB genes were more prevalent among Halobacterota representatives (75 genomes in 7 families; *Haloferacaceae*, *Haloarculaceae*, *Natrialbaceae*, *Halococcaceae*, *Halalkalicoccaceae*, *Haloadaptaceae*, and *Halobacteriaceae*) (Figs. 4 and Supplementary Figure S9). Among investigated RHOs, only tmoA that initiates toluene degradation was present in 5 Sulfolobales and 2 Thermoproteales genomes of the phylum Crenarchaeota (Fig. 5). Detected archaeal tmoA and ladA genes branched separately from bacteria in the phylogenetic trees (Figs. 4 and 5). Apart from alkB, gene duplications were present in several genomes for both tmoA (*Sulfolobus* and *Acidianus* genera) and ladA (*Halopenitus persicus* and *Halopenitus malekzadehii*).

Key enzymes needed to initiate HC degradation were rarely present in archaea (Figs. 3, 4, and 5), indicating that Archaea might not play a significant role in the typically rate-limiting initial degradation of HCs. However, several studies report the ability of halophilic archaeal isolates (e.g., *Halorubrum* sp., *Halobacterium* sp., *Haloferax* sp., *Haloarcula* sp.) to degrade both aliphatic (n-alkanes with chain lengths up to C18 and longer) and aromatic (e.g., naphthalene, phenanthrene, benzene, toluene and

*p*-hydroxybenzoic acid) HCs and use them as their sole source of carbon [59–61]. This may imply that archaea carry alternative and hitherto unknown enzymes for triggering HC degradation. However, there is no complete genome information available for the mentioned isolates to screen them for the presence of alternative degrading enzymes [11]. The *Haloferax* sp., capable of using a wide range of HCs as its sole source of carbon, present in the AnnoTree database (RS\_GCF\_000025685.1), contained none of the key degrading genes. The AnnoTree website chooses representative genomes having completeness of higher than 90%, which reduces the likelihood of incompleteness of the studied genome as a reason for the absence of these genes. Therefore, alternative HC degrading genes that are present in the accessory part of the genomes might be responsible for the observed degradation.

On the other hand, the recent reconstruction of three metagenome-assembled Thermoplasmatota genomes (Poseidonia, MGIIa-L2, MGIIb-N1) from oil-exposed marine water samples (not included in the GTDB release89) contained enzymes involved in alkane (alkB) and xylene (xylM) degradation [46]. Hence as these global genome depositories continue to expand, we may have to revise or update our findings.

A total number of 597 archaeal genomes contain enzymes involved in the degradation of aromatic compounds regarding the conversion of catechol to TCA intermediates. This is observed in the phyla Halobacterota (176 genomes in *Haloferacaceae*, *Haloarculaceae*, *Natrialbaceae*, *Halococcaceae*, *Halobacteriaceae*, *Methanocullaceae*, *Methanoregulaceae*, *Methanosarcinaceae*, *Archaeoglobaceae*, and some other methano-prefixed families), Thermoplasmatota (175 genomes in Poseidoniales, Marine Group III, Methanomassiliicoccales, UBA10834, Acidiprofundales, DHVEG-1, UBA9212), and Crenarchaeota (110 genomes in Nitrosphaerales, Desulfurococcales, Sufolobales, Thermoproteales). This widespread capacity for degrading downstream intermediates in aromatic HC degradation implies that archaea interact closely with bacteria in HC degradation.

## Conclusions

HCs are ubiquitously distributed in the biosphere and do not exclusively originate from oil. In this study, the distribution of genes of key HC degrading enzymes involved in the degradation of certain HCs (aliphatic and aromatic types) is provided at genome resolution for both archaeal and bacterial domains. Over the last decades, extensive environmental genome and metagenome sequencing has significantly increased the number of available microbial genomes and enriched contemporary genomic databases. The genome-based taxonomy using average nucleotide

identity (ANI) or relative evolutionary divergence adopted by the Genome Taxonomy Database; GTDB [62, 63] as a reproducible method has in parallel revised and updated some taxonomic ranks. The order Oceanospirillales, as an example, is a well-known taxon in the marine oil degradation context, and its representatives have been frequently reported as one of the main HC degrading members in response to oil pollution [52, 64, 65]. Nonetheless, this taxonomic rank has been removed from the genome-based taxonomy, and its members have been mainly placed in the order Pseudomonadales [66]. This could potentially cause a communication gap between the existing literature and new research. An updated comprehensive metabolic survey of Bacteria and Archaea for HC degradation potential at genome resolution could thus help bridge this gap. Our extensive survey shows that a greater diversity of bacteria contain genes involved in aliphatic HC degradation compared to aromatic HCs. Few genomes were detected to contain all necessary genes to carry out complete degradation pathways. This reiterates previous findings that microbes generally cooperate for HC degradation by “division of labor” and a community perspective would therefore be crucial for predicting the fate of oil HCs in the ecosystem. According to our results and prevalence of genes for known HC degradation enzymes, archaea could potentially make a small contribution to aerobic HC degradation however, based on genes they carry they could play a significant role in peripheral routes to degrade intermediate compounds produced by bacterial community. We detected HC degradation ability among both *r* and *K* strategists and found signals of gene duplication and horizontal transfer of key HC degrading genes among microbes. This could be an efficient way to increase degradation capability among microbial members and potentially help them adapt to the available pool of HCs in their ecosystem.

## Materials and methods

### Data collection of HC degrading enzymes

Representative compounds from each category of HCs, including saturated aliphatic (short/long-chain alkanes) and alicyclic (cyclohexane/cyclododecane), compounds with mono-aromatic (toluene, phenol, xylene, and benzene), and poly-aromatic (PAHs) (naphthalene, phenanthrene, and biphenyl as representatives) hydrocarbons were selected to survey the distribution of Bacteria and Archaea capable of their degradation under aerobic conditions.

A complete list of enzymes involved in the degradation pathway of mentioned HCs was compiled from previous reports [67–74]. We explored these enzymes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [75], Pfam [76], TIGRFAMs [77], InterPro [78], and UniProt [79]

databases. The accession number of enzymes in each mentioned database, their function, name, reaction (if available), EC number, and additional information are provided in Supplementary Table S1. The workflow overview is provided in Supplementary Figure S10. The CYP153 gene is another essential alkane hydroxylase that is less prevalent but present in HC degrading microorganisms lacking alkB/M [80]. However, there were some limitations in detecting this gene in silico via AnnoTree. A proper model to search for this gene has not been developed (no KEGG/TIGRFam accessions), and the only way to search for this gene through the AnnoTree web server was to use a Pfam domain (PF00067). However, the PF00067 domain is a universal domain present in every CYP (cytochrome p450) gene, making it difficult (almost impossible) to differentiate between potential substrates and confirm whether it is involved in alkane degradation. Due to this limitation, we decided to skip this enzyme and only work on the groups we could confidently annotate.

#### Distribution of HC degrading enzymes among bacterial and archaeal representative genomes

The distribution of the compiled HC degrading enzymes described in Supplementary Table S1 was assessed across domains Bacteria and Archaea using AnnoTree (<http://annotree.uwaterloo.ca>) [12]. AnnoTree database is providing functional annotations for 24,692 genome representatives in the genome taxonomy database (GTDB) release 89. The phylogenetic classification of genomes is derived from the GTDB database (release R89). In total, the annotation information for 18, 10, and 90 enzymes involved in the degradation process of alkane, cycloalkane, and aromatic HCs, respectively, were analyzed. Genome hits were collected at the thresholds of percent identity  $\geq 50$ , e-value cut off  $\leq 1e^{-5}$ , subject/query percent alignment  $\geq 70$  for KEGG annotations, and e-value cut off  $\leq 1e^{-5}$  for Pfam and TIGRFAMs annotations. For each HC degrading enzyme, we first checked KEGG annotations. If there were no KEGG accession numbers for the enzyme, the second priority was TIGRFAMs; otherwise, the Pfam annotation was considered. The table contains information for the distribution of HC degrading enzymes of each pathway present in representative genomes from bacteria and archaea domains, as is shown in Supplementary Table S2 and Supplementary Table S3, respectively.

#### Phylogeny of bacteria and archaea augmented with the abundance of HC degrading enzymes

Evolview, a web-based tool for phylogenetic tree visualization, management, and annotation, was used to present

the distribution view of HC degrading enzymes in representative genomes across bacterial/archaeal phylogenomic trees [81, 82].

The phylogenomic tree of bacteria and archaea in the Newick format, at the phylum level (123 and 14 leaves, respectively), was adopted from the AnnoTree website (November 21<sup>st</sup>, 2020). Trees were uploaded as the reference tree in Evolview. According to the abundance tables of HC degrading enzymes prepared for each degradation pathway, four heatmaps were plotted for bacteria and archaea domains (separately for aliphatic and aromatic compounds).

#### Single gene phylogeny

To provide the evolutionary history of key enzymes in each HC degradation pathway, the protein sequence of that enzyme was manually confirmed by inspecting their conserved domains using the NCBI web CD-Search tool (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) [47]. In this step, we have uploaded all detected genes retrieved from AnnoTree to the NCBI web CD-Search tool and manually checked their annotations and inspected their annotated domains. Validated amino acid sequences were then aligned using Kalign3 software [83], and their phylogenetic tree was reconstructed using FastTree2 [84].

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08906-w>.

**Additional file 1.**

**Additional file 2.**

**Additional file 3.**

**Additional file 4.**

**Additional file 5.**

**Additional file 6.**

**Additional file 7.**

**Additional file 8.**

**Additional file 9: Supplementary Figure S1.** Schematic representation of HC degradation pathways studied in this work. Purple circles show key HC degrading enzymes triggering the degradation. Blue circles are other crucial enzymes. Important intermediate compounds are written in blue. **Supplementary Figure S2.** Distribution of 143512 genomes of the GTDB database release 89 in different phyla. **Supplementary Figure S3.** Distribution of aliphatic hydrocarbon-degrading genes across domain bacteria at the phylum level. In plot A, the color gradient indicates the proportion of degrading members of each phylum to the entire HC degrading community. In plot B, the color gradient shows the percentage of HC degrading members of each phylum. Columns are the name of genes involved in HC degradation, which key ones are represented in red. **Supplementary Figure S4.** Distribution of aromatic hydrocarbon-degrading genes across domain bacteria at the phylum level. In plot A, the color gradient indicates the proportion of degrading members of each phylum to the entire HC degrading community. In plot B, the color gradient shows the percentage of HC degrading members of each phylum. Columns are the name of genes involved

in HC degradation, which key ones are represented in red. Enzymes written in blue are shared among the degradation processes of different aromatic compounds (xylene, phenol and naphthalene). **Supplementary Figure S5.** Distribution of aliphatic hydrocarbon-degrading genes across domain archaea at the phylum level. In plot A, the color gradient indicates the proportion of degrading members of each phylum to the entire HC degrading community. In plot B, the color gradient shows the percentage of HC degrading members of each phylum. Columns are the name of genes involved in HC degradation, which key ones are represented in red. **Supplementary Figure S6.** Distribution of aromatic hydrocarbon-degrading genes across domain archaea at the phylum level. In plot A, the color gradient indicates the proportion of degrading members of each phylum to the entire HC degrading community. In plot B, the color gradient shows the percentage of HC degrading members of each phylum. Columns are the name of genes involved in HC degradation, which key ones are represented in red. Enzymes with blue color are shared among the degradation processes of different aromatic compounds (xylene, phenol and naphthalene). **Supplementary Figure S7.** Network interaction between 18 copies of *xylX* gene in *Immundisolibacter cernigliae* and other genomes with more than two copies of this gene. Only the blast identity values between 50 to 100 percent are shown. Edges are color-coded based on their blast identity. The size of nodes is based on the "Degree," which is determined by the number of edges of each node. Edges in red are versions of *xylX* in *Immundisolibacter cernigliae* that had a higher degree than others. The gene ID of the assigned number of each node is represented in Supplementary Table S7. **Supplementary Figure S8.** Distribution of genome size versus GC content of the studied genomes with key HC degrading genes. **Supplementary Figure S9.** Distribution of aliphatic (A) and aromatic (B) hydrocarbon-degrading genes across domain archaea at the phylum level. Columns show the name of genes involved in HC degradation and are represented in different colors for various compounds. The color gradient for genes of each compound indicates the percentage of HC degrading members of each phylum. **Supplementary Figure S10.** The overview of workflow that has been done in the present study.

#### Acknowledgements

The computational analysis was performed at the Center for High-Performance Computing, School of Mathematics, Statistics, and Computer Science, University of Tehran.

#### Authors' contributions

M.M. devised the study. M.R.S., L.G.M., and M.M., performed the bioinformatics analysis. M.R.S. and M.M. interpreted the data with input from S.M.M.D., S.B., M.A.A., and M.S.. M.R.S. and M.M. drafted the manuscript. All authors read and approved the manuscript.

#### Authors' information

Not applicable.

#### Funding

Open access funding provided by Swedish University of Agricultural Sciences.

#### Availability of data and materials

All genomes used in this study are downloaded from publicly available material deposited in genebank. All the accession numbers and detected annotations are shown in the supplementary material.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

Author declare no conflict of interest.

#### Author details

<sup>1</sup>Extremophile Laboratory, Department of Microbiology, School of Biology, College of Science, University of Tehran, Tehran, Iran. <sup>2</sup>Biotechnology Research Group, Research Institute of Petroleum Industry, Tehran, Iran. <sup>3</sup>Laboratory of Complex Biological Systems and Bioinformatics (CBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. <sup>4</sup>Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), Box 7050, 75007 Uppsala, Sweden.

Received: 20 July 2022 Accepted: 26 September 2022

Published online: 06 October 2022

#### References

- Abdel-Aal HK, Aggour M, Fahim MA. Petroleum and gas field processin. 2003.
- Speight JG. Origin and occurrence. In: Speight JG, Heinemann H, editors. The Chemistry and Technology of Petroleum, Edition, F. CRC Press, Taylor & Francis Group; 2014.
- Liu Y-F, Qi Z-Z, Shou L-B, Liu J-F, Yang S-Z, Gu J-D, et al. Anaerobic hydrocarbon degradation in candidate phylum "Atribacteria"(JS1) inferred from genomics. ISME J Nature Publishing Group. 2019;13:2377–90.
- Liu Q, Tang J, Bai Z, Hecker M, Giesy JP. Distribution of petroleum degrading genes and factor analysis of petroleum contaminated soil from the Dagang Oilfield. China Sci Rep Nature Publishing Group. 2015;5:1–12.
- Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds—from one strategy to four. Nat Rev Microbiol Nature Publishing Group. 2011;9:803–16.
- Peixoto RS, Vermelho AB, Rosado AS. Petroleum-degrading enzymes: bioremediation and new prospects. Enzyme Res. 2011;2011:475193 Hindawi.
- Sierra-Garcia IN, de Oliveira VM. Microbial hydrocarbon degradation: efforts to understand biodegradation in petroleum reservoirs. Biodegrad Technol. 2013;10:55920 InTech.
- Xu X, Liu W, Tian S, Wang W, Qi Q, Jiang P, et al. Petroleum hydrocarbon-degrading bacteria for the remediation of oil pollution under aerobic conditions: a perspective analysis. Front Microbiol. 2018;9:2885 Frontiers.
- Varjani SJ. Microbial degradation of petroleum hydrocarbons. Bioresour Technol Elsevier. 2017;223:277–86.
- 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 Springer. 2015;71:220–8.
- McGenity TJ. Taxonomy. Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes: Springer; 2019.
- Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res Oxford University Press. 2019;47:4442–8.
- Cohen Y. Bioremediation of oil by marine microbial mats. Int Microbiol Springer. 2002;5:189–93.
- Ibraheem IBM. BIODEGRADABILITY OF HYDROCARBONS BY CYANOBACTERIA 1. J Phycol Wiley Online Library. 2010;46:818–24.
- Raghukumar C, Viparty V, David J, Chandramohan D. Degradation of crude oil by marine cyanobacteria. Appl Microbiol Biotechnol Springer. 2001;57:433–6.
- Rodrigues EM, Gerais M, Gerais M, Gerais M. Detection of horizontal transfer of housekeeping and hydrocarbons catabolism genes in bacterial genus with potential to application in bioremediation process. Open Access Libr J. 2018;5:1 Scientific Research Publishing.
- Phale PS, Shah BA, Malhotra H. Variability in assembly of degradation operons for naphthalene and its derivative, carbaryl, suggests mobilization through horizontal gene transfer. Genes (Basel). 2019;10:569 Multidisciplinary Digital Publishing Institute.
- Korshunova AV, Tourova TP, Shestakova NM, Mikhailova EM, Poltarau AB, Nazina TN. Detection and transcription of n-alkane biodegradation genes (alk B) in the genome of a hydrocarbon-oxidizing bacterium Geobacillus subterraneus K. Microbiology Springer. 2011;80:682–91.
- Hashmat AJ, Afzal M, Fatima K, Anwar-ul-Haq M, Khan QM, Arias CA, et al. Characterization of hydrocarbon-degrading bacteria in constructed wetland microcosms used to treat crude oil polluted water. Bull Environ Contam Toxicol Springer. 2019;102:358–64.

20. Kudo H, Hayashi Y, Arai M. Improving hydrocarbon production by engineering cyanobacterial acyl-(acyl carrier protein) reductase. *Biotechnol Biofuels BioMed Central*. 2019;12:1–13.
21. Lea-Smith DJ, Biller SJ, Davey MP, Cotton CAR, Perez Sepulveda BM, Turchyn AV, et al. Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proc Natl Acad Sci*. 2015;112:13591–6 National Acad Sciences.
22. Nie Y, Chi C-Q, Fang H, Liang J-L, Lu S-L, Lai G-L, et al. Diverse alkane hydroxylase genes in microorganisms and environments. *Sci Rep Nature Publishing Group*. 2014;4:1–11.
23. Love CR, Arrington EC, Gosselin KM, Reddy CM, Van Mooy BAS, Nelson RK, et al. Microbial production and consumption of hydrocarbons in the global ocean. *Nat Microbiol Nature Publishing Group*. 2021;6:489–98.
24. Kertesz MA, Kawasaki A, Stolz A. Aerobic Hydrocarbon-Degrading Alphaproteobacteria: Sphingomonadales. In: McGenity T, editors. *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*. Handbook of Hydrocarbon and Lipid Microbiology. Springer; 2019. p. 105–24.
25. diCenzo GC, Mengoni A, Perrin E. Chromids aid genome expansion and functional diversification in the family Burkholderiaceae. *Mol Biol Evol Oxford University Press*. 2019;36:562–74.
26. Izmalkova TY, Sazonova OI, Nagornih MO, Sokolov SL, Kosheleva IA, Boronin AM. The organization of naphthalene degradation genes in *Pseudomonas putida* strain AK5. *Res Microbiol Elsevier*. 2013;164:244–53.
27. Obayori OS, Salam LB. Degradation of polycyclic aromatic hydrocarbons: Role of plasmids. *Academic Journals: Scientific Research and Essays*. 2010;5:4093–106.
28. Mujumdar S, Joshi P, Karve N. Production, characterization, and applications of bioemulsifiers (BE) and biosurfactants (BS) produced by *Acinetobacter* spp.: a review. *J Basic Microbiol*. 2019;59:277–87 Wiley Online Library.
29. Jung J, Park W. *Acinetobacter* species as model microorganisms in environmental microbiology: current state and perspectives. *Appl Microbiol Biotechnol Springer*. 2015;99:2533–48.
30. Chernikova TN, Bargiela R, Toshchakov SV, Shivaraman V, Lunev EA, Yakimov MM, et al. Hydrocarbon-degrading bacteria *Alcanivorax* and *Marinobacter* associated with microalgae *Pavlova lutheri* and *Nannochloropsis oculata*. *Front Microbiol*. 2020;11:572931 *Frontiers Media SA*.
31. Deborah Gnana Selvam A, Thatheyus AJ. Microbial Degradation of Petroleum Hydrocarbons: An Overview. In: Kumar V, Kumar M, Prasad R, editors. *Microbial Action on Hydrocarbons*. Springer; 2018;485–503.
32. Cheffi M, Hentati D, Chebbi A, Mhiri N, Sayadi S, Marqués AM, et al. Isolation and characterization of a newly naphthalene-degrading *Halomonas pacifica*, strain Cnaph3: biodegradation and biosurfactant production studies. *3 Biotech*. 2020;10:1–15 Springer.
33. Li C, Lai Q, Li G, Liu Y, Sun F, Shao Z. Multilocus sequence analysis for the assessment of phylogenetic diversity and biogeography in *Hypomonas* bacteria from diverse marine environments. *PLoS One*. 2014;9:e101394 *Public Library of Science San Francisco, USA*.
34. Kappell AD, Wei Y, Newton RJ, Van Nostrand JD, Zhou J, McLellan SL, et al. The polycyclic aromatic hydrocarbon degradation potential of Gulf of Mexico native coastal microbial communities after the Deepwater Horizon oil spill. *Front Microbiol*. 2014;5:205 *Frontiers Media SA*.
35. Guo G, Tian F, Ding K, Wang L, Liu T, Yang F. Effect of a bacterial consortium on the degradation of polycyclic aromatic hydrocarbons and bacterial community composition in Chinese soils. *Int Biodeterior Biodegradation*. 2017;123:56–62.
36. Gutierrez T, Green DH, Nichols PD, Whitman WB, Semple KT, Aitken MD. *Polycyclovorans algicola* gen. nov., sp. nov., an Aromatic-Hydrocarbon-Degrading Marine Bacterium Found Associated with Laboratory Cultures of Marine Phytoplankton. *Appl Environ Microbiol. American Society for Microbiology*; 2013;79:205–14.
37. Tao Y, Fishman A, Bentley WE, Wood TK. Altering toluene 4-monoxygenase by active-site engineering for the synthesis of 3-methoxycatechol, methoxyhydroquinone, and methylhydroquinone. *J Bacteriol Am Soc Microbiol*. 2004;186:4705–13.
38. Parales RE, Parales JV, Pelletier DA, Ditty JL. Diversity of microbial toluene degradation pathways. *Adv Appl Microbiol Elsevier*. 2008;64:1–73.
39. Murphy CL, Sheremet A, Dunfield PF, Spear JR, Stepanauskas R, Woyke T, et al. Genomic Analysis of the Yet-Uncultured *Binatota* Reveals Broad Methylophilic, Alkane-Degradation, and Pigment Production Capacities. *MBio Am Soc Microbiol*. 2021;12:e00985-e1021.
40. Khomenkov VG, Shevelev AB, Zhukov VG, Zagustina NA, Bezborodov AM, Popov VO. Organization of metabolic pathways and molecular-genetic mechanisms of xenobiotic degradation in microorganisms: A review. *Appl Biochem Microbiol Springer*. 2008;44:117–35.
41. Furukawa K, Suenaga H, Goto M. Biphenyl dioxygenases: functional versatilities and directed evolution. *J Bacteriol Am Soc Microbiol*. 2004;186:5189–96.
42. Ikuma K, Gunsch C. Effect of carbon source addition on toluene biodegradation by an *Escherichia coli* DH5 $\alpha$  transconjugant harboring the TOL plasmid. *Biotechnol Bioeng Wiley Online Library*. 2010;107:269–77.
43. Boronin AM, Kosheleva IA. The role of catabolic plasmids in biodegradation of petroleum hydrocarbons. In: Cao G, Orrù R, editors. *Current Environmental Issues and Challenges*. Springer, Dordrecht; Springer; 2014. p. 159–68.
44. Corteselli EM, Aitken MD, Singleton DR. Description of *Immundisolibacter cernigliae* gen nov., sp. nov., a high-molecular-weight polycyclic aromatic hydrocarbon-degrading bacterium within the class Gammaproteobacteria, and proposal of *Immundisolibacterales* ord. nov. and *Immundisolibacteraceae* f. Int J Syst Evol Microbiol. 2017;67:925 *Microbiology Society*.
45. Schreiber L, Fortin N, Tremblay J, Wasserscheid J, Sanschagrin S, Mason J, et al. In situ microcosms deployed at the coast of British Columbia (Canada) to study dilbit weathering and associated microbial communities under marine conditions. *FEMS Microbiol Ecol*. 2021;97:fiab082 (Oxford University Press).
46. Somee MR, Dastgheib SMM, Shavandi M, Maman LG, Kavousi K, Amoozegar MA, et al. Distinct microbial community along the chronic oil pollution continuum of the Persian Gulf converge with oil spill accidents. *Sci Rep Nature Publishing Group*. 2021;11:1–15.
47. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res Oxford University Press*. 2020;48:D265–8.
48. Okie JG, Poret-Peterson AT, Lee ZMP, Richter A, Alcaraz LD, Eguarte LE, et al. Genomic adaptations in information processing underpin trophic strategy in a whole-ecosystem nutrient enrichment experiment. *Elife*. 2020;9:e49816 *eLife Sciences Publications Limited*.
49. Barbato M, Mapelli F, Crotti E, Daffonchio D, Borin S. Cultivable hydrocarbon-degrading bacteria have low phylogenetic diversity but highly versatile functional potential. *Int Biodeterior Biodegradation Elsevier*. 2019;142:43–51.
50. Kleindienst S, Grim S, Sogin M, Bracco A, Crespo-Medina M, Joye SB. Diverse, rare microbial taxa responded to the Deepwater Horizon deep-sea hydrocarbon plume. *ISME J Nature Publishing Group*. 2016;10:400–15.
51. Bacos HP, Liu Z, Erdner DL. Natural sunlight shapes crude oil-degrading bacterial communities in Northern Gulf of Mexico surface waters. *Front Microbiol*. 2015;6:1325 (Frontiers).
52. Gutierrez T. Marine, Aerobic Hydrocarbon-Degrading Gammaproteobacteria: Overview. In: McGenity T, editors. *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*. Handbook of Hydrocarbon and Lipid Microbiology. Springer; 2017. p. 143–152.
53. Sun X, Kostka JE. Hydrocarbon-degrading microbial communities are site specific, and their activity is limited by synergies in temperature and nutrient availability in surface ocean waters. *Appl Environ Microbiol Am Soc Microbiol*. 2019;85:e00443-e519.
54. Margesin R, Labbe D, Schinner F, Greer CW, Whyte LG. Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine alpine soils. *Appl Environ Microbiol Am Soc Microbiol*. 2003;69:3085–92.
55. Brzeszcz J, Steliga T, Kapusta P, Turkiewicz A, Kaszycki P. r-strategist versus K-strategist for the application in bioremediation of hydrocarbon-contaminated soils. *Int Biodeterior Biodegradation Elsevier*. 2016;106:41–52.
56. Guo Q, Yin Q, Du J, Zuo J, Wu G. New insights into the r/K selection theory achieved in methanogenic systems through continuous-flow and sequencing batch operational modes. *Sci Total Environ*. 2022;807:150732 Elsevier.
57. Wang M, Chen X, Liu X, Fang Y, Zheng X, Huang T, Tang YQ, Ackermann M, Nie Y, Wu XL. Even allocation of benefits stabilizes microbial community engaged in metabolic division of labor. *Cell Reports*. 2022;40(13):111410.



58. Tsoi R, Wu F, Zhang C, Bewick S, Karig D, You L. Metabolic division of labor in microbial systems. *Proc Natl Acad Sci*. 2018;115:2526–31 National Acad Sciences.
59. Krzmarzick MJ, Taylor DK, Fu X, McCutchan AL. Diversity and niche of archaea in bioremediation. *Archaea*. Hindawi; 2018; vol. 2018, Article ID 3194108,17 pages.
60. Somee MR, Dastgheib SMM, Shavandi M, Zolfaghar M, Zamani N, Ventosa A, et al. Halophiles in bioremediation of petroleum contaminants: challenges and prospects. In: Saxena G, Kumar V, Shah M, editors. *Bioremediation for Environmental Sustainability*. Elsevier; 2021. p. 251–91.
61. Fathepure BZ. Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. *Front Microbiol*. 2014;5:173 Frontiers.
62. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*: Nature Publishing Group; 2018.
63. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol*. Nature Publishing Group. 2020;38(9):1079–86.
64. Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Bælum J, Kimbrel J, et al. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J*. 2014;8:1464 Nature Publishing Group.
65. King GM, Kostka JE, Hazen TC, Sobocky PA. Microbial responses to the Deepwater Horizon oil spill: from coastal wetlands to the deep sea. *Ann Rev Mar Sci Annual Reviews*. 2015;7:377–401.
66. Liao H, Lin X, Li Y, Qu M, Tian Y. Reclassification of the taxonomic framework of orders cellvibrionales, oceanospirillales, pseudomonadales, and alteromonadales in class gammaproteobacteria through phylogenomic tree analysis. *Msystems Am Soc Microbiol*. 2020;5:e00543-e620.
67. Pérez-Pantoja D, González B, Pieper DH. Aerobic degradation of aromatic hydrocarbons. In: Timmis KN, editors. *Handbook of Hydrocarbon and Lipid Microbiology*. Springer; 2010. p. 799–837.
68. Abbasian F, Lockington R, Mallavarapu M, Naidu R. A comprehensive review of aliphatic hydrocarbon biodegradation by bacteria. *Appl Biochem Biotechnol Springer*. 2015;176:670–99.
69. Abbasian F, Lockington R, Megharaj M, Naidu R. A review on the genetics of aliphatic and aromatic hydrocarbon degradation. *Appl Biochem Biotechnol Springer*. 2016;178:224–50.
70. Meckenstock RU, Boll M, Mouttaki H, Koelschbach JS, Tarouco PC, Weyrauch P, et al. Anaerobic degradation of benzene and polycyclic aromatic hydrocarbons. *J Mol Microbiol Biotechnol Karger Publishers*. 2016;26:92–118.
71. Rabus R, Boll M, Heider J, Meckenstock RU, Buckel W, Einsle O, et al. Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. *J Mol Microbiol Biotechnol Karger Publishers*. 2016;26:5–28.
72. Espinola F, Dionisi HM, Borglin S, Brislawn CJ, Jansson JK, Mac Cormack WP, et al. Metagenomic analysis of subtidal sediments from polar and subpolar coastal environments highlights the relevance of anaerobic hydrocarbon degradation processes. *Microb Ecol Springer*. 2018;75:123–39.
73. Liang C, Huang Y, Wang H. pahE, a functional marker gene for polycyclic aromatic hydrocarbon-degrading bacteria. *Appl Environ Microbiol Am Soc Microbiol*. 2019;85:e02399-e2418.
74. Liang C, Ye Q, Huang Y, Wang Y, Zhang Z, Wang H. Shifts of the new functional marker gene (pahE) of polycyclic aromatic hydrocarbons (PAHs) degrading bacterial population and its relationship with PAHs biodegradation. *Elsevier; J Hazard Mater*. 2022;6:129305.
75. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res Oxford University Press*. 2000;28:27–30.
76. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res Oxford University Press*. 2016;44:D279–85.
77. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. *Nucleic Acids Res Oxford University Press*. 2003;31:371–3.
78. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res Oxford University Press*. 2001;29:37–40.
79. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. The universal protein resource (UniProt). *Nucleic Acids Res Oxford University Press*. 2005;33:D154–9.
80. van Beilen JB, Funhoff EG, van Loon A, Just A, Kaysser L, Bouza M, et al. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. *Appl Environ Microbiol Am Soc Microbiol*. 2006;72:59–65.
81. Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. Evolview v3: a web-server for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res Oxford University Press*. 2019;47:W270–5.
82. He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res Oxford University Press*. 2016;44:W236–41.
83. Lassmann T. Kalpin 3: multiple sequence alignment of large data sets. *Bioinformatics*. 2019;36:1928–9.
84. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5:e9490 Public Library of Science.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

