



Metagenomic Screening for Lipolytic Genes Reveals an Ecology-Clustered Distribution Pattern

Mingji Lu, Dominik Schneider and Rolf Daniel*

Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg August University of Göttingen, Göttingen, Germany

OPEN ACCESS

Edited by:

Rakesh Sharma,
Institute of Genomics and Integrative
Biology, India

Reviewed by:

Ida Helene Steen,
University of Bergen, Norway
Ilya V. Kublanov,
Winogradsky Institute of Microbiology
(RAS), Russia

*Correspondence:

Rolf Daniel
rdaniel@gwdg.de

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 10 January 2022

Accepted: 28 April 2022

Published: 10 June 2022

Citation:

Lu M, Schneider D and Daniel R
(2022) Metagenomic Screening
for Lipolytic Genes Reveals an
Ecology-Clustered Distribution
Pattern. *Front. Microbiol.* 13:851969.
doi: 10.3389/fmicb.2022.851969

Lipolytic enzymes are one of the most important enzyme types for application in various industrial processes. Despite the continuously increasing demand, only a small portion of the so far encountered lipolytic enzymes exhibit adequate stability and activities for biotechnological applications. To explore novel and/or extremophilic lipolytic enzymes, microbial consortia in two composts at thermophilic stage were analyzed using function-driven and sequence-based metagenomic approaches. Analysis of community composition by amplicon-based 16S rRNA genes and transcripts, and direct metagenome sequencing revealed that the communities of the compost samples were dominated by members of the phyla *Actinobacteria*, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Chloroflexi*. Function-driven screening of the metagenomic libraries constructed from the two samples yielded 115 unique lipolytic enzymes. The family assignment of these enzymes was conducted by analyzing the phylogenetic relationship and generation of a protein sequence similarity network according to an integrated classification system. The sequence-based screening was performed by using a newly developed database, containing a set of profile Hidden Markov models, highly sensitive and specific for detection of lipolytic enzymes. By comparing the lipolytic enzymes identified through both approaches, we demonstrated that the activity-directed complements sequence-based detection, and vice versa. The sequence-based comparative analysis of lipolytic genes regarding diversity, function and taxonomic origin derived from 175 metagenomes indicated significant differences between habitats. Analysis of the prevalent and distinct microbial groups providing the lipolytic genes revealed characteristic patterns and groups driven by ecological factors. The here presented data suggests that the diversity and distribution of lipolytic genes in metagenomes of various habitats are largely constrained by ecological factors.

Keywords: lipolytic enzymes, function-driven metagenomics, sequence-based metagenomics, profile HMM, lipolytic enzyme classification, comparative analysis, compost

INTRODUCTION

Enzymes acting on carboxyl ester bonds in lipids, include esterases (EC 3.1.1.1, carboxylesterases) and true lipases (EC 3.1.1.3, triacylglycerol acyl hydrolases) and are all together called lipolytic enzymes (LEs; Kovacic et al., 2019). Due to the catalytic versatility, LEs have remarkable applications in various processes relevant to food, paper, medical, detergent, and pharmaceutical

industries (Hita et al., 2009; Romdhane et al., 2010; Ferrer et al., 2015; Sarmah et al., 2018). Nowadays, LEs are considered to be one of the most important biocatalysts for biotechnological applications.

In principle, LEs can be classified on the basis of the substrate preference (Sarmah et al., 2018) and sequence similarity (Chen et al., 2016). The latter provides an easy-to-perform way for classification and indication of the similarity and evolutionary relationship between LEs. Arpigny and Jaeger (1999) have elaborated the most widely accepted classification of lipolytic enzymes into eight families (I to VIII). The classification system was based on conserved sequence motifs and biological properties of 53 LEs. A recent update of this system resulted in addition of 11 families (IX to XIX) (Kovacic et al., 2019). Besides the nineteen families, there are claims of novel families, such as Est22 (Li et al., 2017), Est9X (Jeon et al., 2009), LipSM54 (Li et al., 2016), and EstDZ2 (Zarafeta et al., 2016). To avoid an artificial inflation of the number of families, previous claims of novelty have to be confirmed and then the novel families have to be integrated into a classification system (Hitch and Clavel, 2019).

Lipolytic enzymes are ubiquitous among all aspects of life. Most known LEs originate from microorganisms (Kovacic et al., 2019). Environmental microbes, including the so far uncultured species, encode a largely untapped reservoir of novel LEs. Metagenomic function-driven and sequence-based approaches provided access to the genetic resources from so far uncultured and uncharacterized microorganisms (Simon and Daniel, 2009, 2011). LEs are among the most frequent targets in function-based screens of metagenomic libraries derived from diverse habitats, such as compost (Kang et al., 2011; Lu et al., 2019), landfill leachate (Rashamuse et al., 2009), marine sediment (Peng et al., 2011; Zhang et al., 2017), activated sludges (Liaw et al., 2010), and hot springs (López-López et al., 2015).

Most published metagenomic screenings for LEs were enzyme activity-driven and not sequence-based (Ferrer et al., 2015; Berini et al., 2017). Only a few studies explored LEs by sequence-driven approaches, including analysis based on regular expression patterns (Masuch et al., 2015), ancestral sequence reconstruction (Verma et al., 2019), and conserved motifs (Zhang et al., 2009; Barriuso and Jesús Martínez, 2015; Zarafeta et al., 2016). For various reasons, only a very limited number of LEs were identified by these strategies. Sequence-based approaches primarily rely on the reference database to infer functions of newly discovered biomolecules (Hugenholtz and Tyson, 2008; Berini et al., 2017; Quince et al., 2017; Ngara and Zhang, 2018). Taking also the constantly increasing amount of genomic and metagenomic data in the public repositories (Keegan et al., 2016; Chen et al., 2017; Sayers et al., 2019) into account the full diversity of LEs is far from being completely described.

In order to quantitatively analyze LEs distributed in environmental samples, we developed a LE-specific profile Hidden Markov Model (HMM) database. Profile HMMs have been widely adopted for detection of remote homologs (Gibson et al., 2015; Berglund et al., 2017; Walsh et al., 2017) and annotation of general functions in microbial genomes and metagenomes (Skewes-Cox et al., 2014; Reyes et al., 2017;

Bzhalava et al., 2018). However, they have not yet been specifically applied to LEs. Once developed and validated, the database was applied to profile the lipolytic genes in metagenomes from various habitats. Profiling the distribution of LEs among various habitats provides a straightforward approach for their downstream analysis. In this study, two composts were sampled and LEs identified through function-based and sequence-based approaches were compared. The distribution of lipolytic genes in 175 metagenomes was also investigated by sequence-based screening.

MATERIALS AND METHODS

Sample Collection

Compost samples were collected as described previously (Lu et al., 2019). Briefly, two compost piles fermenting mainly wood chips (Pile_1) or kitchen waste (Pile_2) were sampled. Temperatures at the sampling spots were measured, and the two samples were designated as compost55 (55°C for Pile_1) and compost76 (76°C for Pile_2). Approximately 50 g compost per sample was collected in sterile plastic tubes and stored frozen until further use.

Isolation of Nucleic Acids

Total DNA of the compost sample was isolated by using the phenol-chloroform method (Zhou et al., 1996) and MoBio PowerSoil DNA extraction kit as recommended by the manufacturer (MO BIO Laboratories, Hilden, Germany). DNA obtained from these two methods was pooled per sample and stored at –20°C until use.

RNA was extracted by employing the MoBio PowerSoil RNA isolation kit as recommended by the manufacturer (MO BIO Laboratories, Hilden, Germany). Residual DNA was removed by treatment with 2 U Turbo DNase (Applied Biosystems, Darmstadt, Germany) at 37°C for 1 h and recovered by using RNeasy MinElute Cleanup kit as recommended by the manufacturer (Qiagen, Hilden, Germany). RNA yields were estimated by employing a Qubit® Fluorometer as recommended by the manufacturer (Thermo Fisher Scientific, Schwerte, Germany). A PCR reaction targeting the 16S rRNA gene was performed to verify the complete removal of DNA as described by Schneider et al. (2015). Subsequently, the DNA-free RNA was converted to cDNA using the SuperScript™ III reverse transcriptase (Thermo Fisher Scientific, Schwerte, Germany). Briefly, a mixture (14 µl) containing 100 ng of DNA-free RNA in DEPC-treated water, 2 µM of reverse primer (5' – CCGTCAATTCMTTGGAGT – 3') and 10 mM dNTP mix was incubated at 65°C for 5 min and chilled on ice for at least 1 min. Then, 10 µl of cDNA synthesis mix including reaction buffer, 5 mM MgCl₂, 0.01 M DTT, 1 µl 40U RiboLock™ RNase inhibitor (Thermo Fisher Scientific, Schwerte, Germany) and 200U SuperScript™ III reverse transcriptase (Thermo Fisher Scientific, Schwerte, Germany) was added to each RNA/primer mixture of the previous step, and incubated at 55°C for 90 min. The reaction was ended by incubation at 70°C for 15 min.

Amplicon-Based Analysis of Partial 16S rRNA Genes and Transcripts

The PCR amplification of the V3–V5 regions of bacterial 16S rRNA genes and transcripts were performed with the following set of primers comprising the Roche 454 pyrosequencing adaptors (underlined), a key (TCAG), a unique 10-bp multiplex identifier (MID), and template-specific sequence per sample: the forward primer V3for_B (5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-TACGGRAGGCAGCAG-3'), (Liu et al., 2007), and reverse primer V5rev_B 5'-CTATGCGCCTTGCCAGCCCCGCTCAG-MID-CCGTC AATTCMTTTGAGT-3' (Wang and Qian, 2009). The PCR reaction mixture (50 μ l) contained 10 μ l of fivefold reaction buffer, 200 μ M of each of the four deoxynucleoside triphosphates, 0.2 μ M of each primer, 5% DMSO, 1 U of Phusion hot start high-fidelity DNA Polymerase (Finnzymes, Vantaa, Finland) and 50 ng template (DNA or cDNA). The thermal cycling scheme comprised initial denaturation at 98°C for 5 min, 25 cycles of denaturation at 98°C for 45 s, annealing for 45 s at 60°C, and extension at 72°C for 30 s, followed by a final extension period at 72°C for 5 min. All amplicon PCR reactions were performed in triplicate and pooled in equimolar amounts for sequencing. The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA gene and transcript amplicons by using a 454 GS-FLX sequencer and titanium chemistry as recommended by the manufacturer (Roche, Mannheim, Germany).

Quality-filtering and denoising of the recovered 16S rRNA pyrotag reads were performed with the QIIME (1.9.1) software package (Caporaso et al., 2010) by employing the scheme outlined by Schneider et al. (2015). Forward and reverse primer sequences were removed with the *split_libraries.py* script. Pyrosequencing noise was removed with Acacia v01.53 (Bragg et al., 2012) and chimeric sequences were removed with UCHIME (Edgar et al., 2011). Operational taxonomic unit (OTU) determination was performed by employing the *pick_open_reference_otus.py* script at genetic divergence level of 3%. Taxonomic classification of OTUs was performed by *parallel_assign_taxonomy_blast.py* script against the Silva SSU database release 138 (Quast et al., 2013). The *filter_otu_table.py* script was used to remove singletons, extrinsic domain OTUs (i.e., chloroplast, archaeal, and eukaryotic sequences), and unclassified OTUs.

Rarefaction curves was calculated with QIIME software by using *alpha-rarefaction.py*.

Metagenomic Sequencing and Data Processing

The sequencing libraries were constructed and indexed with Nextera DNA Sample Preparation kit and Index kit as recommended by the manufacturer (Illumina, San Diego, CA, United States). Paired-end sequencing was performed using a HiSeq 4000 instrument (2 \times 150 bp) as recommended by the manufacturer (Illumina, San Diego, CA, United States). Raw reads were trimmed with Trimmomatic version 0.36 (Bolger et al., 2014) and verified with FastQC version 0.11.5 (Andrews, 2010). Then, reads were submitted to MG-RAST metagenomics

analysis server and processed by the default quality control pipeline (Keegan et al., 2016). Microbial composition analysis was performed September 2020 using MG-RAST best hit classification tool against the databases of M5RNA (Non-redundant multisource ribosomal RNA annotation) and M5NR (M5 non-redundant protein) with default settings. Functional classification was performed based on clusters of orthologous groups (COGs) and Subsystem categories with default settings. Since we mainly focused on the bacterial community, the baseline for all fractions reported referred to the reads assigned to the bacterial domain.

Construction of Metagenomic Plasmid Libraries and Function-Based Screening for Lipolytic Activity

Lipolytic genes were screened by constructing small-insert plasmid libraries as described by Lu et al. (2019). Briefly, DNA was sheared by sonication for 3 s at 30% amplitude and cycle 0.5 (UP200S Sonicator, Stuttgart, Germany), and size-separated using a 0.8% low-melting point agarose gel. DNA fragments from 6 to 12 kb were recovered by gel extraction using the peqGOLD Gel Extraction kit as recommended by the manufacturer (Peqlab Biotechnologie GmbH, Erlangen, Germany). The metagenomic small-insert library was constructed using the vectors pFLD or pCR-XL-TOPO (Thermo Fisher Scientific, Schwerte, Germany).

Vector pFLD was digested with *PmlI* at 37°C for 2 h and dephosphorylated with 5 U Antarctic phosphatase at 37°C for 30 min as recommended by the manufacturer (NEB, Ipswich, MA, United States). Subsequently, the ends of DNA fragments were blunt-ended and phosphorylated by employing the Fast DNA End Repair kit (Thermo Fisher Scientific, Schwerte, Germany). SureClean was applied to purify DNA or vector between steps as described by the manufacturer (Bioline GmbH, Luckenwalde, Germany). Finally, metagenomic fragments and pFLD vector were ligated using T4 DNA ligase (Thermo Fisher Scientific, Schwerte, Germany) at 16°C, overnight. Metagenomic DNA fragments were cloned into vector pCR-XL-TOPO following the protocol of the manufacturer (Thermo Fisher Scientific, Schwerte, Germany).

To screen for lipolytic activity, *Escherichia coli* TOP10 was used as the host (Dukunde et al., 2017). Library-bearing cells were plated onto LB agar plates (15 g/L) containing 1% (v/v) emulsified tributyrin (Sigma) as the indicator substrate and the appropriate antibiotic (pFLD, 100 μ g/ml Ampicillin; pCR-XL-TOPO, 50 μ g/ml Kanamycin). The quality of the libraries was controlled by checking the average insert sizes and the percentage of insert-bearing *E. coli* clones (Table 1). Cells were incubated on indicator agar at 37°C for 24 h and subsequently for 1–7 days at 30°C. Lipolytic-positive *E. coli* clones were identified by the formation of clear zones (halos) around individual colonies.

The recombinant plasmid DNA derived from positive clones was isolated by using the QIAGEN plasmid mini kit (QIAGEN) and digested with *PmlI* (vector pFLD) or *EcoRI* (vector pCR-XL-TOPO) at 37°C for 2 h. The digestion pattern was analyzed, and phenotype of positive clones was confirmed by transformation of

TABLE 1 | Summary of metagenomic libraries used for lipolytic activity screening in this and other studies.

Environmental sample ^b	Vector type (average insert size in kb)	No. of library-containing clones (confirmed positive hits, No. of hits per million of clones)	Probability (No. of hits per Gb of DNA screened)	References
Compost	Plasmid (5.3)	675,200 (156, 213)	43.6	compost55 (this study)
Compost	Plasmid (5.6)	234,912 (43, 183)	32.7	compost55 (this study)
Compost	Plasmid (6)	281,281 (37, 132)	21.9	compost76 (this study)
Compost	Plasmid (6.2)	140,747 (14, 100)	16.1	compost76 (this study)
Compost	Plasmid (3.2)	21,000 (14, 670)	208	Lämmle et al., 2007
Compost	Fosmid (35)	23,400 (19, 810)	23.2	Kim et al., 2010
Compost	Fosmid (37.5)	1,920 (2, 1040)	27.8	Leis et al., 2015
Compost	Fosmid (- ^a)	13,000 (10, 770)	- ^a	Kang et al., 2011
Compost	plasmid (- ^a)	66,000 (6, 0.90)	- ^a	Popovic et al., 2017
Grassland soil	Plasmid (5.7)	510,808 (2, 0.4)	0.714	Nacke et al., 2011
Grassland soil	Fosmid (27.8)	50,952 (2, 40)	1.41	Nacke et al., 2011
Forest soil	Fosmid (35)	33,700 (8, 240)	6.78	Lee et al., 2004
Forest soil	Plasmid (3.1)	70,000 (3, 42)	13.8	Berlemont et al., 2013
River surface water	BAC (50)	8,000 (1, 120)	2.5	Wu and Sun, 2009
Hot spring biofilm	BAC (50)	68,352 (10, 150)	2.93	Yan et al., 2017
Surface sea water	BAC (70)	20,000 (4, 200)	2.86	Chu et al., 2008
Marine sediment	plasmid (4.5)	29,000 (6, 200)	46.0	Ranjan et al., 2018
Marine sediment	Fosmid (36)	40,000 (19, 480)	13.2	Hu Y. et al., 2010
Marine mud	Fosmid (40)	40,000 (5, 120)	3.12	Gao et al., 2016
Deep-sea hydrothermal vent	Fosmid (35)	18,000 (7, 390)	11.1	Fu et al., 2015
Paper mill sludge	Plasmid (5.1)	15,000 (13, 870)	170	Jia et al., 2019
Activated sludge	Plasmid (5.1)	3,818 (12, 3140)	616	Liaw et al., 2010
Activated sludge	Plasmid (2.5)	40,000 (1, 24)	10.0	Shao et al., 2013
Solar saltern	Fosmid (35)	5,100 (1, 200)	5.60	Jayanath et al., 2018
Oil field soil	Plasmid (3.9)	83,000 (1, 12)	3.09	Fan et al., 2011

^aThis information is not specified in the reference.

^bExcept compost metagenomic libraries, only those included the full library information were listed.

the recovered plasmids from the previous step into the host and rescreening on indicator agar plates. In addition, lipolytic activity toward different triacylglycerides was measured qualitatively by incubating the confirmed lipolytic positive clones on agar plates emulsified with tributyrin (C4), tricaproin (C6), tricapyrin (C8), tricaprin (C10), trilaurin (C12), trimyristin (C14), or tripalmitin (C16). Formation of clearing zones (halos) on agar plates indicated lipolytic activity.

Analysis of Lipolytic Genes From Function-Based Screenings

The plasmids recovered from the confirmed positive clones were pooled in equal amounts (50 ng of each clone) for compost55 and compost76. Then, the two plasmid DNA mixtures were sequenced using an Illumina MiSeq instrument with reagent kit version 3 (2 × 300 cycles) as recommended by the manufacturer (Illumina, San Diego, CA, United States). To remove the vector sequences, raw reads were initially mapped against vector sequences (pFLD or pCR-XL-TOPO) using Bowtie 2 (Langmead and Salzberg, 2012). The unmapped reads were quality-filtered by Trimmomatic v0.30 (Bolger et al., 2014) and assembled into contigs by MetaVelvet v1.2.01 (Namiki et al., 2012) and MIRA 4 (Chevreux et al., 1999). In addition, both

ends of the inserts of each plasmid were sequenced using Sanger technology and the following primers: pFLD504_F (5'-GCCTTACCTGATCGCAATCAGGATTTTC-3') and pFLD706_R (5'-CGAGGAGAGGGTTAGGGATAGGCTTAC-3') for vector pFLD, and M13_Forward (5'-GTAAAACGACGGCCAG-3') and M13_Reverse (5'-CAGGAAACAGCTATGAC-3') for vector pCR-XL-TOPO. The raw Sanger reads were processed with the Staden package (Staden et al., 2003). Finally, the full insert sequence for each plasmid was reconstructed by mapping the processed Sanger reads on the contigs assembled from the Illumina reads. Open reading frames (ORFs) were predicted by MetaGeneMark (Zhu et al., 2010) using default parameters. Lipolytic genes were annotated by searches against NCBI Non-redundant sequence database¹.

Family Classification of Lipolytic Enzymes Revealed From Function-Based Screening

Lipolytic enzymes were clustered according to the classification standard defined by Arpigny and Jaeger (1999). In order to classify LEs identified from function-based screening, we have integrated all the so far reported lipolytic families, including

¹<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>

families I to XIX, and potential novel families reported in recent studies (**Supplementary Table 1**). The neighbor-joining tree and maximum likelihood tree were constructed with LEs identified from this study and reference proteins (**Supplementary Table 2**) using MEGA version 7 (Tamura et al., 2013). The robustness of the tree was tested by bootstrap analysis using 500 replications. The phylogenetic tree was depicted by GraPhlAn (Asnicar et al., 2015). To confirm the classification and group proteins in clusters, a protein sequence similarity network was generated. In a protein sequence similarity network, members in a potential isofunctional group consist of nodes (symbol) that share a sequence similarity larger than a selected value and are connected by edges (line). As similarity increases, edges decrease and finally proteins can be separated into defined clusters (Gerlt et al., 2015). In this study, a protein sequence similarity network was generated by submitting the same sequence dataset used in the phylogenetic analysis to the Enzyme Function Initiative-Enzyme Similarity Tool web server (EFI-EST)² (Atkinson et al., 2009) with an *E*-value cutoff of $\leq 1e^{-10}$ and alignment score ≥ 16 . The resulting network was visualized in Cytoscape 3.2.1 using the organic layout (Shannon et al., 2003). In addition, multiple-sequence alignments were conducted to explore the presence of catalytic residues, and conservative and distinct motifs in each lipolytic family by employing ClustalW (Larkin et al., 2007).

Building Profile Hidden Markov Model Database for Sequence-Based Screening

A search method based on profile HMMs was developed to identify and annotate putative lipolytic genes in metagenomes (**Supplementary Figure 1**). In order to target homologous sequences, profile HMMs were built from multiple sequence alignments, which requires relatedness in the input protein sequences. Thus, consistent to the classification of function-derived LEs, we generally followed the clustering system of Arpigny and Jaeger (1999).

With the exception of LEs belonging to families II and VIII, and patatin-like-proteins, LEs in the other families generally share a conserved α/β -hydrolase fold and a canonical G-x-S-x-G pentapeptide around the catalytic serine (Kovacic et al., 2019). ESTHER is a database dedicated to proteins with α/β -hydrolase-fold and their classifications (Lenfant et al., 2013), containing approximately 60,000 α/β hydrolases grouped in 214 clusters (November 2019). In ESTHER, families I-XIX were integrated into an own classification with corresponding entries³. We thereby designated lipolytic families that were classified and named according to ESTHER database as ELF (abbreviation of ESTHER Lipolytic Families). For lipolytic families that were not incorporated into the 19 families (I-XIX), their corresponding ELFs were determined by searching LEs against ESTHER database. Generally, a LE was assigned to an ELF if its BLASTp top hit (with lowest *e*-value) had $\geq 60\%$ amino acid identity and $\geq 80\%$ query coverage. Protein sequences in all of the

determined ELFs were downloaded from ESTHER database for profile HMM construction.

Firstly, multiple sequence alignments were performed with protein sequences in each ELF, using the following three algorithms and default settings: ClustalW (Thompson et al., 1994), Clustal Omega (Sievers et al., 2011), and Muscle (Edgar, 2004). Subsequently, the three alignment sets were run through *hmmbuild* in HMMER3⁴ to create three sets of profile HMMs. Moreover, profile HMMs supplied in the ESTHER database were downloaded. Finally, four profile HMM databases were constructed by concatenating and compressing the respective set of profile HMMs using *hmmcompress*. Thereafter, we designated the four profile HMM databases with respect to the corresponding alignment algorithm (ClustalW-pHMMs, omega-pHMMs, and muscle-pHMMs) or source (ESTHER-pHMMs). All generated databases are available under <https://github.com/mingji-lu/database-for-lipolytic-enzymes>.

For families II, VIII and patatin-like-proteins, profile HMMs were retrieved directly from Pfam database (Finn et al., 2014) using the searching keywords of “GDSL,” “beta-lactamase,” and “patatin,” respectively. The profile HMM database was constructed as described above and designated as Pfam-pHMMs, specifying for LEs in families II and VIII, and patatin-like-proteins.

Validating Profile Hidden Markov Model Database

The prediction sensitivity and specificity of the profile HMM databases were evaluated using four datasets. Dataset 1, LEs recruited in the UniProtKB database using as search strategy the EC numbers 3.1.1.1 or 3.1.1.3, and protein length between 200 and 800 amino acids. Only the prokaryotic LEs were selected for analysis (**Supplementary Table 3A**). Dataset 2 comprises LEs reported in literature. Most of these enzymes were obtained through metagenomic approaches and biochemically characterized, and with a confirmed lipolytic family assignment by constructing a multiple sequence alignment and/or phylogenetic tree (**Supplementary Table 3B**). Dataset 3 includes protein sequences predicted by MetaGeneMark (Zhu et al., 2010) from identified inserts harboring functional lipolytic genes (**Supplementary Table 3C**). Dataset 4 comprises randomly selected protein sequences (not recruited from ESTHER database) that were annotated in UniProt or NCBI database as non-lipolytic proteins but with sequence similarity to LEs (**Supplementary Table 3D**). Proteins in the four datasets were screened against the profile HMM databases successively with *hmmsearch* using an *E*-value cutoff of $\leq 1e^{-10}$. The sensitivity and specificity of each database were evaluated by the recalls and false positive returns. In addition, we compared our method (profile HMMs) with the similarity-based pairwise sequence alignment method (BLAST; Altschul et al., 1990). The database for BLAST-based searching was built with the same dataset used for profile HMM construction. BLASTp was performed at an *E*-value cutoff of $\leq 1e^{-10}$.

²<http://efi.igb.illinois.edu/efi-est/index.php>

³http://bioweb.supagro.inra.fr/ESTHER/Arpigny_Jaeger.table

⁴<http://hmmmer.org>

In order to improve the accuracy for assigning proteins to lipolytic families and distinguishing “true” LEs from the non-lipolytic proteins, protein sequences were annotated by two methods and combined for final assignment. Briefly, putative lipolytic proteins (PLPs) identified by screening against the selected profile HMM database (one from ClustalW-pHMMs, omega-pHMMs, muscle-pHMMs, and ESTHER-pHMMs) were further searched against the ESTHER database (all entries were included) by BLASTp using an E -value cutoff of $\leq 1e^{-10}$ (Supplementary Figure 1). A PLP was assigned to a lipolytic family only if it was annotated into the same ELF by *hmmscan* and BLASTp. Otherwise, according to the BLAST results, the remaining PLPs were either annotated as “unassigned” PLPs or non-lipolytic proteins (Supplementary Figure 1). In principle, PLPs with the best Blast hits were affiliated to the miscellaneous ESTHER families (functions were not determined, including *5_AlphaBeta_hydrolase*, *6_AlphaBeta_hydrolase*, *Abhydrolase_7*, and *AlphaBeta_hydrolase*), or other ESTHER families (with <60% identity or <70% query coverage) were classified as unassigned PLPs. The remaining PLPs with the best Blast hits showing $\geq 60\%$ amino acid identity and $\geq 70\%$ query coverage to the non-lipolytic ESTHER families were classified as non-lipolytic proteins.

Family annotation of PLPs obtained by screening against Pfam-pHMMs were confirmed by a further scan against the CATH HMMs database (Knudsen and Wiuf, 2010) using the GitHub repository *cath-tools-genomescan*⁵. PLPs were assigned to lipolytic families VIII and II, or patatin-like-proteins only if the PLP was assigned to the specific FunFams (functional families) dedicated to lipolytic-related activities, which were inferred from the functionally characterized LEs and gene ontology (GO) annotations (Supplementary Table 4). Additionally, based on our literature search, the LEs in family VIII were generally restricted to PLPs with sequence length between 350 and 450 amino acids. In other cases, the PLP was grouped into non-lipolytic proteins. For the unassigned PLPs, these sequences show low similarity to any ESTHER family with known function or CATH FunFams, and hence, could contain novel lipolytic or non-lipolytic proteins. Non-lipolytic proteins were excluded from the downstream analysis.

Sequence-Based Screening for Putative Lipolytic Genes

Sequence-based screening for putative lipolytic genes in the two compost metagenomes were performed as described above. Briefly, the processed metagenomic short reads were assembled into contigs with MetaSPAdes version 3.10.1 (Bankevich et al., 2012). Then, protein sequences were deduced from PROKKA v1.14.5 annotation (Seemann, 2014). In order to obtain full-length lipolytic genes, only proteins with an amino acid sequence length between 200 and 800 amino acids were retained. Subsequently, the resulting protein sequences were screened against the selected profile HMM databases using *hmmscan* (Eddy, 2011) with an E -value cutoff of $\leq 1e^{-10}$. Identified PLPs were further assigned into different lipolytic families as described

above (Supplementary Figure 1). Moreover, the lipolytic family classification of assigned PLPs was confirmed by constructing the protein sequence similarity network (Atkinson et al., 2009). The taxonomic origins of PLP-encoding genes and their corresponding contigs were determined using Kaiju web server (Menzel et al., 2016)⁶. Taxonomic distributions of assigned PLPs in each lipolytic family were visualized via Circos software (Krzywinski et al., 2009).

Comparative Analysis of Metagenomic Datasets

A total of 175 assembled metagenomes from 15 different habitats were retrieved from the Integrated Microbial Genomes and Microbiomes database (IMG/M). These included metagenomes from anaerobic digester active sludges (ADAS, $n = 9$), agriculture soils (AS, $n = 10$), composts (COM, $n = 18$), grassland soils (GS, $n = 11$), human gut systems (HG, $n = 16$), hypersaline mats (HM, $n = 7$), hydrocarbon resource environments (HRE, $n = 6$), hot springs (HS, $n = 14$), landfill leachates (LL, $n = 10$), marine sediments (MS, $n = 12$), marine waters (MW, $n = 10$), oil reservoirs (OR, $n = 13$), river waters (RW, $n = 11$), tropical forest soils (TFS, $n = 14$), and wastewater bioreactors (WB, $n = 13$) (Supplementary Table 5). Data processing including open reading frame prediction in assembled contigs and taxonomic assignment of the corresponding deduced protein sequences were conducted by the IMG/M built-in pipelines (Chen et al., 2017). The protein sequences were downloaded from IMG/M database and used in the sequence-based screening as described above (Supplementary Figure 1).

For comparative analysis, the abundance of PLP-encoding genes in each metagenome were normalized according to the method described by Kaminski et al. (2015). The normalized count is in units of LPGM (Lipolytic hits Per Gigabase per Million mapped genes). Unless otherwise stated, LPGM values were used for all calculations. Heatmap was built in R v3.5.2 (R Core Team, 2016) with the function *heatmap.2* using the “Heatplus” package (Ploner, 2015). The heatmap hierarchical clustering was performed with “vegan” package (*vegdist* = “bray,” *data.dist* = “ward.D”). Non-metric multidimensional scaling (NMDS) was also performed with the “vegan” package (Oksanen et al., 2018). The analysis of similarities (ANOSIM) was performed with 9,999 permutations using PAST 4 (Hammer et al., 2001). The taxonomic affiliation of PLPs was retrieved from IMG/M. Association networks between habitats and phylogenetic distribution of PLPs at genus level were generated by mapping significant point biserial correlation values with the “indicspecies” package in R (De Cáceres, 2013). Only genera with significant correlation coefficients ($P \leq 0.05$) were included. The resulting bipartite networks were visualized with Cytoscape v3.5 by using the *edge-weighted spring embedded layout* algorithm, whereby the habitats were source nodes, genera target nodes and edges (lines connecting nodes) weighted positive associations between genera and specific habitat or habitats combinations.

In addition, due to the ambiguity of unassigned PLPs, all analyses were performed successively using two datasets: (1) only

⁵<https://github.com/UCLOrengoGroup/cath-tools-genomescan>

⁶<http://kaiju.binf.ku.dk/server>

assigned PLPs, as the consideration of excluding the potential non-lipolytic ones, (2) assigned and unassigned PLPs combined (total PLPs), in order to include all the possible lipolytic ones. This paper mainly focuses on the assigned PLPs for the sake of accuracy, but the comparative analysis of total PLPs was also performed.

RESULTS AND DISCUSSION

Phylogenetic and Functional Profile of Microbes in the Compost Metagenomes

During the heating-up process of composting, the succession of microorganisms plays a key role in degrading organic matter (Dougherty et al., 2012). In this study, the bacterial community compositions in two compost samples with different pile core temperatures of 55 (compost55) and 76°C (compost76) were revealed by amplicon-based sequencing of 16S rRNA genes (DNA-based, total community) and transcripts (RNA-based, potentially active community) (Supplementary Figures 2, 3, respectively). To extend the taxonomic analysis, the environmental DNA from both metagenomes were also directly sequenced (Supplementary Table 6). Generally, the bacterial community determined by direct sequencing were consistent with that derived from 16S rRNA gene-based analysis. The bacterial phyla *Actinobacteria*, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Chloroflexi* were predominant (relative abundance >5% each) in compost55 and compost76 (Supplementary Figures 3, 4, respectively). This is in agreement with previous studies of bacterial communities in thermophilic composts (Ryckeboer et al., 2003; Antunes et al., 2016; Yu et al., 2018; Zhou et al., 2018). Differences were detected, which were derived mainly from the different feedstock composition (wood chips vs. kitchen waste) and composting conditions (core temperature 55 vs. 76°C). *Actinobacteria* was the most abundant phylum (>25%) in compost55 (Supplementary Figures 3, 4), which is accordance with the bacterial communities in composts using mainly plant material as feedstock (Yu et al., 2007; Zhang et al., 2014). In compost76, members of the *Firmicutes* were most abundant (>55%), which was also reported for composts harboring high-nitrogen feedstock, such as animal manure and kitchen waste (Niu et al., 2013; Antunes et al., 2016; Ma et al., 2018; Zhou et al., 2019). The 16S rRNA gene and transcript analysis (Supplementary Table 7) revealed presence of genera (>1%) in compost55 such as *Brockia*, *Rhodothermus*, *Thermobispora*, *Longispora*, *Geobacillus*, *Filomicrobium*, and *Thermomonospora*, and in compost76 such as *Symbiobacterium*, *Calditerricola*, and *Thermaerobacter*. The detected genera were among the typical bacterial taxa previously identified in composting processes (Ryckeboer et al., 2003; Antunes et al., 2016; Yu et al., 2018; Zhou et al., 2018).

Additionally, the metagenomic reads were searched against the COG and SEED subsystem databases to assess the functions prominent in compost microbes (Supplementary Figure 5). In principle, compost55 and compost76 share similar metabolic patterns (Supplementary Figure 5). Particularly, the broad diversity and abundance of gene functions in carbohydrate

metabolism and transport (COG) and carbohydrates (SEED subsystem) indicated that composts were potential candidates for exploring biocatalysts (Hu X. et al., 2010; Leis et al., 2015; Wang et al., 2016; Egelkamp et al., 2019). Notably, the COG category of lipid transport and metabolism as well as the subsystems category of fatty acids, lipids, and isoprenoids were more abundant in the compost55 community than in the compost76 community, suggesting a higher possibility to identify lipolytic genes in the compost55 metagenome.

Function-Based Screening of Lipolytic Enzymes in Compost Metagenomes

In this study, four metagenomic libraries were prepared to probe the diversity of LEs from compost microbes by the function-driven approach using tributyrin-containing agar (Table 1). Overall, approximately 4.89 and 2.56 Gbp of cloned compost DNA were screened, yielding 199 and 51 positive clones for compost55 and compost76, respectively. Previous studies have used various vectors such as BACs, fosmids and plasmids for function-based screening of LEs from different bioresources (Lee et al., 2004; Lämmle et al., 2007; Kim et al., 2010; Nacke et al., 2011; Berlemont et al., 2013; Shao et al., 2013; Leis et al., 2015; Jia et al., 2019). The hit rate to recover a lipolytic-positive clone ranged from 0.714 to 208 per Gb of cloned DNA (Table 1). Among the compost metagenomic libraries, the targeting probability toward a LE in our study ranged from 16.1 to 43.6 per Gb and is generally consistent with the values from other studies (Lämmle et al., 2007; Kim et al., 2010; Leis et al., 2015). In addition, the probabilities in metagenomic libraries from compost and sludge are generally higher than those from other environments, such as grassland, forest soil, and river water (Wu and Sun, 2009; Nacke et al., 2011; Berlemont et al., 2013). According to Liaw et al. (2010), the probability and/or hit rate for discovering a lipolytic clone is largely attributed to the sample source. Other studies further suggested that samples subjected to specific enrichment processes, such as composting and waste treatment procedures, usually results in a high hit rate (Mayumi et al., 2008; Kang et al., 2011; Popovic et al., 2017).

The insert sizes of the recovered plasmids (250 in total) with a confirmed phenotype ranged from 1,038 to 12,587 bp. In all inserts, at least one putative gene showing similarities to known genes encoding lipolytic enzymes was detected. In total, 210 and 60 lipolytic genes were identified from compost55 and compost76 derived libraries, respectively. To identify unique and full-length LEs, the amino acid sequences deduced from the corresponding lipolytic genes were clustered at 100% identity. This resulted in 115 (92 for compost55, 23 for compost76, with 7 shared by both samples) unique and full-length LEs (Supplementary Table 8). The length of the unique LEs ranged from 223 to 707 amino acids, with calculated molecular masses from 23.9 to 72.3 kDa (Supplementary Table 9). Forty-one of the deduced enzymes showed the highest similarity to esterases/lipases from uncultured bacteria (40) or archaea (1, EstC55-13). The remaining 74 were most similar to LEs from cultured bacteria. Seven of the 41 LEs from uncultured prokaryotes showed the highest identities (53–65%) to lipolytic enzymes obtained during function-based

screening of metagenomes derived from marine sediment (Hu Y. et al., 2010), forest topsoil (Lee et al., 2004), mountain soil (Ko et al., 2012), activated sludge (Liaw et al., 2010), wheat field (Stroobants et al., 2015), and compost (Okano et al., 2015). In the remaining 34 cases, the matching esterases/lipases were mainly detected by sequence-based metagenomic surveys of composts (15 LEs), soil (7 LEs), marine sediment (6 LEs), and marine water (3 LEs).

Functionally Derived Lipolytic Enzymes Are Affiliated to Various Lipolytic Enzyme Families

The LEs identified through function-based screening were grouped into families based on the classification system reported by Arpigny and Jaeger (1999). With the increasing amount of reports on LEs, claims of new families have been reported (Arpigny and Jaeger, 1999; Jeon et al., 2011a; Wang et al., 2013; Esteban-Torres et al., 2014; Fang et al., 2014; Rahman et al., 2016; Castilla et al., 2017). In this study, we integrated 29 so-called “novel” families into the classification system for phylogenetic analysis. As shown in the phylogenetic tree (Figure 1 and Supplementary Figure 6), LEs were assigned to 12 families, including families I, II, III, IV, V, VII, VIII, XVII, EM3L4 (Jeon et al., 2011b), FLS18 (Hu Y. et al., 2010), EstGS (Nacke et al., 2011), LipT (Chow et al., 2012), patatin-like-proteins and tannases (Supplementary Table 8). The majority of the LEs were affiliated to families V (25 LEs), VIII (21 LEs), IV (15 LEs), I (8 LEs), and patatin-like-proteins (9 LEs). Noteworthy, 7 LEs could not be classified into any known lipolytic family, indicating new branches of LEs. In agreement with previous studies (Arpigny and Jaeger, 1999; Glogauer et al., 2011; Akmoussi-Toumi et al., 2018), the “true lipases,” which can hydrolyze long-chain substrates ($\geq C_{10}$) were all affiliated to family I (Figure 1). The remaining LEs exhibiting a preference for short-chain substrates ($< C_{10}$) were esterases.

To verify the classification result, a protein sequence similarity network was built (Figure 1). The network visualizes relationships among evolutionarily related proteins and is usually considered as an approach complementary to the phylogenetic analysis (Atkinson et al., 2009; Gerlt et al., 2015). At a threshold of 1×10^{-16} , the network produced clusters that almost matched all the lipolytic families, with the same classification results as obtained by phylogenetic analysis (Figure 1).

Multiple sequence alignments revealed the catalytic residues and conserved motifs in each family (Supplementary Figure 7). For LEs that harbor the canonical α/β -hydrolase fold, the catalytic triad is consistently composed of a nucleophilic serine, an aspartic acid/glutamic acid and a histidine residue (Nardini and Dijkstra, 1999). Most of these LEs contain the conserved motif Gly-x-Ser-x-Gly in which the catalytic serine is embedded (Supplementary Figure 7). Alternatively, three LEs in family I show variations of this conserved motif. The variations were Ala-x-Ser-x-Gly, Thr-x-Ser-x-Gly (Diamond et al., 2019) and Ser-x-Ser-x-Gly (Dalcin Martins et al., 2018; Supplementary Figure 7).

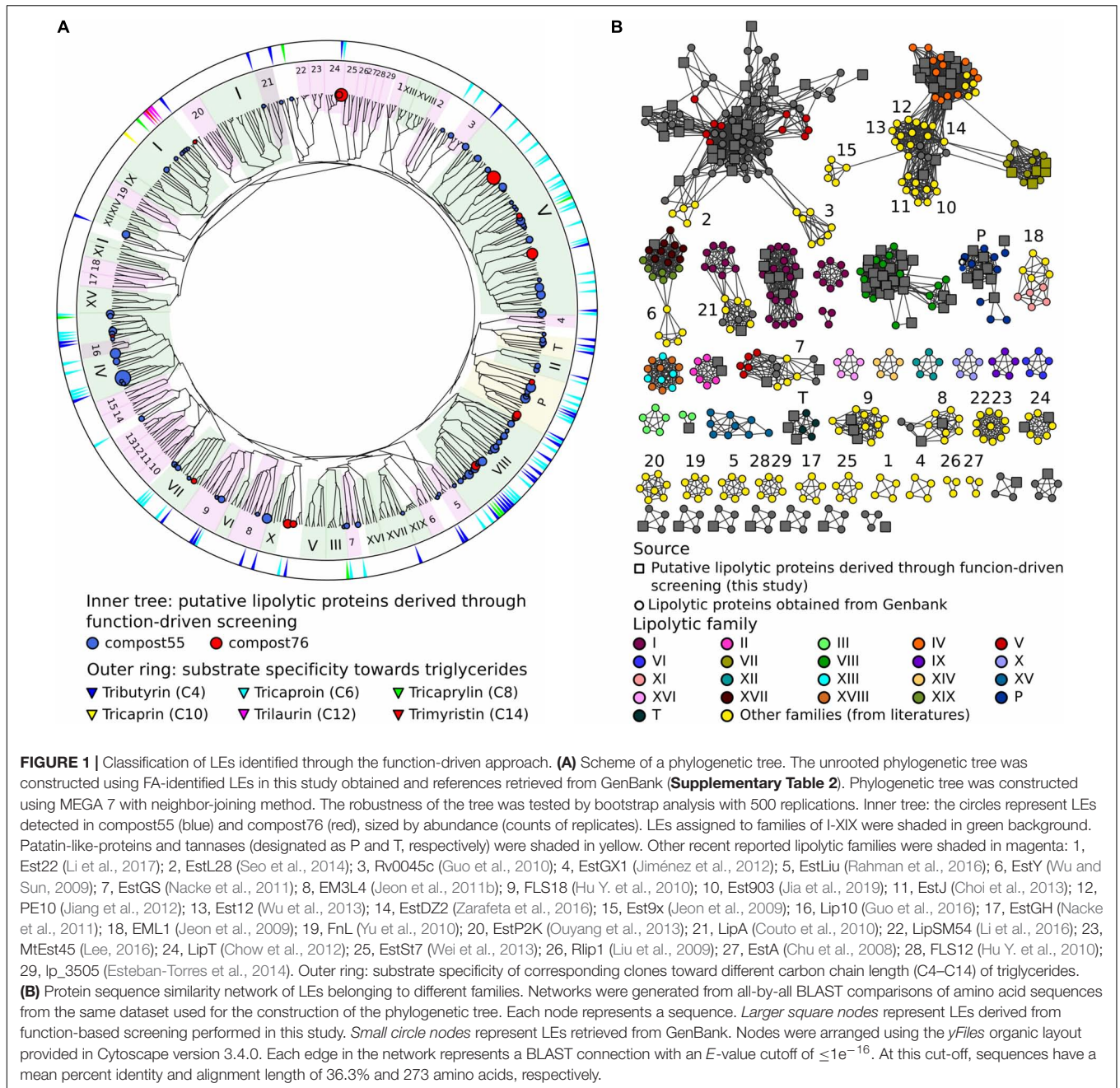
Family II LEs share a canonical α/β -hydrolase fold, which is characterized by a conserved hydrophobic core consisting of five β -strands and at least four α -helices (Akoh et al., 2004). As shown in Supplementary Figure 7A, there are four conservative regions and one conserved residue in each region (serine, glycine, asparagine, and histidine, respectively), which is essential for catalysis (Akoh et al., 2004; Hong et al., 2008). The structures of family VIII enzymes show remarkable sequence similarities to β -lactamases and penicillin-binding proteins (Bornscheuer, 2002). Site-directed mutagenesis demonstrated that the catalytic triad is composed of serine and lysine located in a Ser-X-X-Lys motif, and a tyrosine (Supplementary Figure 7; Biver and Vandenberg, 2013; Kovacic et al., 2019). The patatin-like-proteins display an α/β -hydrolase fold, in which a central six-stranded beta-sheet is sandwiched between alpha-helices front and back (Banerji and Flieger, 2004). Unlike the catalytic triad of Ser-Asp/Glu-His for most lipolytic proteins, the catalytic Ser-Asp dyad is responsible for the catalytic activity of patatin-like-proteins. In addition, they also contained the Gly-x-Ser-x-Gly motif with the catalytic serine embedded (Supplementary Figure 7).

Development of a Lipolytic Enzyme Profile Hidden Markov Model Database for Sequence-Based Screening

Profile HMMs are statistical models that convert patterns, motifs and other properties from a multiple sequence alignment into a set of position-specific hidden states, i.e., frequencies, insertions, and deletions (Reyes et al., 2017). Profile HMMs are sensitive in detecting remote homologs. They have been used to detect, e.g., viral protein sequences (Skewes-Cox et al., 2014; Bzhalava et al., 2018), antibiotic resistance genes (Gibson et al., 2015), GDSL esterase/lipase family genes (Li et al., 2019) in metagenomes.

In this study, a total of 32 ELFs were used for profile HMM database construction (Supplementary Table 10). Subsequently, four profile HMM databases (Omega-pHMMs, Muscle-pHMMs, ClustalW-pHMMs, ESTHER-pHMMs)⁷ specific for LEs affiliated to α/β hydrolase superfamily were constructed. Each database consists of 32 profile HMMs (Supplementary Table 11). The prediction sensitivity and specificity of the four databases were evaluated using four datasets (Table 2). All of the four databases obtained high recalls for the datasets 1, 2, and 3 (Table 2), with the highest ones for omega-pHMMs (4,446 in total), followed by muscle-pHMMs (4,444), ClustalW-pHMMs (4,425), and ESTHER-pHMMs (4,425). Noteworthy, omega-pHMMs did not identify any false positive LEs for dataset 3. Thus, omega-pHMMs was chosen for downstream screening. In addition, we compared omega-pHMMs with the pairwise sequence alignment method (BLASTp) for their ability to predict LEs. The omega-pHMM database exhibited improved sensitivity for datasets 1, 2, and 3. In total, 135 more LEs were identified using omega-pHMMs than BLASTp (Table 2).

⁷<https://github.com/mingji-lu/database-for-lipolytic-enzymes>

**TABLE 2 |** Comparison of profile HMM databases based on different alignment tools to detect LEs.

Datasets ^a	Nr. of LEs (α/β hydrolase)	Nr. of LEs (non- α/β hydrolase)	Recall of LEs (α/β hydrolase)					Recall of LEs (non- α/β hydrolase)	
			Omega-pHMMs	Muscle-pHMMs	ClustalW-pHMMs	ESTHER-pHMMs	BLASTp	Pfam-pHMMs	
Dataset 1	4382	554	4243	4244	4228	4225	4122	554	
Dataset 2	130	32	125	125	121	124	117	32	
Dataset 3	80	36	78	75	76	76	70	36	
Dataset 4	68	0	56	55	53	53	51	0	

^aDataset 1, LEs from UniProt database; Dataset 2, recently reported LEs; Dataset 3, MetaGeneMark-predicted proteins from inserts conferring lipolytic activity; Dataset 4, potential non-lipolytic proteins with homology to LEs.

The accuracy of omega-pHMMs for lipolytic family assignment was also assessed. For datasets 2 and 3, we achieved high precision of annotating LEs into the known lipolytic families but not for LEs from novel families (**Supplementary Table 12**). Dataset 4 included non-lipolytic proteins, such as epoxide hydrolases, dehalogenases and haloperoxidases, which exhibited some amino acid sequence similarity (20–25%) to LEs in subfamilies V.1 and V.2 (Arpigny and Jaeger, 1999). Our “homology-based” method only differentiated part of these non-lipolytic homologs from “true” LEs (**Table 2**). To improve the annotation accuracy, putative lipolytic proteins (PLPs) were further searched against the entire ESTHER database by BLASTp. By combining the annotations from both methods (**Supplementary Figure 1**), these “novel” LEs in datasets 2 and 3 were correctly identified as “unassigned,” in terms of not assigned to any known ELF (**Supplementary Table 12**). Moreover, almost all of the non-lipolytic proteins (>92%) in dataset 4 were distinguished from LEs (**Supplementary Table 12**).

To identify LEs affiliated to families VIII and II, and patatin-like proteins, enzymes were successively screened against Pfam-pHMMs and CATH HMMs database. For the first three datasets, all the LEs in the three families were correctly identified by screening against Pfam-pHMMs (**Table 2** and **Supplementary Table 13**).

As demonstrated in other sequence-based metagenomic approaches (Liu et al., 2015b; Maimanakovs et al., 2016; Azziz et al., 2019), our screening strategy is also vastly dependent on the completeness and accuracy of the reference databases (ESTHER and CATH database in this study). Hence, PLPs exhibiting closest similarity to members affiliated to the miscellaneous ESTHER families or no ESTHER/CATH hits returned, were classified into the “unassigned” group in this study (**Supplementary Table 12**). This might have resulted in an underestimation of assigned lipolytic proteins (**Supplementary Table 13**).

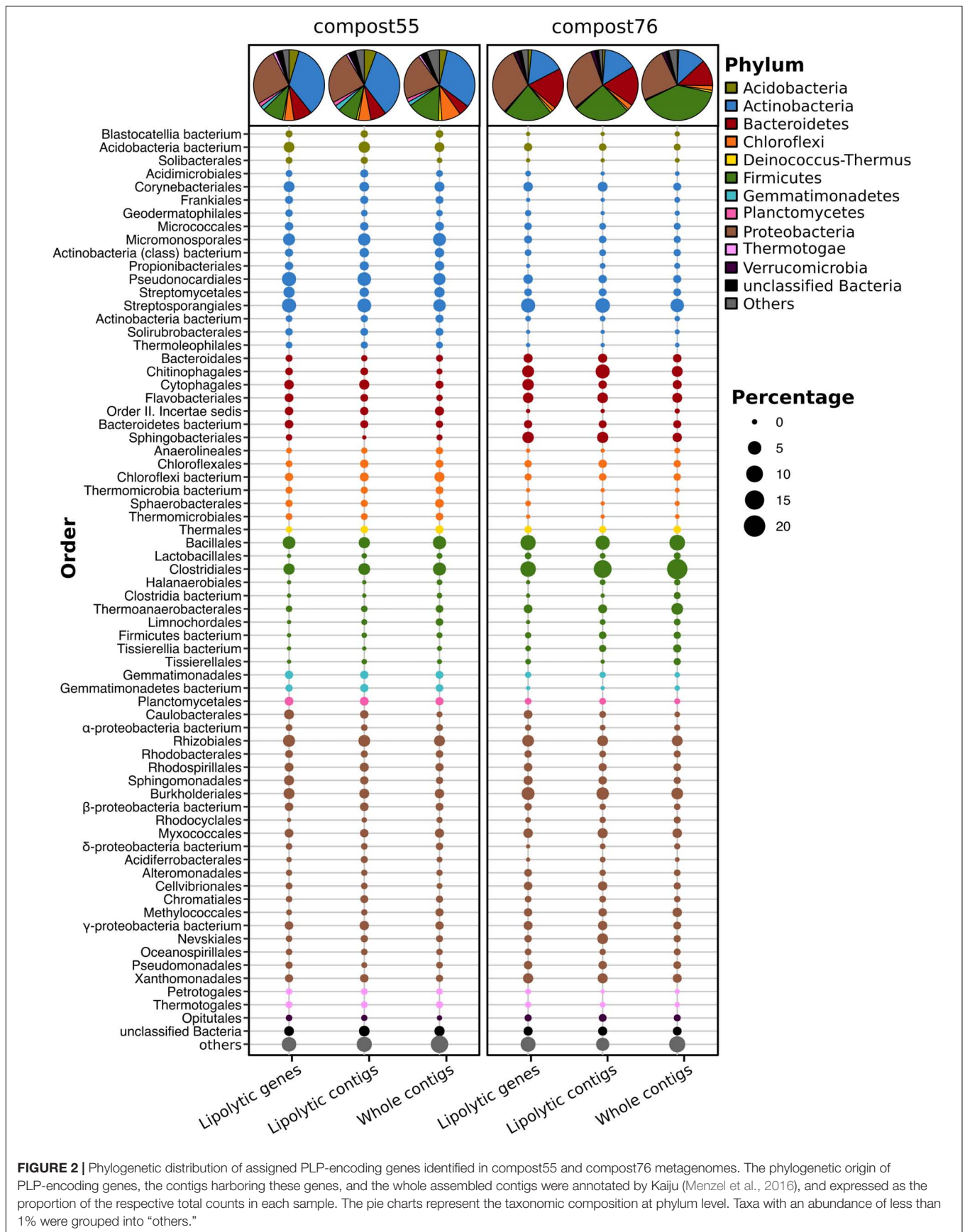
Sequence-Based Screening Confirmed Compost Metagenomes as Reservoir for Putative Lipolytic Genes

Initial screening of the assembled metagenomes of compost55 and compost76 resulted in the identification of 4,157 and 2,234 PLPs, respectively. Among them, 1,234 and 759 were further assigned into 28 and 26 families, respectively. The assigned PLPs belonged mainly to family VIII, hormone-sensitive lipase-like proteins, patatin-like proteins, II, A85-Feruloyl-Esterase, Carb_B_Bacteria and homoserine transacetylase (**Supplementary Figure 8**). The family assignment was also verified by constructing a protein sequence similarity network (**Supplementary Figure 9**). The large number of unassigned PLPs (2,460 for compost55 and 1,208 for compost76) indicated the presence of candidates for novel lipolytic families. The assigned PLPs were generally of bacterial origin (>95%), and mainly affiliated to the phyla (>5%) *Actinobacteria*, *Proteobacteria*, *Firmicutes*, and *Bacteroidetes* (**Figure 2**). The corresponding contigs

were also taxonomically assigned and exhibited a similar phylogeny as seen for the embedded PLP-encoding gene sequences (**Figure 2**).

Members of the *Actinobacteria* have been reported as important biomass degraders (Ryckebøer et al., 2003; Hubbe et al., 2010; Lewin et al., 2016; Wang et al., 2016). In this study, 34.7 (compost55) and 15.8% (compost76) of the assigned PLPs originated from *Actinobacteria*. At genus level, the assigned PLPs were affiliated to *Mycobacterium*, *Actinomadura*, *Thermomonospora*, *Streptomyces*, *Micromonospora*, *Pseudonocardia*, and *Thermobifida* (**Supplementary Table 14**). Members of these genera have been reported as producers of lipases/esterases (Wei et al., 1998; Alisch et al., 2004; Chahinian et al., 2005; Guo et al., 2010; Hu X. et al., 2010; Brault et al., 2012; Mander et al., 2014; Sriyapai et al., 2015). Moreover, some of the corresponding families, such as *Micromonosporaceae*, *Streptomycetaceae*, and *Thermomonosporaceae*, are commonly found in thermophilic composts (Schloss et al., 2003; Blaya et al., 2016; Lima-Junior et al., 2016). *Proteobacteria* are also an abundant source of the assigned PLPs in compost55 (26.2%) and compost76 (31.4%) (**Figure 2**). Popovic et al. (2017) identified 80 LEs, of which 65% were proteobacterial origin by screening of 16 metagenomic DNA libraries prepared from seawater, soils, compost and wastewater. In our study, lipolytic genes exhibited high taxonomic diversity at genus level, they were distributed across 97 and 111 genera for compost55 and compost76, respectively (**Supplementary Table 14**). The assigned PLPs affiliated to *Firmicutes* originated mainly from *Clostridiales* and *Bacillales* (**Figure 2**). By analyzing the microbial diversity and metabolic potential of compost metagenomes, members of *Clostridiales* and *Bacillales* were shown to play key roles in degradation of different organic compounds (Martins et al., 2013; Antunes et al., 2016). *Bacteroidetes* is the fourth most abundant phylum of assigned PLPs in compost55 (8.4%) and compost76 (18.8%) (**Figure 2**). At genus level, the assigned PLPs derived mainly from *Rhodothermus* in compost55, and *Sphingobacterium*, *Flavobacterium*, *Niastella*, and *Flaviumibacter* in compost76 (**Supplementary Table 14**). Members of these genera are known as important fermenters during composting (Neher et al., 2013; Antunes et al., 2016; Lapébie et al., 2019).

The phylogenetic distribution of assigned PLPs in each sample, to some extent, corresponded well to the taxonomic composition revealed from the whole contigs (**Figure 2**) but with minor differences in the rank abundance order. The 16S rRNA gene amplicon (**Supplementary Figure 3**) and metagenomic datasets (**Supplementary Figure 4**) also showed a composition of dominant orders similar to that deduced from lipolytic genes/contigs (**Figure 2**). Wang et al. (2016) showed that the phylogenetic distribution of CAZyme genes in the rice straw-adapted compost consortia was in accordance to its microbial composition. Mapping resistance gene dissemination between humans and their environment by Pehrsson et al. (2016) revealed that resistomes across habitats were generally structured by bacterial phylogeny along ecological gradients.



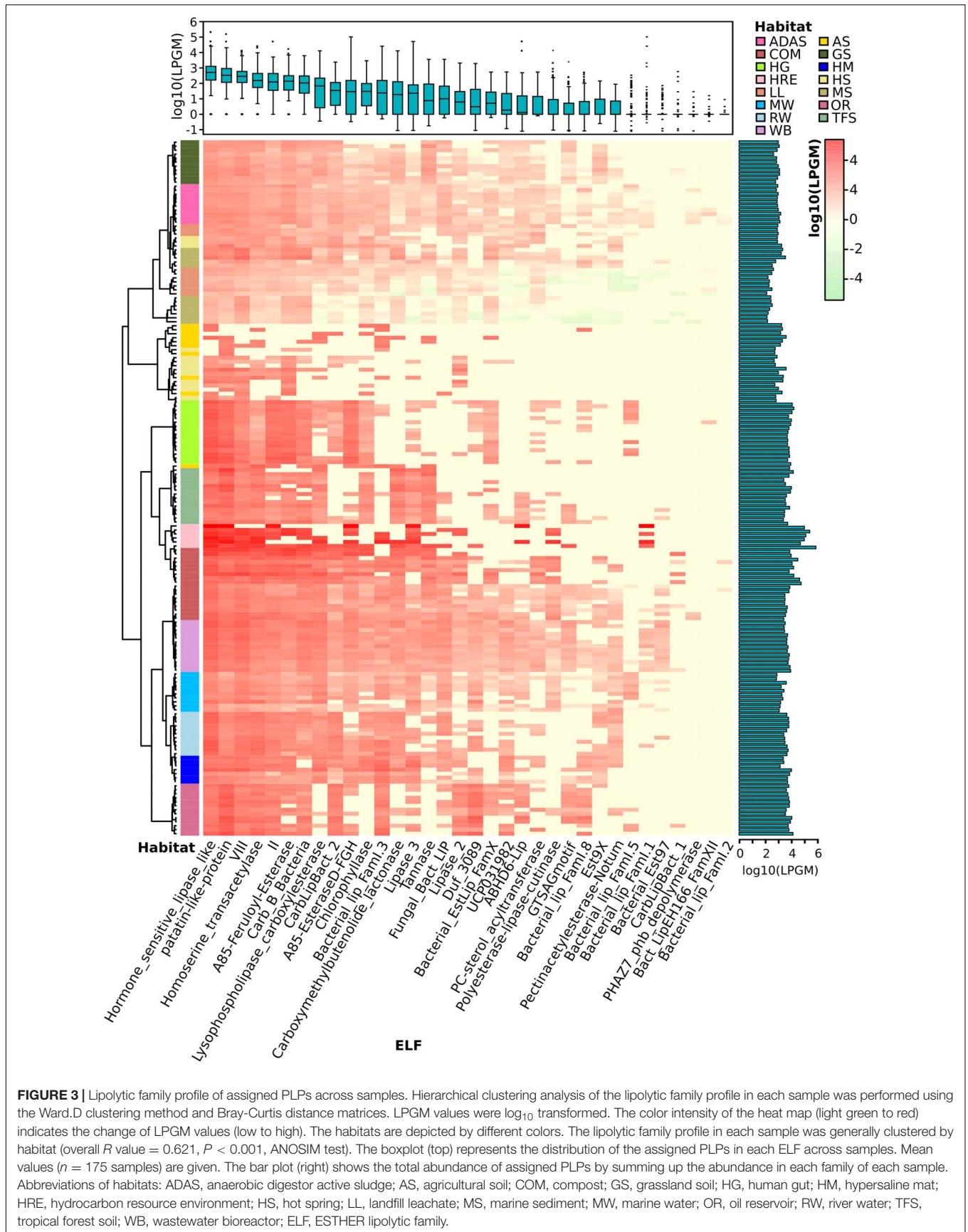


FIGURE 3 | Lipolytic family profile of assigned PLPs across samples. Hierarchical clustering analysis of the lipolytic family profile in each sample was performed using the Ward.D clustering method and Bray-Curtis distance matrices. LPGM values were log₁₀ transformed. The color intensity of the heat map (light green to red) indicates the change of LPGM values (low to high). The habitats are depicted by different colors. The lipolytic family profile in each sample was generally clustered by habitat (overall *R* value = 0.621, *P* < 0.001, ANOSIM test). The boxplot (top) represents the distribution of the assigned PLPs in each ELF across samples. Mean values (*n* = 175 samples) are given. The bar plot (right) shows the total abundance of assigned PLPs by summing up the abundance in each family of each sample. Abbreviations of habitats: ADAS, anaerobic digester active sludge; AS, agricultural soil; COM, compost; GS, grassland soil; HG, human gut; HM, hypersaline mat; HRE, hydrocarbon resource environment; HS, hot spring; LL, landfill leachate; MS, marine sediment; MW, marine water; OR, oil reservoir; RW, river water; TFS, tropical forest soil; WB, wastewater bioreactor; ELF, ESTHER lipolytic family.

Comparison Between Function-Driven and Sequence-Based Screening of Lipolytic Enzymes

Metagenomics allows tapping into the rich genetic resources of so far uncultured microorganisms (Simon and Daniel, 2011) through function-driven or sequence-based approaches. The function-driven strategy targets a particular activity of metagenomic library-bearing hosts (Ngara and Zhang, 2018). In this way, we identified 13 novel LEs (**Supplementary Table 9** and **Supplementary Figure 7B**), which confirmed functional screening as a valuable approach for discovering entirely novel classes of genes and enzymes, particularly when the function could not be predicted based on DNA sequence alone (Reyes-Duarte et al., 2012; Lam et al., 2015; Villamizar et al., 2017).

The sequence-based screening strategy is also frequently used due to the easy access to a wealth of metagenome sequence data and continuous advances in bioinformatics (Chan et al., 2010; Liu et al., 2015c; Maimanakis et al., 2016). The hit rate for LEs was higher by sequence-based than by function-based screening, but the sequence-based derived hits need to be functionally verified. By mapping the metagenomic short reads to the functional screening-derived lipolytic genes, 63 genes out of 115 lipolytic genes in total had a coverage of 100% and 88 of $\geq 99\%$ (**Supplementary Table 15**). Blast-based comparison between lipolytic genes derived from function-driven and sequence-based approaches indicated that 31 genes from each approach exhibited 100% sequence identity (**Supplementary Table 16**).

Function-driven screenings are generally constrained by factors, such as labor-intensive operation, limitations of the employed host systems and low hit rate (Simon and Daniel, 2011). However, function-based approaches are activity-directed, and sequence- and database-independent, thus, they bear the potential to discover entirely novel genes for biomolecules of interest (Rabausch et al., 2013; Lam et al., 2015). Sequencing-based screening, on the other hand, is effective in identifying sequences and potential genes encoding targeted biomolecules in metagenomes. Sequence-based screens largely rely on the used search algorithms, and quality and content of the reference databases to infer the functions of discovered candidate genes (Ngara and Zhang, 2018). Thus, the best way to explore novel molecules is to combine the two approaches (Barriuso and Jesús Martínez, 2015).

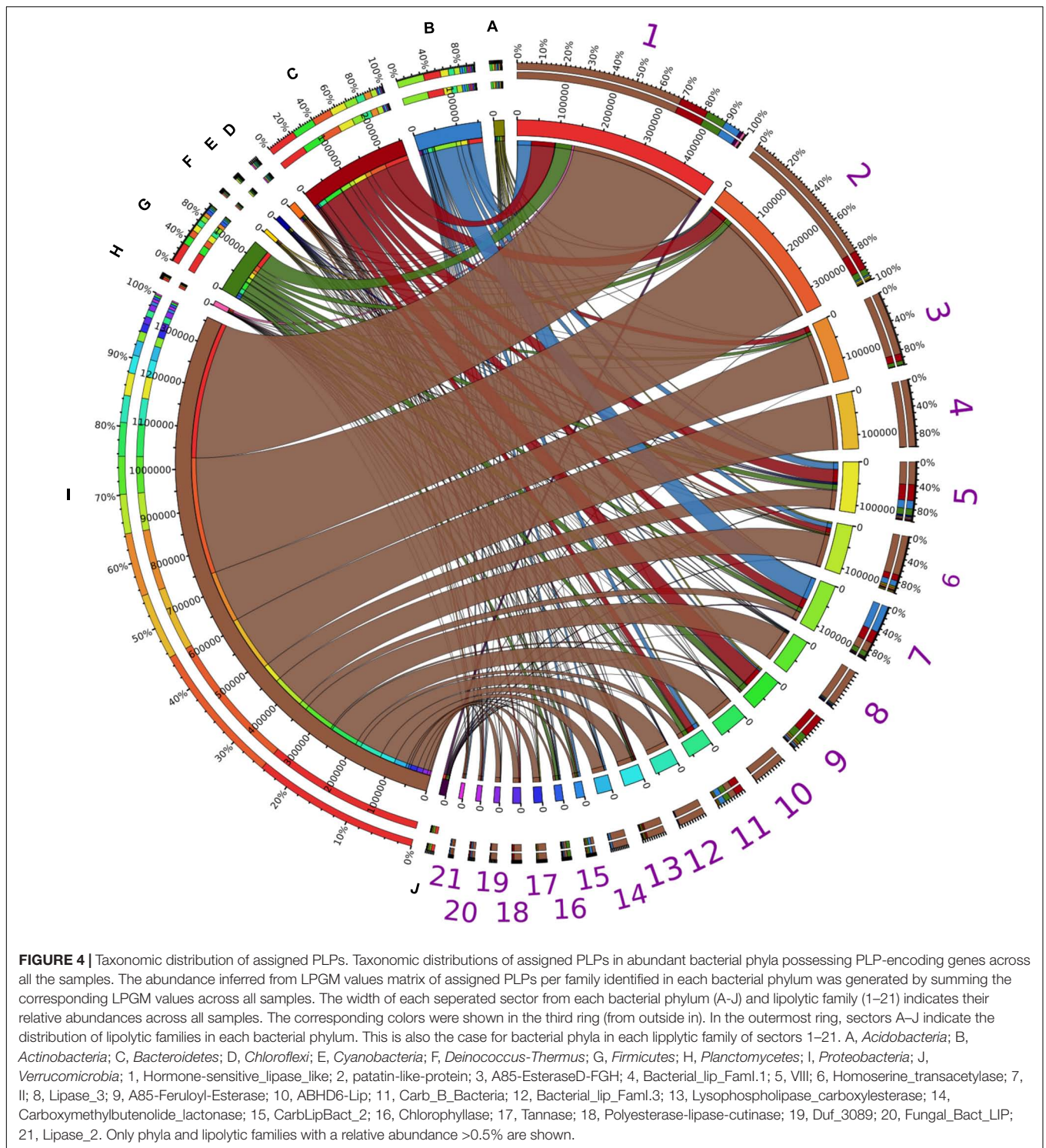
Assigned Putative Lipolytic Proteins Are Distributed by Ecological Factors

In this study, 175 metagenomes representing various ecology niches were selected for sequence-based searching of PLPs. In total, we have screened approx. 1.23 billion genes in assembled metagenomes and recovered approx. 0.22 million (absolute counts) PLP-encoding genes. The assigned PLPs (34% of the total counts) were normalized to LPGM values for comparative analysis. In accordance with the function-based screening, samples subjected to certain enrichment processes, particularly lipid-related, tend to have a higher hit rate (**Figure 3**). For example, samples with high LPGM values were derived from

a hydrocarbon resource environment and an oil reservoir that are enriched with oil-degrading microbes (Liu et al., 2015a, 2018; Hu et al., 2016; Vigneron et al., 2017), and composts and wastewater bioreactors that are reservoir for microbes degrading organic compounds (Dougherty et al., 2012; Silva et al., 2012; Antunes et al., 2016; Berini et al., 2017). Intriguingly, samples from human gut systems were also candidates for LEs (LPGM values $> 7,500$). The human intestinal microorganisms play an important role in degrading diet components into metabolizable molecules (Wang et al., 2015). The function- and sequence-based study of human gut metagenomes have proved that the human gut microbiome is a rich source for various carbohydrate active enzymes (Li et al., 2009; Turnbaugh et al., 2009; Tasse et al., 2010; Moore et al., 2011).

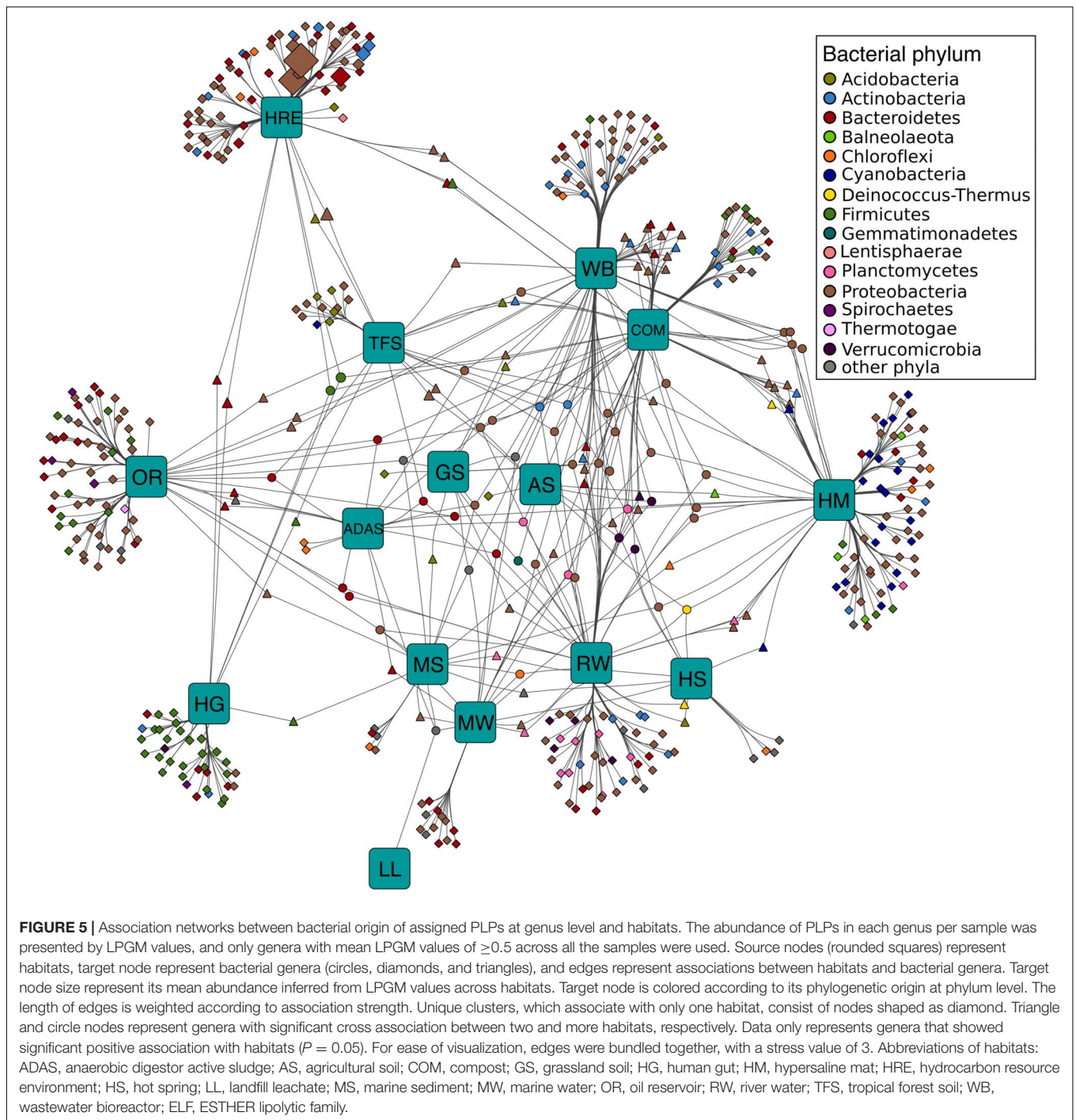
Overall, the assigned PLPs were classified into 34 lipolytic families (**Figure 3**). Members of the hormone-sensitive_lipase_like and patatin-like-protein families were most abundant (average LPGM values across samples $> 2,000$), followed by families of A85-EsteraseD-FGH, VIII and Bacterial_lip_FamI.1 (average LPGM values > 700 ; **Figure 3**). However, no family was shared by all samples. Nevertheless, members from families of hormone-sensitive-lipase-like, patatin-like-proteins, VIII, homoserine transacetylase, II and A85-Feruloyl-Esterase were detected in more than 90% of samples (**Figure 3**). Enzymes belonging to families of PHAZ7_phb_depolymerase, Bact_LipEH166_FamXII, and Bacterial_lip_FamI.2 were not or only rarely detected ($< 6\%$ of all samples) and showed a low abundance (LPGM values < 1). The prevalence and abundance of a lipolytic family revealed by the sequence-based screening are dependent on the distribution of the corresponding target genes in the microbial consortia (Wang et al., 2016). Taking members from the “abundant” family hormone-sensitive_lipase_like as example, the corresponding genes are widely distributed in more than 1,200 species as recorded in the ESTHER database so far. This was, somehow, also reflected by the function-based screening, in which a large proportion of the identified LEs belonged to the hormone-sensitive_lipase_like family. In contrast, according to the ESTHER database, only 23, 8 and 6 species harboring LEs were affiliated to the “rare” families like PHAZ7_phb_depolymerase, PC-sterol_acyltransferase and Bact_LipEH166_FamXII, respectively.

To investigate the distribution of assigned PLPs that cause the observed lipolytic family profiles across samples and habitats, a matrix with LPGM values representing the abundance of PLPs per lipolytic family identified in each metagenome was generated. The lipolytic family profiles clustered by habitats (**Figure 3**), which was confirmed by NMDS (**Supplementary Figure 10**). ANOSIM (Clarke, 1993) was used to pairwise compare the multivariate (group) differences of lipolytic family profiles between habitats. A *R* value-based matrix was generated among habitats (**Supplementary Figure 10**), a high *R* value (between 0 and 1) indicated a high group dissimilarity between two habitats. Generally, each habitat exhibited a distinctive pattern of lipolytic family profiles (overall *R* value = 0.6168; **Supplementary Table 17**). For example, PLPs detected in agricultural soils were only present in eight lipolytic families with



low abundances. In contrast, PLPs in composts were detected in almost all lipolytic families, and with remarkably high abundance in families such as hormone-sensitive_lipase_like, patatin-like-protein and VIII (Supplementary Figure 11). Notably, the lowest group dissimilarity was observed between the habitats compost and wastewater bioreactor ($R = 0.1941$, $P < 0.001$,

ANISOM; Supplementary Figure 10). The analysis of lipolytic profiles across habitats allows selecting suitable habitats for function-based screening, e.g., targeting LEs of a specific family or with some properties for desired applications. Metagenomes from composts are promising for recovering LEs in families Lysophospholipase_carboxylesterase (family VI), CarbLipBact_2



(family XIII-2/XVIII), and CarbLipBact_1 (family XIII-1) (Supplementary Figure 12).

The Phylogenetic Distribution of Assigned Putative Lipolytic Proteins

More than 98% of the assigned PLPs were encoded by bacterial community members. Although LEs are widely encoded in various microbial genomes (Hausmann and Jaeger, 2010;

Ramnath et al., 2016; Kovacic et al., 2019), the assigned PLPs were mainly derived from the bacterial phyla *Proteobacteria* (66.5%), *Bacteroidetes* (12.5%), *Actinobacteria* (7.7%), *Firmicutes* (6.7%) (Figure 4). This is consistent with the taxonomic origin of reference LEs in ESTHER database (Supplementary Figure 13). Moreover, enzymes from members of *Proteobacteria* were dominant in almost all lipolytic families (Figure 4). At genus level, the phylogenetic origins of assigned PLPs were scattered across approx. 2,000 bacterial genera

(**Supplementary Figure 14**), with enriched abundance in the genera *Acinetobacter*, *Pseudomonas*, *Bacteroides*, *Bradyrhizobium*, and *Mycobacterium* (average LPGM values across samples >180). Many of the LEs from these genera were described as exoenzymes (Rudek and Haque, 1976; Gilbert, 1993; Snellman and Colwell, 2004; Guo et al., 2010; Liu et al., 2018). Notably, a similar taxonomic enrichment at genus level was also observed for the reference LEs in ESTHER database as 960 LEs were encoded by *Mycobacterium*, 410 by *Pseudomonas*, 260 by *Bacteroides*, 166 by *Acinetobacter*, and 164 by *Bradyrhizobium* species (**Supplementary Table 18**).

The taxonomic origin of assigned PLPs at genus level varied significantly across habitats (overall R value = 0.821, $P < 0.01$), especially for the human gut system, oil reservoir and hydrocarbon resource environment (**Supplementary Figure 15**). The average R value was 0.98, 0.97, and 0.94, respectively (**Supplementary Table 19**). The lowest dissimilarity was observed between compost and wastewater bioreactor (R value = 0.2317, $P < 0.001$, ANISOM).

Habitats Harboring Prevalent and Distinct Microbial Clusters Are Main Drivers of Putative Lipolytic Protein Distribution

Bipartite association networks have been used to identify microbial taxa responsible for shifts in community structures (Hartmann et al., 2015; Dukunde et al., 2019). In this study, a bipartite association network was constructed to visualize the associations between bacterial members at genus level that harbor lipolytic genes and habitats or habitat combinations (**Figure 5**). 225 of the total 712 genera were not significantly separated in abundance and frequency by habitat. These belonged mainly to *Proteobacteria* (82 genera), *Bacteroidetes* (43 genera), *Firmicutes* (33 genera), and *Actinobacteria* (25 genera) (**Supplementary Table 20**). These non-significant genera were conserved across different habitats, generally represented the “indigenous group” (Hartmann et al., 2015; Wemheuer et al., 2017), and formed the core microbiota harboring lipolytic genes. This core microbiota was also an indication of the prevalence of lipolytic genes across microbes and habitats (Bornscheuer, 2002; Hasan et al., 2006; Barriuso and Jesús Martínez, 2015; Berini et al., 2017). In contrast, the significant indicators, with respect to the “characteristic group” (Rime et al., 2016; Dukunde et al., 2019), highlighted the bacterial genera that were responsible for the change of assigned PLPs distribution across habitats (**Figure 5**). Particularly, the indicators associated with only one habitat defined the distinctiveness of microbiota in each habitat (Hartmann et al., 2015). In this study, the unique-associated indicators accounted for 76% of all significant indicators (**Supplementary Table 20**). This strongly resembled the ANISOM result, in which the high overall R value (0.8199) suggested a significant distinctiveness of the phylogenetic origins of assigned PLPs across habitats (**Supplementary Table 19**). With respect to each habitat, a high ratio of unique-associated indicators to the total significant genera in a habitat generally indicated a high R value (Pearson’s r correlation = 0.6672,

$P < 0.01$, linear regression; **Supplementary Figure 16**). For example, out of the 75 indicators that were significantly associated to the habitat hydrocarbon resource environment, 65 were unique-associated indicators with a mean R value of 0.93 (**Supplementary Table 19**). This is also the case for the habitats oil reservoir (60 out of 75; mean R value = 0.96) and human gut system (35 out of 41; mean R value = 0.97).

Only a small fraction of the indicators exhibited cross associations between two (14% of the total indicators) or more (10%) habitats. Nevertheless, the 29 cross-associated indicators between habitats compost and wastewater bioreactor explained the low dissimilarity of phylogenetic distributions of assigned PLPs between the two habitats ($R = 0.2317$, $P < 0.001$, ANISOM).

Similar to the “indigenous group,” the “characteristic group” consisted mainly of genera affiliated to *Proteobacteria* (224 genera), *Bacteroidetes* (72), *Firmicutes* (49), and *Actinobacteria* (36). Among them, proteobacterial genera largely characterized the major habitats, such as tropical forest soil (83%), wastewater bioreactor (67%), hypersaline mat (52%), hydrocarbon resource environment (51%), oil reservoir (51%), compost (50%), marine water (46%), river water (45%), and grassland soil (42%), whereas *Bacteroidetes* and *Firmicutes* characterized the human gut system (68%) and the active sludge of an anaerobic digester (53%) (**Supplementary Table 20**). Noteworthy, the unique-associated indicators affiliated to *Cyanobacteria* were primarily enriched in the hypersaline mat (95% indicators), which is also the case for *Planctomycetes* and *Verrucomicrobia* in river water (88 and 80%, respectively). Pehrsson et al. (2016) detected a link between microbial community structure and functional gene repertoire. This link could be extended to the distribution pattern of indicators in our study. For example, various studies have proved that the microbes in human gut systems were dominated by *Firmicutes* (Mahowald et al., 2009; Vital et al., 2014; Rinninella et al., 2019), which in turn leads to the *Firmicutes*-dominated indicators for lipolytic genes (**Figure 5**). Among all the habitats, only hypersaline mats were featured by the *Cyanobacteria*-dominated oxygenic layer for photosynthesis (Sørensen et al., 2005; Lindemann et al., 2013), which explained that almost all the *Cyanobacteria* indicators were associated with the hypersaline mat (**Figure 5**).

CONCLUSION

In this study, two compost samples (compost55 and compost76) were used for metagenomic screening of potential lipolytic genes. Through the function-driven screening, 115 unique LEs were identified and assigned into 12 known lipolytic families. In addition, 7 LEs were not assigned to any known family, indicating new branches of lipolytic families. Our results show that functional screening is a promising approach to discover novel lipolytic genes, particularly for targeted genes, whose function is not predicted based on DNA sequence alone. For sequence-based screening, we have developed a search and annotation strategy specific for putative lipolytic genes in metagenomes (**Supplementary Figure 1**). Our profile HMM-based searching methods yielded higher sensitivity (recall) for

LEs than the BLASTp-derived counterpart. The annotation method also remarkably increased the specificity and accuracy in distinguishing lipolytic from non-lipolytic proteins. With this sequence-based strategy, we identified the putative lipolytic genes within the two compost metagenomes. Analysis of the taxonomic origin of these genes indicated a potential link between microbial taxa and their functional traits. By comparing the lipolytic hits identified by function-driven and sequence-based screening, we conclude that the best way for exploring and exploiting LEs is to combine both approaches.

In addition, assembled metagenomes from samples of various habitats were used for comparative analysis of the PLP distribution. We profiled the lipolytic family and phylogenetic origin of assigned PLPs for each sample. The two profiles were generally driven by ecological factors, i.e., the habitat. Moreover, the habitat also determined the conserved and distinctive microbial groups harboring the putative lipolytic genes.

Putative lipolytic proteins were also mainly enriched in the bacterial phyla *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Firmicutes* (**Supplementary Figure 17**). The profile of the phylogenetic total PLP distribution in each sample clustered also by habitats (**Supplementary Figures 18–20**). The bipartite association network identified the conserved and distinctive microbial groups harboring PLP-encoding genes among the habitats (**Supplementary Tables 21, 22**). Thus, our study provided a sequence-based strategy for effective identification and annotation of potential lipolytic genes in assembled metagenomes. More importantly, through this strategy, the overview of how the lipolytic genes distributed ecologically (in various habitats), functionally (in different lipolytic enzyme families), and phylogenetically (in diverse microbial groups) is an advantage for novel and/or industrially relevant LE identification.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SRR13115019; <https://www.ncbi.nlm.nih.gov/>,

SRR13115018; <https://www.ncbi.nlm.nih.gov/>, SAMN06859928; <https://www.ncbi.nlm.nih.gov/>, SAMN06859946; <https://www.ncbi.nlm.nih.gov/>, SAMN06859935; <https://www.ncbi.nlm.nih.gov/>, SAMN06859953; <https://www.ncbi.nlm.nih.gov/genbank/>, MW408002-MW408112.

AUTHOR CONTRIBUTIONS

RD conceived and supervised the study. ML planned and performed the experiments, analyzed the data, and wrote the first draft of the manuscript. DS planned sequencing and advised analysis of sequencing data. ML and DS validated data. All authors interpreted the results, and reviewed and revised the manuscript.

FUNDING

We acknowledge the support of ML by “Erasmus Mundus Action 2 – Lotus I Project” and Open Access Publication Funds of the University of Göttingen. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

ACKNOWLEDGMENTS

We thank Silja Brady and Mechthild Bömeke for providing technical assistance and support. We acknowledge support by the Open Access Publication Funds of the University of Göttingen. A version of the manuscript is available as preprint (Lu et al., 2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.851969/full#supplementary-material>

REFERENCES

- Akmoussi-Toumi, S., Khemili-Talbi, S., Ferioun, I., and Kebbouche-Gana, S. (2018). Purification and characterization of an organic solvent-tolerant and detergent-stable lipase from *Haloferax mediterranei* CNCMM 50101. *Int. J. Biol. Macromol.* 116, 817–830. doi: 10.1016/j.ijbiomac.2018.05.087
- Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004). GDSL family of serine esterases/lipases. *Prog. Lipid Res.* 43, 534–552. doi: 10.1016/j.plipres.2004.09.002
- Alisch, M., Feuerhack, A., Müller, H., Mensak, B., Andreas, J., and Zimmermann, W. (2004). Biocatalytic modification of polyethylene terephthalate fibres by esterases from actinomycete isolates. *Biocatal. Biotransform.* 22, 347–351. doi: 10.1080/10242420400025877
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Antunes, L. P., Martins, L. F., Pereira, R. V., Thomas, A. M., Barbosa, D., Lemos, L. N., et al. (2016). Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics. *Sci. Rep.* 6:38915. doi: 10.1038/srep38915
- Arpigny, J. L., and Jaeger, K.-E. (1999). Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* 343, 177–183. doi: 10.1042/0264-6021:3430177
- Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3:e1029. doi: 10.7717/peerj.1029
- Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4:e4345. doi: 10.1371/journal.pone.0004345
- Azziz, G., Giménez, M., Romero, H., Valdespino-Castillo, P. M., Falcón, L. I., Ruberto, L. A. M., et al. (2019). Detection of presumed genes encoding beta-lactamases by sequence based screening of metagenomes derived from

- Antarctic microbial mats. *Front. Environ. Sci. Eng.* 13:44. doi: 10.1007/s11783-019-1128-1
- Banerji, S., and Flioger, A. (2004). Patatin-like proteins: A new family of lipolytic enzymes present in bacteria? *Microbiology* 150, 522–525. doi: 10.1099/mic.0.26957-0
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barriuso, J., and Jesús Martínez, M. (2015). In silico metagenomes mining to discover novel esterases with industrial application by sequential search strategies. *J. Microbiol. Biotechnol.* 25, 732–737. doi: 10.4014/jmb.1406.06049
- Berglund, F., Marathe, N. P., Österlund, T., Bengtsson-Palme, J., Kotsakis, S., Flach, C. F., et al. (2017). Identification of 76 novel B1 metallo- β -lactamases through large-scale screening of genomic and metagenomic data. *Microbiome* 5:134. doi: 10.1186/s40168-017-0353-8
- Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017). Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* 364:fnx211. doi: 10.1093/femsle/fnx211
- Berlemont, R., Spee, O., Delsaute, M., Lara, Y., Schuldes, J., Simon, C., et al. (2013). Novel organic solvent-tolerant esterase isolated by metagenomics: insights into the lipase/esterase classification. *Rev. Argent. Microbiol.* 45, 3–12.
- Biver, S., and Vandenbol, M. (2013). Characterization of three new carboxylic ester hydrolases isolated by functional screening of a forest soil metagenomic library. *J. Ind. Microbiol. Biotechnol.* 40, 191–200. doi: 10.1007/s10295-012-1217-7
- Blaya, J., Marhuenda, F. C., Pascual, J. A., and Ros, M. (2016). Microbiota characterization of compost using omics approaches opens new perspectives for phytophthora root rot control. *PLoS One* 11:e0158048. doi: 10.1371/journal.pone.0158048
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bornscheuer, U. T. (2002). Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol. Rev.* 26, 73–81. doi: 10.1111/j.1574-6976.2002.tb00599.x
- Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G. W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9, 435–426. doi: 10.1038/nmeth.1990
- Brault, G., Shareck, F., Hurtubise, Y., Lépine, F., and Doucet, N. (2012). Isolation and characterization of EstC, a new cold-active esterase from *Streptomyces coelicolor* A3(2). *PLoS One* 7:e32041. doi: 10.1371/journal.pone.0032041
- Bzhalava, Z., Hultin, E., and Dillner, J. (2018). Extension of the viral ecology in humans using viral profile hidden Markov models. *PLoS One* 13:e0190938. doi: 10.1371/journal.pone.0190938
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Castilla, A., Panizza, P., Rodríguez, D., Bonino, L., Díaz, P., Irazoqui, G., et al. (2017). A novel thermophilic and halophilic esterase from *Janibacter* sp. R02, the first member of a new lipase family (Family XVII). *Enzyme Microb. Technol.* 98, 86–95. doi: 10.1016/j.enzmictec.2016.12.010
- Chahinian, H., Ali, Y. B., Abousalham, A., Petry, S., Mandrich, L., Manco, G., et al. (2005). Substrate specificity and kinetic properties of enzymes belonging to the hormone-sensitive lipase family: Comparison with non-lipolytic and lipolytic carboxylesterases. *Biochim. Biophys. Acta* 1738, 29–36. doi: 10.1016/j.bbap.2005.11.003
- Chan, W. Y., Wong, M., Guthrie, J., Savchenko, A. V., Yakunin, A. F., Pai, E. F., et al. (2010). Sequence- and activity-based screening of microbial genomes for novel dehalogenases. *Microb. Biotechnol.* 3, 107–120. doi: 10.1111/j.1751-7915.2009.00155.x
- Chen, I. M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017). IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, 507–516. doi: 10.1093/nar/gkw929
- Chen, Y., Black, D. S., and Reilly, P. J. (2016). Carboxylic ester hydrolases: Classification and database derived from their primary, secondary, and tertiary structures. *Protein Sci.* 25, 1942–1953. doi: 10.1002/pro.3016
- Chevreur, B., Wetter, T., and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Comp. Sci. Biol. Proc. Germ. Conf. Bioinform.* 99, 45–56.
- Choi, J. E., Kwon, M. A., Na, H. Y., Hahm, D. H., and Song, J. K. (2013). Isolation and characterization of a metagenome-derived thermoalkaliphilic esterase with high stability over a broad pH range. *Extremophiles* 17, 1013–1021. doi: 10.1007/s00792-013-0583-z
- Chow, J., Kovacic, F., Dall, A. Y., Krauss, U., Fersini, F., Schmeisser, C., et al. (2012). The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS One* 7:e47665. doi: 10.1371/journal.pone.0047665
- Chu, X., He, H., Guo, C., and Sun, B. (2008). Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl. Microbiol. Biotechnol.* 80, 615–625. doi: 10.1007/s00253-008-1566-3
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x
- Couto, G. H., Glogauer, A., Faoro, H., Chubatsu, L. S., Souza, E. M., and Pedrosa, F. O. (2010). Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genet. Mol. Res.* 9, 514–523. doi: 10.4238/vol9-1gmr738
- Dalcin Martins, P., Danczak, R. E., Roux, S., Frank, J., Borton, M. A., Wolfe, R. A., et al. (2018). Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* 6:138. doi: 10.1186/s40168-018-0522-4
- De Cáceres, M. (2013). How to use the indicpecies package (ver. 1.7.1). *R Proj.* 2013:29.
- Diamond, S., Andeer, P. F., Li, Z., Crits-Christoph, A., Burstein, D., Anantharaman, K., et al. (2019). Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat. Microbiol.* 4, 1356–1367. doi: 10.1038/s41564-019-0449-y
- Dougherty, M. J., D'haeseleer, P., Hazen, T. C., Simmons, B. A., Adams, P. D., and Hadi, M. Z. (2012). Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnol.* 12:38. doi: 10.1186/1472-6750-12-38
- Dukunde, A., Schneider, D., Lu, M., Brady, S., and Daniel, R. (2017). A novel, versatile family IV carboxylesterase exhibits high stability and activity in a broad pH spectrum. *Biotechnol. Lett.* 39, 577–587. doi: 10.1007/s10529-016-2282-1
- Dukunde, A., Schneider, D., Schmidt, M., Veldkamp, E., and Daniel, R. (2019). Tree species shape soil bacterial community structure and function in temperate deciduous forests. *Front. Microbiol.* 10:1519. doi: 10.3389/fmicb.2019.01519
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Egelkamp, R., Zimmermann, T., Schneider, D., Hertel, R., and Daniel, R. (2019). Impact of nitriles on bacterial communities. *Front. Environ. Sci.* 7:103. doi: 10.3389/fenvs.2019.00103
- Esteban-Torres, M., Santamaría, L., de las Rivas, B., and Muñoz, R. (2014). Characterisation of a cold-active and salt-tolerant esterase from *Lactobacillus plantarum* with potential application during cheese ripening. *Int. Dairy J.* 39, 312–315. doi: 10.1016/j.idairyj.2014.08.004
- Fan, X., Liu, X., Wang, K., Wang, S., Huang, R., and Liu, Y. (2011). Highly soluble expression and molecular characterization of an organic solvent-stable and thermotolerant lipase originating from the metagenome. *J. Mol. Catal. B: Enzymatic* 72, 319–325. doi: 10.1016/j.molcatb.2011.07.009
- Fang, Z., Li, J., Wang, Q., Fang, W., Peng, H., Zhang, X., et al. (2014). A novel esterase from a marine metagenomic library exhibiting salt tolerance ability. *J. Microbiol. Biotechnol.* 24, 771–780. doi: 10.4014/jmb.1311.11071
- Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2015). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* 9, 22–34. doi: 10.1111/1751-7915.12309

- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Fu, L., He, Y., Xu, F., Ma, Q., Wang, F., and Xu, J. (2015). Characterization of a novel thermostable patatin-like protein from a Guaymas basin metagenomic library. *Extremophiles* 19, 829–840. doi: 10.1007/s00792-015-0758-x
- Gao, W., Wu, K., Chen, L., Fan, H., Zhao, Z., Gao, B., et al. (2016). A novel esterase from a marine mud metagenomic library for biocatalytic synthesis of short-chain flavor esters. *Microb. Cell Fact.* 15:41. doi: 10.1186/s12934-016-0435-5
- Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., et al. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* 1854, 1019–1037. doi: 10.1002/cncr.27633. Percutaneous
- Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 9, 207–216. doi: 10.1038/ismej.2014.106
- Gilbert, E. J. (1993). *Pseudomonas* lipases: Biochemical properties and molecular cloning. *Enzyme Microb. Technol.* 15, 634–645. doi: 10.1016/0141-0229(93)90062-7
- Glogauer, A., Martini, V. P., Faoro, H., Couto, G. H., Müller-Santos, M., Monteiro, R. A., et al. (2011). Identification and characterization of a new true lipase isolated through metagenomic approach. *Microb. Cell Fact.* 10:54. doi: 10.1186/1475-2859-10-54
- Guo, H., Zhang, Y., Shao, Y., Chen, W., Chen, F., and Li, M. (2016). Cloning, expression and characterization of a novel cold-active and organic solvent-tolerant esterase from *Monascus ruber* M7. *Extremophiles* 20, 451–459. doi: 10.1007/s00792-016-0835-9
- Guo, J., Zheng, X., Xu, L., Liu, M. Z., Xu, K., Li, S., et al. (2010). Characterization of a novel esterase Rv0045c from *Mycobacterium tuberculosis*. *PLoS One* 5:e13143. doi: 10.1371/journal.pone.0013143
- Hammer, O., Harper, D. A. T., and Ryan, P. D. (2001). Past: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electr.* 4:4.
- Hartmann, M., Frey, B., Mayer, J., Mäder, P., and Widmer, F. (2015). Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J.* 9, 1177–1194. doi: 10.1038/ismej.2014.210
- Hasan, F., Shah, A. A., and Hameed, A. (2006). Industrial applications of microbial lipases. *Enzyme Microb. Technol.* 39, 235–251. doi: 10.1016/J.ENZMICTEC.2005.10.016
- Hausmann, S., and Jaeger, K.-E. (2010). “Lipolytic enzymes from bacteria,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed. K. N. Timmis (Berlin: Springer), 1099–1126. doi: 10.1007/978-3-540-77587-4_77
- Hita, E., Robles, A., Camacho, B., González, P. A., Esteban, L., Jiménez, M. J., et al. (2009). Production of structured triacylglycerols by acidolysis catalyzed by lipases immobilized in a packed bed reactor. *Biochem. Eng. J.* 46, 257–264. doi: 10.1016/J.BEJ.2009.05.015
- Hitch, T. C. A., and Clavel, T. (2019). A proposed update for the classification and description of bacterial lipolytic enzymes. *PeerJ* 7:e7249. doi: 10.7717/peerj.7249
- Hong, J. K., Choi, H. W., Hwang, I. S., Kim, D. S., Kim, N. H., Choi, D. S., et al. (2008). Function of a novel GDSL-type pepper lipase gene, CaGLIP1, in disease susceptibility and abiotic stress tolerance. *Planta* 227, 539–558. doi: 10.1007/s00425-007-0637-5
- Hu, P., Tom, L., Singh, A., Thomas, B. C., Baker, B. J., Piceno, Y. M., et al. (2016). Genome-resolved metagenomic analysis reveals roles for candidate phyla and other microbial community members in biogeochemical transformations in oil reservoirs. *MBio* 7, e1669–e1615. doi: 10.1128/mBio.01669-15
- Hu, X., Thumarat, U., Zhang, X., Tang, M., and Kawai, F. (2010). Diversity of polyester-degrading bacteria in compost and molecular analysis of a thermoactive esterase from *Thermobifida alba* AHK119. *Appl. Microbiol. Biotechnol.* 87, 771–779. doi: 10.1007/s00253-010-2555-x
- Hu, Y., Fu, C., Huang, Y., Yin, Y., Cheng, G., Lei, F., et al. (2010). Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS Microbiol. Ecol.* 72, 228–237. doi: 10.1111/j.1574-6941.2010.00851.x
- Hubbe, M. A., Nazhad, M., and Sánchez, C. (2010). Composting as a way to convert cellulosic biomass and organic waste into high-value soil amendments: a review. *BioResources* 5, 2808–2854. doi: 10.15376/biores.5.4.2808-2854
- Hugenholtz, P., and Tyson, G. W. (2008). Microbiology: Metagenomics. *Nature* 455, 481–483.
- Jayanath, G., Mohandas, S. P., Kachiprath, B., Solomon, S., Sajeevan, T. P., Bright Singh, I. S., et al. (2018). A novel solvent tolerant esterase of GDSSG motif subfamily from solar saltern through metagenomic approach: Recombinant expression and characterization. *Int. J. Biol. Macromol.* 119, 393–401. doi: 10.1016/j.ijbiomac.2018.06.057
- Jeon, J. H., Kim, J. T., Kim, Y. J., Kim, H. K., Lee, H. S., Kang, S. G., et al. (2009). Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl. Microbiol. Biotechnol.* 81, 865–874. doi: 10.1007/s00253-008-1656-2
- Jeon, J. H., Kim, S.-J., Lee, H. S., Cha, S.-S., Lee, J. H., Yoon, S.-H., et al. (2011a). Novel metagenome-derived carboxylesterase that hydrolyzes β -lactam antibiotics. *Appl. Environ. Microbiol.* 77, 7830–7836. doi: 10.1128/AEM.05363-11
- Jeon, J. H., Kim, J. T., Lee, H. S., Kim, S. J., Kang, S. G., and Choi, S. H. (2011b). Novel lipolytic enzymes identified from metagenomic library of deep-sea sediment. *Evid. Based Compl. Alternat. Med.* 2011:271419. doi: 10.1155/2011/271419
- Jia, M. L., Zhong, X. L., Lin, Z. W., Dong, B. X., and Li, G. (2019). Expression and characterization of an esterase belonging to a new family via isolation from a metagenomic library of paper mill sludge. *Int. J. Biol. Macromol.* 126, 1192–1200. doi: 10.1016/j.ijbiomac.2019.01.025
- Jiang, X., Huo, Y., Cheng, H., Zhang, X., Zhu, X., and Wu, M. (2012). Cloning, expression and characterization of a halotolerant esterase from a marine bacterium *Pelagibacterium halotolerans* B2T. *Extremophiles* 16, 427–435. doi: 10.1007/s00792-012-0442-3
- Jiménez, D. J., Montaña, J. S., Álvarez, D., and Baena, S. (2012). A novel cold active esterase derived from Colombian high Andean forest soil metagenome. *World J. Microbiol. Biotechnol.* 28, 361–370. doi: 10.1007/s11274-011-0828-x
- Kaminski, J., Gibson, M. K., Franzosa, E. A., Segata, N., Dantas, G., and Huttenhower, C. (2015). High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput. Biol.* 11:e1004557. doi: 10.1371/journal.pcbi.1004557
- Kang, C. H., Oh, K. H., Lee, M. H., Oh, T. K., Kim, B. H., and Yoon, J. H. (2011). A novel family VII esterase with industrial potential from compost metagenomic library. *Microb. Cell Fact.* 10:41. doi: 10.1186/1475-2859-10-41
- Keegan, K. P., Glass, E. M., and Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods Mol. Biol.* 1399, 207–233. doi: 10.1007/978-1-4939-3369-3_13
- Kim, Y. H., Kwon, E. J., Kim, S. K., Jeong, Y. S., Kim, J., Yun, H. D., et al. (2010). Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. *Biochem. Biophys. Res. Commun.* 393, 45–49. doi: 10.1016/j.bbrc.2010.01.070
- Knudsen, M., and Wiuf, C. (2010). The CATH database. *Hum. Gen.* 4, 207–212. doi: 10.1186/1479-7364-4-3-207
- Ko, K. C., Rim, S. O., Han, Y., Shin, B. S., Kim, G. J., Choi, J. H., et al. (2012). Identification and characterization of a novel cold-adapted esterase from a metagenomic library of mountain soil. *J. Ind. Microbiol. Biotechnol.* 39, 681–689. doi: 10.1007/s10295-011-1080-y
- Kovacic, F., Babic, N., Krauss, U., and Jaeger, K. (2019). “Classification of Lipolytic Enzymes from Bacteria,” in *Aerobic Utilization of Hydrocarbons, Oils and Lipids. Handbook of Hydrocarbon and Lipid Microbiology*, ed. F. Rojo (Cham: Springer), 1–35. doi: 10.1007/978-3-319-50418-6
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., and Charles, T. C. (2015). Current and future resources for functional metagenomics. *Front. Microbiol.* 6:1196. doi: 10.3389/fmicb.2015.01196
- Lämmle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., et al. (2007). Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J. Biotechnol.* 127, 575–592. doi: 10.1016/j.jbiotec.2006.07.036

- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lapébie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019). Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* 10, 1–7. doi: 10.1038/s41467-019-10068-5
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lee, S. W., Won, K., Lim, H. K., Kim, J. C., Choi, G. J., and Cho, K. Y. (2004). Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl. Microbiol. Biotechnol.* 65, 720–726. doi: 10.1007/s00253-004-1722-3
- Lee, Y.-S. (2016). Isolation and characterization of a novel cold-adapted esterase, MtEst45, from *Microbulbifer thermotolerans* DAU221. *Front. Microbiol.* 7:218. doi: 10.3389/fmicb.2016.00218
- Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V. T. T., Lauinger, B., et al. (2015). Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6:275. doi: 10.3389/fmicb.2015.00275
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., and Chatonnet, A. (2013). ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* 41, 423–429. doi: 10.1093/nar/gks1154
- Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et al. (2016). Evolution and ecology of Actinobacteria and their bioenergy applications. *Annu. Rev. Microbiol.* 70, 235–254. doi: 10.1146/annurev-micro-102215-095748
- Li, H., Han, X., Qiu, W., Xu, D., Wang, Y., Yu, M., et al. (2019). Identification and expression analysis of the GDSL esterase/lipase family genes, and the characterization of SaGLIP8 in *Sedum alfredii* Hance under cadmium stress. *PeerJ* 7:e6741. doi: 10.7717/peerj.6741
- Li, L. L., McCorkle, S. R., Monchy, S., Taghavi, S., and van der Lelie, D. (2009). Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol. Biofuels* 2, 1–11. doi: 10.1186/1754-6834-2-10
- Li, M., Yang, L. R., Xu, G., and Wu, J. P. (2016). Cloning and characterization of a novel lipase from *Stenotrophomonas maltophilia* GS11: The first member of a new bacterial lipase family XVI. *J. Biotechnol.* 228, 30–36. doi: 10.1016/j.jbiotec.2016.04.034
- Li, P. Y., Yao, Q. Q., Wang, P., Zhang, Y., Li, Y., Zhang, Y. Q., et al. (2017). A novel subfamily esterase with a homoserine transacetylase-like fold but no transferase activity. *Appl. Environ. Microbiol.* 83, e131–e117. doi: 10.1128/AEM.00131-17
- Liaw, R.-B., Cheng, M.-P., Wu, M.-C., and Lee, C.-Y. (2010). Use of metagenomic approaches to isolate lipolytic genes from activated sludge. *Bioresour. Technol.* 101, 8323–8329. doi: 10.1016/j.biortech.2010.05.091
- Lima-Junior, J. D., Viana-Niero, C., Conde Oliveira, D. V., Machado, G. E., Rabello, M. C., da, S., et al. (2016). Characterization of mycobacteria and mycobacteriophages isolated from compost at the São Paulo Zoo Park Foundation in Brazil and creation of the new mycobacteriophage Cluster U. *BMC Microbiol.* 16:111. doi: 10.1186/s12866-016-0734-3
- Lindemann, S. R., Moran, J. J., Stegen, J. C., Renslow, R. S., Hutchison, J. R., Cole, J. K., et al. (2013). The epsomitic phototrophic microbial mat of Hot Lake, Washington: Community structural responses to seasonal cycling. *Front. Microbiol.* 4:323. doi: 10.3389/fmicb.2013.00323
- Liu, K., Wang, J., Bu, D., Zhao, S., McSweeney, C., Yu, P., et al. (2009). Isolation and biochemical characterization of two lipases from a metagenomic library of China Holstein cow rumen. *Biochem. Biophys. Res. Commun.* 385, 605–611. doi: 10.1016/j.bbrc.2009.05.110
- Liu, S. P., Liu, R. X., Zhang, L., and Shi, G. Y. (2015b). Sequence-based screening and characterization of cytosolic mandelate oxidase using oxygen as electron acceptor. *Enzyme Microb. Technol.* 69, 24–30. doi: 10.1016/j.enzmictec.2014.11.001
- Liu, S. P., Liu, R. X., Zhang, L., and Shi, G. Y. (2015c). Sequence-based screening and characterization of cytosolic mandelate oxidase using oxygen as electron acceptor. *Enzyme Microb. Technol.* 69, 24–30. doi: 10.1016/j.enzmictec.2014.11.001
- Liu, J. F., Sun, X. B., Yang, G. C., Mbadanga, S. M., Gu, J. D., and Mu, B. Z. (2015a). Analysis of microbial communities in the oil reservoir subjected to CO₂-flooding by using functional genes as molecular biomarkers for microbial CO₂ sequestration. *Front. Microbiol.* 6:236. doi: 10.3389/fmicb.2015.00236
- Liu, Y. F., Galzerani, D. D., Mbadanga, S. M., Zaramela, L. S., Gu, J. D., Mu, B. Z., et al. (2018). Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome* 6:5. doi: 10.1186/s40168-017-0392-1
- Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D., and Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* 35:e120. doi: 10.1093/nar/gkm541
- López-López, O., Knapik, K., Cerdán, M. E., and González-Siso, M. I. (2015). Metagenomics of an alkaline hot spring in Galicia (Spain): Microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* 6:1291. doi: 10.3389/fmicb.2015.01291
- Lu, M., Dukunde, A., and Daniel, R. (2019). Biochemical profiles of two thermostable and organic solvent-tolerant esterases derived from a compost metagenome. *Appl. Microbiol. Biotechnol.* 103, 3421–3437. doi: 10.1007/s00253-019-09695-1
- Lu, M., Schneider, D., and Daniel, R. (2021). Metagenomic screening for lipolytic genes reveals an ecology-clustered distribution pattern. *bioRxiv* 2021, 429111. doi: 10.1101/2021.02.01.429111
- Ma, S., Fang, C., Sun, X., Han, L., He, X., and Huang, G. (2018). Bacterial community succession during pig manure and wheat straw aerobic composting covered with a semi-permeable membrane under slight positive pressure. *Bioresour. Technol.* 259, 221–227. doi: 10.1016/j.biortech.2018.03.054
- Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., et al. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5859–5864. doi: 10.1073/pnas.0901529106
- Maimanakis, J., Chow, J., Gaßmeyer, S. K., Güllert, S., Busch, F., Kourist, R., et al. (2016). Sequence-based screening for rare enzymes: new insights into the world of AMDases reveal a conserved motif and 58 novel enzymes clustering in eight distinct families. *Front. Microbiol.* 7:1332. doi: 10.3389/fmicb.2016.01332
- Mander, P., Yoo, H. Y., Kim, S. W., Choi, Y. H., Cho, S. S., and Yoo, J. C. (2014). Transesterification of waste cooking oil by an organic solvent-tolerant alkaline lipase from *Streptomyces* sp. CS273. *Appl. Biochem. Biotechnol.* 172, 1377–1389. doi: 10.1007/s12010-013-0610-7
- Martins, L. F., Antunes, L. P., Pascon, R. C., de Oliveira, J. C. F., Digiampietri, L. A., Barbosa, D., et al. (2013). Metagenomic analysis of a tropical composting operation at the São Paulo zoo park reveals diversity of biomass degradation functions and organisms. *PLoS One* 8:e61928. doi: 10.1371/journal.pone.0061928
- Masuch, T., Kusnezowa, A., Nilewski, S., Bautista, J. T., Kourist, R., and Leichert, L. I. (2015). A combined bioinformatics and functional metagenomics approach to discovering lipolytic biocatalysts. *Front. Microbiol.* 6:1110. doi: 10.3389/fmicb.2015.01110
- Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., and Nakajima-Kambe, T. (2008). Identification and characterization of novel poly(dl-lactic acid) depolymerases from metagenome. *Appl. Microbiol. Biotechnol.* 79, 743–750. doi: 10.1007/s00253-008-1477-3
- Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:1257. doi: 10.1038/ncomms11257
- Moore, A. M., Munck, C., Sommer, M. O. A., and Dantas, G. (2011). Functional metagenomic investigations of the human intestinal microbiota. *Front. Microbiol.* 2:188. doi: 10.3389/fmicb.2011.00188
- Nacke, H., Will, C., Herzog, S., Nowka, B., Engelhaupt, M., and Daniel, R. (2011). Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol. Ecol.* 78, 188–201. doi: 10.1111/j.1574-6941.2011.01088.x
- Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40:e155. doi: 10.1093/nar/gks678
- Nardini, M., and Dijkstra, B. W. (1999). α/β Hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.* 9, 732–737. doi: 10.1016/S0959-440X(99)00037-8
- Neher, D. A., Weicht, T. R., Bates, S. T., Leff, J. W., and Fierer, N. (2013). Changes in bacterial and fungal communities across compost recipes, preparation methods, and composting times. *PLoS One* 8:e79512. doi: 10.1371/journal.pone.0079512

- Ngara, T. R., and Zhang, H. (2018). Recent advances in function-based metagenomic screening. *Gen. Prot. Bioinform.* 16, 405–415. doi: 10.1016/j.gpb.2018.01.002
- Niu, Q., Qiao, W., Qiang, H., and Li, Y. Y. (2013). Microbial community shifts and biogas conversion computation during steady, inhibited and recovered stages of thermophilic methane fermentation on chicken manure with a wide variation of ammonia. *Bioresour. Technol.* 146, 223–233. doi: 10.1016/j.biortech.2013.07.038
- Okano, H., Hong, X., Kanaya, E., Angkawidjaja, C., and Kanaya, S. (2015). Structural and biochemical characterization of a metagenome-derived esterase with a long N-terminal extension. *Protein Sci.* 24, 93–104. doi: 10.1002/pro.2591
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2018). *vegan: Community Ecology Package. R Package Version 2.5-4*.
- Ouyang, L. M., Liu, J. Y., Qiao, M., and Xu, J. H. (2013). Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl. Biochem. Biotechnol.* 169, 15–28. doi: 10.1007/s12010-012-9949-4
- Pehrsson, E. C., Tsukayama, P., Patel, S., Mejía-Bautista, M., Sosa-Soto, G., Navarrete, K. M., et al. (2016). Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533, 212–216. doi: 10.1038/nature17672
- Peng, Q., Zhang, X., Shang, M., Wang, X., Wang, G., Li, B., et al. (2011). A novel esterase gene cloned from a metagenomic library from neritic sediments of the South China Sea. *Microb. Cell Fact.* 10:95. doi: 10.1186/1475-2859-10-95
- Ploner, A. (2015). *Heatplus: Heatmaps with Row and/or Column Covariates and Colored Clusters. R Package Version 2.26.0*. Available online at: <https://github.com/alexploner/Heatplus>
- Popovic, A., Hai, T., Tchigvintsev, A., Hajjigahsemi, M., Nocek, B., Khusnutdinova, A. N., et al. (2017). Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep.* 7:44103. doi: 10.1038/srep44103
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi: 10.1038/nbt.3935
- Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., et al. (2013). Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol.* 79, 4551–4563. doi: 10.1128/AEM.01077-13
- Rahman, M. A., Culsum, U., Tang, W., Zhang, S. W., Wu, G., and Liu, Z. (2016). Characterization of a novel cold active and salt tolerant esterase from *Zunongwangia profunda*. *Enzyme Microb. Technol.* 85, 1–11. doi: 10.1016/j.enzmictec.2015.12.013
- Ramnath, L., Sithole, B., and Govinden, R. (2016). Classification of lipolytic enzymes and their biotechnological applications in the pulping industry. *Can. J. Microbiol.* 63, 179–192. doi: 10.1139/cjm-2016-0447
- Ranjan, R., Yadav, M. K., Suneja, G., and Sharma, R. (2018). Discovery of a diverse set of esterases from hot spring microbial mat and sea sediment metagenomes. *Int. J. Biol. Macromol.* 119, 572–581. doi: 10.1016/j.ijbiomac.2018.07.170
- Rashmuse, K., Magomani, V., Ronneburg, T., and Brady, D. (2009). A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous β -lactamase activity on nitrocefin. *Appl. Microbiol. Biotechnol.* 83, 491–500. doi: 10.1007/s00253-009-1895-x
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reyes, A., Alves, J. M. P., Durham, A. M., and Gruber, A. (2017). Use of profile hidden Markov models in viral discovery: current insights. *Adv. Genomics Genet.* 7, 29–45. doi: 10.2147/AGG.S136574
- Reyes-Duarte, D., Ferrer, M., and García-Arellano, H. (2012). Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods Mol. Biol.* 861, 101–113. doi: 10.1007/978-1-61779-600-5_6
- Rime, T., Hartmann, M., and Frey, B. (2016). Potential sources of microbial colonizers in an initial soil ecosystem after retreat of an alpine glacier. *ISME J.* 10, 1625–1641. doi: 10.1038/ismej.2015.238
- Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., et al. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 7:14. doi: 10.3390/microorganisms7010014
- Romdhane, I. B.-B., Fendri, A., Gargouri, Y., Gargouri, A., and Belghith, H. (2010). A novel thermoactive and alkaline lipase from *Talaromyces thermophilus* fungus for use in laundry detergents. *Biochem. Eng. J.* 53, 112–120. doi: 10.1016/j.BEJ.2010.10.002
- Rudek, W., and Haque, R. U. (1976). Extracellular enzymes of the genus *Bacteroides*. *J. Clin. Microbiol.* 4, 458–460. doi: 10.1128/jcm.4.5.458-460.1976
- Ryckeboer, J., Mergaert, J., Vaes, K., Klammer, S., Clercq, D., Coosemans, J., et al. (2003). A survey of bacteria and fungi occurring during composting and self-heating processes. *Ann. Microbiol.* 53, 349–410.
- Sarmah, N., Revathi, D., Sheelu, G., Yamuna Rani, K., Sridhar, S., Mehtab, V., et al. (2018). Recent advances on sources and industrial applications of lipases. *Biotechnol. Prog.* 34, 5–28. doi: 10.1002/btpr.2581
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., and Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Res.* 47, 94–99. doi: 10.1093/nar/gky989
- Schloss, P. D., Hay, A. G., Wilson, D. B., and Walker, L. P. (2003). Tracking temporal changes of bacterial community fingerprints during the initial stages of composting. *FEMS Microbiol. Ecol.* 46, 1–9. doi: 10.1016/S0168-6496(03)00153-3
- Schneider, D., Engelhaupt, M., Allen, K., Kurniawan, S., Krashevska, V., Heinemann, M., et al. (2015). Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in sumatra (Indonesia). *Front. Microbiol.* 6:1339. doi: 10.3389/fmicb.2015.01339
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Seo, S., Lee, Y. S., Yoon, S. H., Kim, S. J., Cho, J. Y., Hahn, B. S., et al. (2014). Characterization of a novel cold-active esterase isolated from swamp sediment metagenome. *World J. Microbiol. Biotechnol.* 30, 879–886. doi: 10.1007/s11274-013-1496-9
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shao, H., Xu, L., and Yan, Y. (2013). Isolation and characterization of a thermostable esterase from a metagenomic library. *J. Ind. Microbiol. Biotechnol.* 40, 1211–1222. doi: 10.1007/s10295-013-1317-z
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Silva, C. C., Hayden, H., Sawbridge, T., Mele, P., Kruger, R. H., Rodrigues, M. V. N., et al. (2012). Phylogenetic and functional diversity of metagenomic libraries of phenol degrading sludge from petroleum refinery wastewater treatment system. *AMB Exp.* 2:18. doi: 10.1186/2191-0855-2-18
- Simon, C., and Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* 85, 265–276. doi: 10.1007/s00253-009-2233-z
- Simon, C., and Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Appl. Environ. Microbiol.* 77, 1153–1161. doi: 10.1128/AEM.02345-10
- Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., and DeRisi, J. L. (2014). Profile Hidden Markov Models for the detection of viruses within metagenomic sequence data. *PLoS One* 9:e105067. doi: 10.1371/journal.pone.0105067
- Snellman, E. A., and Colwell, R. R. (2004). Acinetobacter lipases: Molecular biology, biochemical properties and biotechnological potential. *J. Ind. Microbiol. Biotechnol.* 31, 391–400. doi: 10.1007/s10295-004-0167-0
- Sørensen, K. B., Canfield, D. E., Teske, A. P., and Oren, A. (2005). Community composition of a hypersaline endoevaporitic microbial mat. *Appl. Environ. Microbiol.* 71, 7352–7365. doi: 10.1128/AEM.71.11.7352-7365.2005
- Sriyapai, P., Kawai, F., Siripoke, S., Chansiri, K., and Sriyapai, T. (2015). Cloning, expression and characterization of a thermostable esterase HydS14 from *Actinomadura* sp. strain S14 in *Pichia pastoris*. *Int. J. Mol. Sci.* 16, 13579–13594. doi: 10.3390/ijms160613579
- Staden, R., Judge, D. P., and Bonfield, J. K. (2003). “Managing sequencing projects in the GAP4 environment,” in *Introduction to bioinformatics: a theoretical and practical approach*, eds D. D. Womble and S. A. Krawetz (New York, NY: Humana Press), 327–344. doi: 10.1385/1-59259-335-6:327
- Stroobants, A., Martin, R., Roosens, L., Portetelle, D., and Vandenberg, M. (2015). New lipolytic enzymes identified by screening two metagenomic libraries

- derived from the soil of a winter wheat field. *Biotechnol. Agron. Soc. Environ.* 19, 125–131. doi: 10.1111/jam.12597
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., et al. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi: 10.1101/gr.108332.110
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Turnbaugh, P. J., Hamady, M., Yatsunencko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi: 10.1038/nature07540
- Verma, S., Kumar, R., and Meghwanishi, G. K. (2019). Identification of new members of alkaliphilic lipases in archaea and metagenome database using reconstruction of ancestral sequences. *3 Biotech* 9, 1–8. doi: 10.1007/s13205-019-1693-9
- Vigneron, A., Alsop, E. B., Lomans, B. P., Kyrpides, N. C., Head, I. M., and Tesmetzis, N. (2017). Succession in the petroleum reservoir microbiome through an oil field production lifecycle. *ISME J.* 11, 2141–2154. doi: 10.1038/ismej.2017.78
- Villamizar, G. A. C., Nacke, H., and Daniel, R. (2017). Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity. *Methods Mol. Biol.* 1539, 249–260. doi: 10.1007/978-1-4939-6691-2_16
- Vital, M., Howe, A. C., and Tiedje, J. M. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* 5, e889–e814. doi: 10.1128/mBio.00889-14
- Walsh, C. J., Guinane, C. M., O’ Toole, P. W., and Cotter, P. D. (2017). A profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome. *PeerJ* 5:e3254. doi: 10.7717/peerj.3254
- Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., et al. (2016). Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9:22. doi: 10.1186/s13068-016-0440-2
- Wang, S., Wang, K., Li, L., and Liu, Y. (2013). Isolation and characterization of a novel organic solvent-tolerant and halotolerant esterase from a soil metagenomic library. *J. Mol. Catal. B: Enzym.* 95, 1–8. doi: 10.1016/j.molcatb.2013.05.015
- Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., and Zheng, S.-S. (2015). Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* 21, 803–814. doi: 10.3748/wjg.v21.i3.803
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4:e7401. doi: 10.1371/journal.pone.0007401
- Wei, T., Feng, S., Shen, Y., He, P., Ma, G., Yu, X., et al. (2013). Characterization of a novel thermophilic pyrethroid-hydrolyzing carboxylesterase from *Sulfolobus tokodaii* into a new family. *J. Mol. Catal. B: Enzym.* 97, 225–232. doi: 10.1016/j.molcatb.2013.07.022
- Wei, Y., Swenson, L., Castro, C., Derewenda, U., Minor, W., Arai, H., et al. (1998). Structure of a microbial homologue of mammalian platelet-activating factor acetylhydrolases: *Streptomyces exfoliatus* lipase at 1.9 Å resolution. *Structure* 6, 511–519. doi: 10.1016/S0969-2126(98)00052-5
- Wemheuer, F., Kaiser, K., Karlovsky, P., Daniel, R., Vidal, S., and Wemheuer, B. (2017). Bacterial endophyte communities of three agricultural important grass species differ in their response towards management regimes. *Sci. Rep.* 7:40914. doi: 10.1038/srep40914
- Wu, C., and Sun, B. (2009). Identification of novel esterase from metagenomic library of Yangtze River. *J. Microbiol. Biotechnol.* 19, 187–193. doi: 10.4014/jmb.0804.292
- Wu, G., Zhang, S., Zhang, H., Zhang, S., and Liu, Z. (2013). A novel esterase from a psychrotrophic bacterium *Psychrobacter celer* 3Pb1 showed cold-adaptation and salt-tolerance. *J. Mol. Catal. B: Enzym.* 98, 119–126. doi: 10.1016/j.molcatb.2013.10.012
- Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., and Bai, L. (2017). Discovery and characterization of a novel lipase with transesterification activity from hot spring metagenomic library. *Biotechnol. Rep.* 14, 27–33. doi: 10.1016/j.btre.2016.12.007
- Yu, H., Zeng, G., Huang, H., Xi, X., Wang, R., Huang, D., et al. (2007). Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation* 18, 793–802. doi: 10.1007/s10532-007-9108-8
- Yu, S., Yu, S., Han, W., Wang, H., Zheng, B., and Feng, Y. (2010). A novel thermophilic lipase from *Fervidobacterium nodosum* Rt17-B1 representing a new subfamily of bacterial lipases. *J. Mol. Catal. B: Enzym.* 66, 81–89. doi: 10.1016/j.molcatb.2010.03.007
- Yu, Z., Tang, J., Liao, H., Liu, X., Zhou, P., Chen, Z., et al. (2018). The distinctive microbial community improves composting efficiency in a full-scale hyperthermophilic composting plant. *Bioresour. Technol.* 265, 146–154. doi: 10.1016/j.biortech.2018.06.011
- Zarafeta, D., Moschidi, D., Ladoukakis, E., Gavrilov, S., Chrysina, E. D., Chatziioannou, A., et al. (2016). Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Sci. Rep.* 6:38886. doi: 10.1038/srep38886
- Zhang, X., Zhong, Y., Yang, S., Zhang, W., Xu, M., Ma, A., et al. (2014). Diversity and dynamics of the microbial community on decomposing wheat straw during mushroom compost production. *Bioresour. Technol.* 170, 183–195. doi: 10.1016/j.biortech.2014.07.093
- Zhang, Y., Hao, J., Zhang, Y. Q., Chen, X. L., Xie, B., Bin, et al. (2017). Identification and characterization of a novel salt-tolerant esterase from the deep-sea sediment of the South China Sea. *Front. Microbiol.* 8:441. doi: 10.3389/fmicb.2017.00441
- Zhang, Y., Pengjun, S., Wanli, L., Kun, M., Yingguo, B., Guozeng, W., et al. (2009). Lipase diversity in glacier soil based on analysis of metagenomic DNA fragments and cell culture. *J. Microbiol. Biotechnol.* 19, 888–897. doi: 10.4014/jmb.0812.695
- Zhou, G., Xu, X., Qiu, X., and Zhang, J. (2019). Biochar influences the succession of microbial communities and the metabolic functions during rice straw composting with pig manure. *Bioresour. Technol.* 272, 10–18. doi: 10.1016/j.biortech.2018.09.135
- Zhou, H., Gu, W., Sun, W., and Hay, A. G. (2018). A microbial community snapshot of windrows from a commercial composting facility. *Appl. Microbiol. Biotechnol.* 102, 8069–8077. doi: 10.1007/s00253-018-9201-4
- Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322. doi: 10.1128/aem.62.2.316-322.1996
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lu, Schneider and Daniel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.