A proposed update for the classification and description of bacterial lipolytic enzymes (#37278)

First submission

Guidance from your Editor

Please submit by **19 May 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the location described by the author.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

5 Table file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

📄 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

- Original primary research within [Scope of the journal](#).
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- ℹ️ Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	Example
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. ...* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

A proposed update for the classification and description of bacterial lipolytic enzymes

Thomas C A Hitch [Corresp., 1] , **Thomas Clavel** [1]

[1] Institute of Medical Microbiology, University hospital RWTH Aachen, Aachen, Germany

Corresponding Author: Thomas C A Hitch
Email address: thitch@ukaachen.de

Bacterial lipolytic enzymes represent an important class of proteins: they provide their host species with access to additional resources and have multiple applications within the biotechnology sector. Since the formalisation of lipolytic enzymes into families and subfamilies, advances in molecular biology have led to the discovery of lipolytic enzymes unable to be classified via the existing system. Utilising sequence-based comparison methods, we have integrated these novel families within the classification system proposed by Arpigny and Jaeger (1999) and recently updated by Kovacic *et al* (2019) so that it now consists of 35 families and 11 true lipase subfamilies. Representative sequences for each family and subfamily have been defined as well as methodology for accurate comparison of novel sequences against the reference proteins, facilitating the future assignment of novel proteins. Both the code and protein sequences required for integration of additional families are available at:
https://github.com/thh32/Lipase_reclassification

1 A proposed update for the classification and
2 description of bacterial lipolytic enzymes

3 Thomas C. A. Hitch[1,*] and Thomas Clavel[1]

4 [1]Functional Microbiome Research Group, Institute of Medical Microbiology, University hospital of RWTH Aachen,
5   Germany.

6 *To whom correspondence should be addressed.

7 Contact: thitch@ukaachen.de

8
9

10 Abstract

11
12   Bacterial lipolytic enzymes represent an important class of proteins: they provide
13 their host species with access to additional resources and have multiple applications
14 within the biotechnology sector. Since the formalisation of lipolytic enzymes into families
15 and subfamilies, advances in molecular biology have led to the discovery of lipolytic
16 enzymes unable to be classified via the existing system. Utilising sequence-based
17 comparison methods, we have integrated these novel families within the classification
18 system proposed by Arpigny and Jaeger (1999) and recently updated by Kovacic *et al*
19 (2019) so that it now consists of 35 families and 11 true lipase subfamilies.
20 Representative sequences for each family and subfamily have been defined as well as
21 methodology for accurate comparison of novel sequences against the reference
22 proteins, facilitating the future assignment of novel proteins. Both the code and protein
23 sequences required for integration of additional families are available at:
24 https://github.com/thh32/Lipase_reclassification
25
26

27 1. Introduction

28
29   Bacterial lipolytic enzymes are involved in a diverse range of microbial functions
30 from metabolism of fat for energy (1) to pathogenesis (2). These enzymes are also of
31 interest to the biotechnology industry for production of biodiesel and biopolymers (3).
32 The alpha-beta hydrolase family contains both lipases (Enzyme commission (EC):
33 3.1.1.3) and carboxylesterases (EC: 3.1.1.1), both of which can catalyse the hydrolysis
34 of 4-nitrophenyl-bound fatty acids (4,5). Lipases are defined as those which display
35 some ability to catalyse the release of free fatty acids from long-chain triacylglycerols,
36 such as those found in plant oil and animal fat, however activity may be highest against
37 short or medium chain fatty acids (3,6,7). This separates the lipases from the esterases,
38 which may have activity against p-nitrophenyl (pNP)-bound fatty acids (a standard test
39 for substrate specificity) but are unable to act upon triacylglycerol molecules (8,9). A
40 further subset of lipases are the true lipases which are defined as lipases which display
41 preferential activity against long-chain fatty acids (>C10) (10). The gold standard
42 classification system of bacterial lipolytic enzymes was originally published by Arpigny

43 and Jaeger in 1999 and defined eight families based on the biochemical properties and
44 sequence similarity of 53 proteins (11).
45
46   Under the Arpigny and Jaeger (1999) classification system, true lipases were
47 placed within a singular family (family I), within which sub-families are defined by unique
48 lipase groups (3,11). Originally, six sub-families of true lipases were described based on
49 shared features such as source genus, requirement of a chaperone for folding, motifs
50 and sequence similarity (11). Jaeger and Eggert (2002) redefined these families by
51 splitting family I.5 into two new families, family I.5 and family I.6, and moving the
52 previous members of family I.6 into family I.7 (3). In the 20 years since this bacterial
53 lipolytic enzyme classification system was published, more than twice as many 'novel'
54 families have been proposed (5,6,20–24,12–19). A recent update to this system
55 expanded the system to include some (11 of 30) of the recently published novel lipases
56 within 19 families and 8 true lipase sub-families, however many lipolytic proteins remain
57 unclassified (25).
58
59   Many claims of identifying novel families of lipolytic enzymes have been made,
60 however no exhaustive comparison has been conducted to verify and combine these
61 results. This has led to an inconsistent characterisation system and multiple publications
62 claiming discovery of the same family (24). Examples of this are family X (6,20) and
63 family XV (21,22), both of which have been coined twice. Rao *et al* (2013) attempted to
64 incorporate previously published families into their analysis, however their literature
65 search did not include multiple families (15,17–20,26,27). Ferrer *et al* (2015) also
66 attempted to integrate additional families into the Arpigny and Jaeger (1999)
67 classification system. Both previous attempts at including additional families relied on
68 the authors claims of 'novelty' without reanalysis of these claims and, as discussed
69 above, incomplete literature searches mean additional 'novel' families have been
70 missed from the analysis. The possibility exists that multiple publications may describe
71 members of the same 'novel' family under different names, artificially inflating the
72 number of families. Therefore, before additional families can be incorporated into the
73 system and provided with a family number, previous claims of novelty must be
74 confirmed and integration of these proteins into a singular classification system must
75 occur.
76
77   Alternative classification systems for bacterial lipolytic enzymes have been
78 published alongside online databases dedicated to these new classification systems,
79 including MELDB (28) and LIPABASE (29). Both websites have since been deactivated
80 preventing their continued use. The simplicity of the Arpigny and Jeager (1999)
81 classification system is that by providing type proteins for each family, the system can
82 continue to be utilised by the research community. Additional databases such as the
83 Lipase Engineering Database (30) and ESTHER (31) classify similar groups of
84 enzymes, however these are not specific to bacterial lipolytic enzymes and as with
85 MELDB and LIPABASE, their continued availability is not guaranteed.
86
87   In this paper, the novelty of lipolytic enzymes was confirmed via pairwise
88 sequence comparison to the current type proteins. This allowed us to expand the

89 current classification system and remove spurious claims of novelty and provide a
90 consistent nomenclature to these proteins.
91

2. Methods

93

2.1. Literature search

95   Multiple methods were used to identify post-1999 claims of novel lipases and
96 lipase families. The key terms 'novel, bacterial, lipase, esterase, carboxylesterase,
97 family' were used to search both the Pubmed (www.ncbi.nlm.nih.gov/pubmed/) and
98 Google scholar (https://scholar.google.de/) databases. Each identified paper was
99 screened to confirm that: (i) the protein was of bacterial/metagenomic origin; (ii) the
100 lipolytic activity of the proposed sequence was confirmed experimentally; (iii) the studied
101 lipolytic protein was indeed novel or assigned to a previously defined family.

102

2.2. Sequence comparison

104   Type proteins for both the families described by Arpigny and Jaeger (1999) and
105 lipolytic proteins identified by the literature search were compared pairwise via ClustalW
106 (32). Sequence identity was calculated using a python implementation of the "Identity
107 and similarity" tool within the Sequence Manipulation Suite (33) and has been made
108 available as a reproducible Jupyter notebook (Pairwise_analysis.ipynb) within the
109 Github repository given in the abstract. Proteins were considered likely to be part of the
110 same family at 60% sequence identity, a threshold that is 3 standard deviations away
111 from the mean similarity between the type proteins of the original 8 families (30.9 ±
112 9.6% similarity). Proteins identified by sequence comparison to be representatives of
113 the same family had their biochemical properties compared to identify if these were
114 consistent or highly divergent. Transmembrane regions were detected within each
115 protein using TMHMM (34). Proteins were also annotated with an ESTHER family using
116 the online HMMscan tool, minimum bitscore of 50 (31), allowing for cross-classification
117 system comparison.
118
119

2.3. Taxonomic occurrence

121   To provide insight into the evolutionary diversity or specificity of each family, the
122 NCBI genome database (downloaded November 2017) (35) containing 72,462
123 genomes was annotated against the type proteins for each family using BLASTP (36).
124 Matches were separated into two categories: (i) strict matches required high-scoring-
125 pairs to share >80% identity across >80% of both the query and subject sequences,
126 providing an overall identity of >64%. Minimum identity of 60% was selected based on
127 previous research on functional annotation transfer (37); (ii) An expanded genome
128 search using reduced strictness (>60% identity across >80% of the query sequence,
129 resulting in a total identity of >48%) was conducted to provide insight into the potential
130 host range of metagenome-derived proteins with no close matches. The coverage and
131 identity used in the expanded genome search remain more stringent than those
132 previously identified to be optimum for protein family identification and clustering (38).
133 Once annotated, genomes were linked to their taxonomic lineage using the NCBI

134   taxonomy information, implemented in MGkit (39), and the Genome Taxonomy
135   database (GTDB), version 86 (40). GTDB annotations were used to confirm the
136   taxonomic assignment of genomes when the NCBI assignment was uncertain.
137
138   **2.4. Origin check of metagenomic enzymes**
139   Metagenomic approaches capture both eukaryotic and prokaryotic DNA. To
140   prevent the inclusion of eukaryotic enzymes into this system, enzymes captured by
141   metagenomic approaches were screened against the NCBI genome database. If a strict
142   match was identified, the bacterial origin of the enzyme was assured. For those
143   enzymes without a genome match, possibly due to the host species being as-yet
144   uncultured, the sequences were compared against the NR database using BLASTP.
145   The taxonomic assignment of the top 10 matches was examined and if all were
146   bacterial, the enzyme was included within the classification system, else they were
147   excluded.
148
149
150   **3. Results and discussion**
151   **3.1. Overview**
152   Pairwise sequence comparison of the accepted type proteins from Arpigny and
153   Jaeger (1999) to each proposed new family retrieved from the literature identified a
154   single match (>60% similarity), Supplementary Table S1. The HZ lipase proposed by
155   Masomian *et al* (2016) shared 69.5% similarity with family I.5 (U78785) and 57.0%
156   similarity with family I.6 (M1271) (9). Due to this, the HZ lipase was assigned as a
157   member of family I.5. Choi *et al* (2004) also proposed a novel lipase (EstI) that was
158   identified to share 62.5% similarity with family I.6 (M12715), which indicates it is a
159   member of this subfamily.
160
161   Pairwise comparison of the proposed families type proteins against one-another
162   was conducted to identify if any families have been discovered multiple times
163   (Supplementary Table S2). We identified that Est30 (13), EstUT1 (41) and EstOF4 (42)
164   share >60% similarity, forming a single family, as did EstGtA2 (43), bMGL (44) and LipS
165   (45).
166
167   As this paper aims to act as an expand upon the existing Arpigny and Jaeger
168   (1999) classification system which was updated very recently (25), all proposed families
169   were included within the existing framework. We remained consistent with the recent
170   update to the original classification system proposed by Kovacic *et al* (2019), however
171   we propose a few corrections. In the aforementioned update, LipG was identified as the
172   type protein for family XI, however its preference for long-chain fatty acids (C14-16) and
173   activity against triglycerides (tricaprylin) identify it as a true lipase (14). Secondly, EstD
174   and LipR were both placed together under family X, however these proteins share
175   <30.1% similarity and have different biochemical characteristics, suggesting they
176   represent separate families. This is supported by each proteins catalytic serine being
177   contained within different motifs. Due to this we propose dividing family X into family X
178   and family XXVIII according the numbering proposed in the present study. Finally, in the
179   updated classification system by Kovacic *et al*, family XVIII was based on EstUT1,

180 separate from Est30 and EstOF4 against which it shares however high (>60%)
181 similarity. Due to this we have collapsed family XVIII into family XIII. These changes
182 mean that family I to XIX remain consistent with the published updated classification
183 system, apart from family XI and XVIII. To prevent confusion regarding our changes up
184 to family XIX we have provided descriptions for each family starting at family IX (see
185 Table 1 and individual descriptions provided below).
186
187
188

3.2. Proposed family I subfamilies

189
190

3.2.1. Family I, subfamily 9
192 LipG, the type protein for family I.9, was isolated during a metagenomic study of
193 korean tidal flat sediment (14). The GxSxG motif (where x represents any amino acid)
194 was identified as GHSLG. Genome analysis was unable to find similar proteins within
195 the current genome database (Supplementary Table S3). However, sequence matches
196 within the NR database confirmed its bacterial origin (Supplementary Table S4).
197 Maximum activity was identified against C14 and C16, however activity decreased by
198 65% against C18. Both Ca^{2+} and Mn^{2+} were identified to increase the activity of LipG
199 and chelation of divalent metal ions via ethylenediaminetetraacetic acid (EDTA)
200 significantly reduced activity confirming the necessity of metal ions. The effect of
201 temperature and pH have yet to be reported for LipG. LipG was annotated as belonging
202 to 'Lipase_3' under the ESTHER system.
203

3.2.2. Family I, subfamily 10
205 RlipE1 was isolated from the rumen of Chinese Holstein cattle using a
206 metagenomic bacterial artificial chromosome (BAC) library (46). Whilst no genomes
207 contained similar proteins in our analysis, the closest matches within the NR database
208 were bacterial in origin, indicating that RlipE1 originated from bacteria. The catalytic
209 Serine was identified within GLSMG. Activity against triglycerides along with maximum
210 activity against C16 confirmed that RlipE1 is a true lipase. Optimum conditions were
211 identified as 30°C and pH 7-8. The effect of metal ions has yet to be examined. RlipE1
212 was annotated as 'A85-Feruloyl-Esterase' under the ESTHER system.
213

3.2.3. Family I, subfamily 11
215 MPlaG is a phospholipase isolated from tidal flat sediment in korea (47). The
216 catalytic serine is present within the pentapeptide GHSLG. Activity against triglycerides
217 and a preference for long-chain fatty acids identified MPlaG as a true lipase. Optimum
218 conditions were determined as pH 8 and 25°C. Addition of Ca^{2+} increased activity 10-
219 fold whilst Cu^{2+}, Zn^{2+}, and Ni^{2+} all decreased activity. EDTA (5mM) removed all activity,
220 suggesting Ca^{2+} is essential for activity. Expanded genomic analysis identified possible
221 members of this enzyme family within the *Oceanospirillaceae*. The source of isolation of
222 MPlaG along with its potential host identified by genome analysis suggest that family
223 I.11 may be specific to the marine environment. MPlaG had no significant match to a
224 ESTHER family.
225

3.3. Proposed families

3.3.1. Family IX

Family IX was characterised in 2001 with the type protein phaZ7 (36 kDa), isolated from *Paucimonas lemoignei* (DSMZ7445) (12). phaZ7 appears to be host-specific as it was not detected within any deposited genome, including that of *Paucimonas lemoignei* NCTC10937. phaZ7 was identified as a carboxyesterase specific for the degradation of poly((R)-3-hydroxybutyrate) (PHB). However, no sequence similarity was noted between it any known PHB depolymerase. The presence of the AHSMG motif within phaZ7 along with its optimum temperature of 65°C support its identification as a thermophilic estrase. Activity was also increased within alkaline environments, with the optimum being pH 10 and remaining stable for 6.5 hours in solutions up to pH 12. Addition of Ca^{2+} was identified to significantly increase the activity of phaZ7 and in combination with EDTA, which by itself reduced activity by 70%, activity returned to baseline levels. Whilst resistant to the effects of phenylmethylsulfonyl fluoride (PMSF), a serine hydrolase inhibitor, activity was significantly reduced by addition of another serine hydrolase inhibitor, diisopropyl fluorophosphates (DFP), confirming the identification of phaZ7 as a serine hydrolase. This family is represented in the ESTHER system as 'PHAZ7_phb_depolymerase'.

3.3.2. Family X

EstD (44.5 kDa) was isolated from *Thermotoga maritime* (6), a hyperthermophile isolated from geothermally heated marine sediment (48). EstD contains the GHSLG variant of the GxSxG motif. Activity was reduced by 96% when treated with PMSF. Both highly similar and reduced threshold matches to EstD were specific to the *Thermotoga* genus. EstD activity increased linearly with temperature to a maximum of 100°C at which it has a half life of 1 hour. The optimum pH of EstD was 7 with a maximum activity against C5 to C8. EstD activity is not significantly affected by the presence or absence of metal ions or by treatment with EDTA, suggesting metal ions may have a minor effect but are not essential for activity. EstD was annotated as 'Bacterial_EstLip_FamX' under the ESTHER system.

3.3.3. Family XI

The type protein for family XI, EstA (33 kDa) was identified via lipase screening of a bacterial artificial chromosome (BAC) library created using DNA fragments from the South China Sea (26). Analysis of EstA and its closest relatives identified the GxSxG variant, GHSMG, within all members. PMSF reduced activity of EstA by 97%. Strict genome analysis was unable to identify any matching sequences but the expanded genome search identified matches were found within 16 taxonomic families and 41 genera. These results suggest that while the exact protein is likely to be present in a uncultured species, variants exist within a diverse range of hosts. EstA is specific for short- and medium-chain fatty acids with maximum activity against C6, decreasing till C10 past which no activity is detected. The optimum temperature and pH for EstA was determined to be 45°C and 6.5. The metal ions Mg^{2+}, Mn^{2+} and Co^{2+} increase EstA activity while Fe^{2+}, Fe^{3+} and Zn^{2+} reduced activity. The effect of EDTA was not tested.

271 EstA was annotated as belonging to the 'A85-EsteraseD-FGH' family within the
272 ESTHER classification system.

273

3.3.4. Family XII

275   Functional analysis of a fosmid library created using DNA from intertidal flats
276 which experience periodic changes in salinity, temperature and pH identified LipEH166
277 (42 kDa), the type protein for family XII (16). LipEH166 contains the catalytic serine
278 within GHSLG. Family XII is preferentially active against medium length fatty acids, C8
279 to C14, with maximum activity against C12. As the fosmid containing LipEH166
280 displayed lipolytic activity on a tributyrin agar plate however the purified proteins activity
281 was never tested against a long-chain triglyceride such as triolein, it has been
282 characterised as a non-true lipase. The optimum temperature and pH for LipEH166 are
283 30°C and 8-9 respectively. The presence of Ca^{2+}, Mg^{2+} and Mn^{2+} increased activity
284 whilst Cu^{2+} and Zn^{2+} were inhibitory and chelation with EDTA removed activity,
285 confirming LipEH166 as a metalloprotein. Genomic analysis assigned LipEH166 to the
286 *Marinobacter* which are known to reside in marine sediment and produce emulsifying
287 agents (49). Expanded genomic analysis identified similar enzymes within three
288 families, *Oceanospirillaceae*, *Alteromonadaceae* and *Oleiphilaceae* which are known to
289 exist in marine environments (49–51). A single transmembrane region was detected
290 from position 7-26. LipEH166 was annotated as belonging to the ESTHER family
291 'Bact_LipEH166_FamXII'

292

3.3.5. Family XIII

294   Ewis *et al* (2004) identified Est30 (30 kDa), the type protein for family XIII, within
295 *Geobacillus stearothermophilus* ATCC7954 (13). EstUT1, identified by Samoylova *et al*
296 (2018), was determined to be the second member of this family due to sharing 65%
297 sequence similarity and key biochemical characteristics (41). EstOF4 was also
298 identified to be a member of this family based on a protein similarity of 85.6% and
299 60.9% to Est30 and EstUT1 respectively (42). The GxSxG motif was identified within all
300 members as GxSLG and treatment with PMSF inhibited enzymatic activity. Optimum
301 conditions for family XIII are 50-80°C and pH 8-9. All members were determined to be
302 most active against short and medium-chain fatty acids and display no activity against
303 long-chain fatty acids. EDTA had no effect on either lipases, suggesting activity is
304 independent of metal ions. Strict genome analysis identified the presence of proteins
305 similar to Est30 within four genera (*Anoxybacillus, Bacillus, Geobacillus,*
306 *Parageobacillus*) belonging to the *Bacillaceae*, in which both EstOF4 and EstUT1 was
307 also identified. Expanded genome analysis identified matches in five taxonomic families
308 and 31 genera according to the NCBI taxonomy. This suggests that members of this
309 enzymatic family are present within a wide range of taxonomic groups. All three
310 members of this family were assigned to 'CarbLipBact_1' under the ESTHER
311 classification system.

312

3.2.6. Family XIV

314   EstA3 (50 kDa), the type protein for family XIV, was identified within
315 *Thermoanaerobacter tengcongensis*, a thermophilic bacteria isolated from a hot spring
316 in China (5). The catalytic serine is present within the CHSMG pentapeptide. Family XIV

317 was identified to be monophyletic, occurring only within genomes from the
318 *Thermoanaerobacteraceae*. Optimum activity was achieved at 70°C and pH 9.5.
319 Maximum activity was identified against C6, although high activity was also observed
320 against C4 suggesting preference for short-chain fatty acids. Family XIV is inhibited by
321 all tested metal ions (Mg^{2+}, Ca^{2+}, Sn^{2+}, Co^{2+}, Mn^{2+}, Ag^+, Ni^{2+}, Fe^{3+}, Pb^{2+}, Hg^{2+}, Zn^{2+},
322 Cu^{2+}) and not effected by EDTA suggesting metal ions are not required for activity.
323 Activity was also not effected by addition of PMSF due to the active serine being 'buried'
324 within the structure of EstA3. The weak annotation (bitscore of 54.3) of EstA3 to the
325 ESTHER family 'Lipase_2' suggests it represents a separate family, not currently
326 included within the ESTHER system.
327
328 **3.3.7. Family XV**
329   EstGtA2 (~30 kDa) was identified within the thermophile *Geobacillus*
330 *thermodenitrificans* (43). Both bMGL (44) and LipS (45) were identified as members of
331 this family via sequence similarity to EstGtA2 of 92.0% and 65.9% respectively.
332 Maximum activity was identified against C8-C12 and optimum conditions of 50-75°C
333 and pH 6-8.. Sequence analysis identified the catalytic serine within the GLSMG
334 pentapeptide for all members. Strict genome analysis identified EstGtA2 as
335 monophyletic to the *Bacillaceae*, occurring within the genera *Geobacillus* and
336 *Anoxybacillus*. This was identical to the results for bMGL. LipS was identified to have a
337 unique taxonomic distribution to both other members of this family as strict matches
338 were identified only within *Symbiobacterium* with the expanded search identifying
339 additional matches within *Rudaea.* All three members of this family were assigned to
340 'CarbLipBact_2' under the ESTHER classification system.
341
342 **3.2.8. Family XVI**
343   Mining the genome of *Stenotrophomonas maltophila* GS11 led to the
344 identification of LipSM54 (52.8 kDa), the type protein for family XVI (52). The catalytic
345 serine was identified within GISYG and PMSF reduced activity by 95%. A single
346 transmembrane helix was identified from position 5-24. Family XVI shows preferential
347 activity against medium length fatty acids with maximum activity against C8 and C10 in
348 both pNP ester and triacylglycerol forms, confirming LipSM54 as a true lipase. Optimum
349 activity was detected at 35 and pH 8. While Mn2+ slightly increased activity of
350 LipSM54, Li^+, Fe^{2+}, Hg^{2+}, Ni^{2+}, Zn^{2+}, Co^{2+} and Cu^{2+} inhibited activity by up to 80%.
351 Addition of EDTA reduced activity by up to 50%, however addition of Ca^{2+} and Mg^{2+}
352 reduced the effect of EDTA. Family XVI was identified to be monophyletic to the
353 *Xanthomonadaceae*, specifically the *Lysobacter* and *Stenotrophomonas* genera.
354 LipSM54 was annotated as a member of the ESTHER family Peptidase_S15.
355
356 **3.3.9. Family XVII**
357   Mining the genome of *Janibacter* sp. R02 which was isolated from the Antartic
358 soil led to the discovery of LipJ2 (44 kDa), the type protein for family XVII (23). Whilst
359 the strain was identified to hydrolyse olive oil, LipJ2 showed a preference towards short-
360 rather than long-chain fatty acids, prevented it being characterised as a true lipase.
361 Optimum conditions for LipJ2 were determined to be pH 7-8 and 80°C. While Na^{2+} and
362 K^+ did not significantly effect the activity of LipJ2 separately, together they

363 synergistically increased activity by 250% compared to the baseline. The catalytic
364 serine (Ser[219]) was identified within the pentapeptide GYSQG. An expanded search of
365 family XVII members identified matches within only the genus *Janibacter*. LipJ2 was
366 annotated as a member of 'Fungal-Bact_LIP' within the ESTHER system.
367
368 **3.2.10. Family XVIII**
369   PlaB, a phospholipase isolated from *Legionella pneumophila*, is the type protein
370 for family XVIII (15). Family XVIII contains its catalytic triad serine within the motif
371 THSTG, replacing the first glycine in the common GxSxG motif with threonine. Genome
372 analysis identified family XVIII as monophyletic to the *Legionellaceae*, specifically the
373 *Legionella* and *Fluoribacter* genera. Due to the clinical relevance of *Legionella*
374 *pneumophila,* research on PlaB has focused on its possible role as a virulence factor via
375 hydrolysis of phosphatidylcholine within eukaryotic membranes. PlaB was annotated to
376 its own ESTHER family called 'PlaB'.
377
378 **3.3.11. Family XIX**
379   Cultivation of lipase positive isolates from waste-treatment sludge, determined by
380 Rhodamine B-olive oil plates, identified *Stenotrophomonas maltophila* Psi-1 and the
381 identification of LipSm (40.7 kDa) (24). LipSm activity was tested against two pNP-
382 bound fatty acids (C4 and C12) with maximum activity against C12. Optimum activity
383 was achieved at 30-40°C and pH 7 and all tested metal ions (Ca^{2+}, Mg^{2+}, Mn^{2+}, Na^{2+})
384 inhibited activity. Sequence analysis identified the presence of the catalytic serine within
385 the pentapeptide GHSQG and similar proteins only in the *Stenotrophomonas*. The
386 ability of purified LipSm to release free fatty acids from olive oil was not tested,
387 preventing its classification as a true lipase. LipSm was annotated as belonging to the
388 'Fungal-Bact_LIP' family under the ESTHER classification system.
389
390
391 **3.3.12. Family XX**
392   Hu *et al* (2010) isolated two lipolytic enzymes sharing identity with a putative
393 poly(3-hydroxybutyrate) depolymerase, FLS18C and FLS18D (~32 kDa), from the South
394 China Sea forming the novel family XX (18). As biochemical characterisation has only
395 been conducted for FLS18D, it was selected as the type protein. The catalytic serine
396 was identified within the conserved GHSMG motif. Activity was highest against C4, with
397 decreasing activity until C10, suggesting preferential hydrolysis of short-chain fatty
398 acids. Optimum activity was achieved at 45°C at pH 8. The effect of metal ions on the
399 activity of FLS18D has not yet been studied. Genome analysis was unable to identify
400 similar proteins within any sequenced bacteria but bacterial origin was identified via
401 BLAST annotation against the NR database. Both proteins within this family were
402 annotated as members of the 'Abhydrolase_5' family within ESTHER.
403
404 **3.3.13. Family XXI**
405   Yu *et al* (2010) identified the type protein, FnL (~30 kDa), for family XXI within
406 *Fervidobacterium nodosum Rt17-B1*, an anaerobic thermophile with optimum growth at
407 80°C (8). Genome analysis identified FnL as specific to the *Fervidobacterium*. The
408 pentapeptide sequence AxSxG identified within FnL and its closest relatives in the

409 original analysis. Treatment with PMSF reduced activity to 5%, confirming FnL as a
410 serine hydrolase. FnL showed the highest activity towards medium length fatty acids,
411 with highest activity on C10. Whilst initially published as a subfamily of true lipases, its
412 preference to medium-chain fatty acids identifies it as a novel family of non-true lipases.
413 FnL showed the highest activity at 70°C and pH of 9.0. Activity was significantly affected
414 by metal ion concentrations with significantly reduced activity with Zn^{2+} and Cu^{2+} and
415 increased activity with Mg^{2+}, Ca^{2+}, Na^+, Ni^{2+} and Co^{2+}. Chelation using EDTA reduced
416 activity by only 6%, suggesting metal ions are not essential for FnL activity. FnL had no
417 significant match within the ESTHER classification system.

418
419 **3.3.14. Family XXII**
420   Functional analysis of a metagenomic library produced from the rhizosphere
421 microbiome of multiple plants identified EstD2 (53 kDa), the type protein for family XXII
422 (17). EstD2 contains the GxSxG motif in the form of GHSQG. Maximum activity was
423 identified against C4 and decreased as the fatty acid length increased. Optimum
424 conditions were determined to be 35°C and pH 7-9. EstD2 activity is highly sensitive to
425 metal ions, as inhibition occurred via addition of Mn^{2+}, Zn^{2+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+}, Ni^{2+}
426 and Rb^{2+} but not Ca^{2+} and Mg^{2+}. EDTA reduced activity by up to 68%, indicating that
427 metal ions may be required for activity of family XXII enzymes. No similar proteins were
428 identified within the genome database suggesting this family is present within currently
429 uncultured bacteria, however BLAST annotation against the NCBI-NR database
430 confirmed it as bacterial in origin. EstD2 was annotated as a member of the 'Duf_3089'
431 family within the ESTHER system.

432
433 **3.3.15. Family XXIII**
434   The type protein for family XXIII is LipA (32 kDa) and was identified within a
435 fosmid library created from Brazilian mangrove sediment (53). Genome analysis
436 identified no similarity to proteins from sequenced bacteria, however annotation against
437 the NR database confirmed it as bacterial in origin. LipA has the pentapeptide sequence
438 of AHSMG, fitting with the AxSxG motif. LipA was shown to have the highest activity on
439 C10 and optimum pH and temperature of 7.5-8.5 and 35°C respectively. LipA was
440 annotated as belonging to the ESTHER family 'Bacterial_lipase'.

441
442 **3.2.16. Family XXIV**
443   Rv0045c (35.5 kDa) is an esterase isolated from *Mycobacterium tuberculosis* (4).
444 The pentapeptide GMSLG contains the catalytic serine. Maximum activity was identified
445 against C6 and optimum conditions of 39°C and pH 8. Genome analysis identified this
446 family of lipolytic enzymes as specific to the *Mycobacterium*. Crystallography identified
447 that the active site of Rv0045c is highly conserved and consists of only a serine and
448 histidine, not the traditional catalytic triad (54). The reliance of Rv0045c on the active
449 serine has been confirmed by inactivation with PMSF (55). Rv0045c was annotated to
450 the ESTHER family 'Epoxide-hydrolase_like', however similar similarity was identified to
451 multiple other families, suggesting it may represent its own ESTHER family.

452
453 **3.3.17. Family XXV**

454   Analysis of a metagenomic library from a deep-sea sediment for lipolytic activity
455 identified Est3L4 (~37 kDa), the type protein for family XXV (19). Optimum conditions
456 for activity were identified as 35°C, pH 7.5 and 1.5M NaCl. The addition of NaCl
457 consistently improved activity, suggesting this family has adapted to the saline
458 environment of the deep-sea. Family XXV is affected by metal ions as Mn^{2+}, Mg^{2+}, Ca^{2+},
459 Cu^{2+} significantly increased activity. Reduced activity was identified with addition of Co^{2+}
460 and addition of Fe^{2+} inhibited all activity. EDTA was shown to have no effect, suggesting
461 that metal ions are not essential for activity. Addition of PMSF significantly reduced
462 activity. Est3L4 was identified as specific for long-chain fatty acids with maximum
463 activity against C16. No similar proteins were identified within the genome database. A
464 single transmembrane helix was detected at position 9-31. Est3L4 was annotated as a
465 member of the 'Esterase_phb' ESTHER family.
466

467 **3.3.18. Family XXVI**
468   A metagenomic library of German soil samples identified EstGS (40.7 kDa)
469 containing the GxSxG motif in the form of GHSFG (27). Maximum activity was identified
470 against C6, with high activity also against C4. The effect of factors such as temperature,
471 pH and metal ions were not studied. No matching proteins were identified via genome
472 analysis but bacterial origin was confirmed via annotation against the NCBI-NR
473 database. EstGS was identified as a member of the ESTHER family 'UCP031982'.
474

475 **3.3.19. Family XXVII**
476   Alongside the discovery of family XXVII, Nacke *et al* (2011) identified the type
477 protein for family XXVII, EstGH (45.6 kDa), via metagenomic analysis of German soil
478 (27). The catalytic serine of EstGH was contained within GHSLG and expanded
479 genome analysis identified similar proteins within the *Mycobacterium* which is known to
480 exist within soil samples (56). EstGH had no significant match to any ESTHER family,
481

482 **3.3.20. Family XXVIII**
483   The high lipolytic activity of *Rhodococcus* sp. strain CR-53, isolated from soil, led
484 to the discovery of LipR (43 kDa), the type protein for family XXVIII (20). LipR was
485 identified to be active against medium length fatty acids with maximum activity against
486 C10. High activity was determined on Rhodamine-triolein-supplemented agar plates.
487 Optimum conditions were determined as 40°C and pH 7. High concentrations of Mg^{2+},
488 Mn^{2+} and Ca^{2+} metal ions (10mM) were shown to increase activity whilst low
489 concentrations (1mM) of Fe^{2+}, Zn^{2+}, Cu^{2+} and Ag^+ significantly inhibited the lipolytic
490 activity of LipR. The catalytic serine (Ser^{212}) was present within the pentapeptide
491 GYSGG. Family XXVIII was identified to be monophyletic, occurring specifically within
492 members of the genus *Rhodococcus*. LipR was determined to contain a single
493 transmembrane helix from position 12-34. LipR was annotated as a member of the
494 ESTHER family 'Fungal-Bact_LIP'.
495

496 **3.3.21. Family XXIX**
497   EstGX1 (22.4 kDa) was isolated via metagenomic analysis of high Andean forest
498 soil from Columbia (57). The catalytic serine was located within the pentapeptide
499 GPSGG. Maximum activity was identified against C4 with optimum conditions of 40°C

500 and pH 8. Whilst no genomic match was identified, annotation against NR identified
501 EstGX1 as being bacterial in origin. EstGX1 had no match to any ESTHER family.

502

503 **3.2.22. Family XXX**
504   Metagenomic fosmid analysis of Chinese soil samples identified EstP2K (25 kDa)
505 (58). Sequence analysis determined the catalytic serine was located within AHSLG and
506 genomic analysis identified no matching sequences. Maximum activity was detected
507 against C8 with optimum conditions of 45-55°C and pH 7.5. Metal ions were identified to
508 have a large effect on the activity of EstP2K with Mg^{2+} doubling the activity whilst Co^{2+},
509 Cu^{2+}, Zn^{2+}, Ca^{2+}, Pb^{2+}, Ni^{2+}, Mn^{2+}, Hg^{2+}, Fe^{2+} and Ag^+ all decreased the activity.
510 Chelation using EDTA halved the enzymatic activity of EstP2K, suggesting Mg^{2+} is
511 required for activity. Activity was reduced to 1.5% with the addition of PMSF, confirming
512 EstP2K as a serine hydrolase. EstP2K had a weak match (bitscore = 70.3) to the
513 ESTHER family 'Lipase_2'.

514

515 **3.3.23. Family XXXI**
516   Genome mining of *Psychrobacter celer* 3Pb1, a known psychrophile, identified
517 the presence of Est12 (35kDa) (59). Sequence analysis identified the catalytic serine
518 within GHSAG and genome analysis determined this family as monophyletic to the
519 *Psychrobacter*. Maximum activity was identified against C4 with optimum conditions of
520 35°C and pH 7-8. Est12 was identified to retain activity when exposed to high
521 concentrations of NaCl (<4.5M) for 13 hours, indicating it is salt tolerant. EDTA was
522 shown to reduce activity by 15%, suggesting metal ions are beneficial for activity but
523 may not be essential. PMSF completely inactivated Est12 when added at 5mM. Est12
524 was annotated as a member of the 'Hormone-sensitive_lipase_like' family in the
525 ESTHER system.

526

527 **3.3.24. Family XXXII**
528   Est9x (32 kDa) was isolated from marine metagenomic library of the South China
529 sea (60). Sequence analysis identified the catalytic serine within GHSAG and genome
530 analysis identified highly similar proteins within the *Glaciecola*, indicating that the
531 original host may have been a member of this genus. PMSF significantly reduced
532 activity, confirming Est9x as a serine hydrolase. Maximum activity was identified against
533 C2 with decreasing activity till C14. Optimum conditions were identified as 65°C and pH
534 8. Activity was removed via addition of Co^{2+} whilst Ni^{2+}, Cu^{2+}, Zn^{2+} all decreased activity
535 by 50% and Mn^{2+} and Mg^{2+} only slightly decreased activity. The inability of EDTA to
536 effect the activity of Est9x confirmed it as not being a metalloprotein. Est9x belongs to
537 its own ESTHER family 'Est9X'.

538

539 **3.3.25. Family XXXIII**
540   Bioinformatic analysis of a metagenomic dataset from the Solnechny hot spring
541 in Russia led to the identification of EstDZ2 (29.4 kDa), the type protein for family XXXIII
542 (22). EstDZ2 activity was specific for short-chain fatty acids with maximum activity
543 against C4. Optimum conditions were identified as 55°C and pH 8. Metal ions and
544 EDTA were shown to have minimal effect on activity, indicating family XXXIII are not
545 metalloproteins. Sequence analysis identified the catalytic serine within GHSAG and an

546 expanded search identified no similar proteins within the genome database. Activity was
547 significantly reduced by addition of PMSF, confirming its classification as a serine
548 hydrolase. EstDZ2 was annotated as a member of the ESTHER family 'Hormone-
549 sensitive_lipase_like'.
550
551 **3.3.26. Family XXXIV**
552   Est22 (44.56 kDa) was isolated from a metagenomic library of deep-sea
553 sediment. Activity is specific to short length fatty acids as maximum activity was
554 reported against C4 and no activity occurred past C10. Optimum conditions were
555 identified as 60°C, pH 9 and 0.5M NaCl. The inability of EDTA to alter Est22 activity
556 suggests that metal ions are not required for activity; hence their effect was not tested.
557 Sequence analysis identified the pentapeptide GPSMG containing the catalytic serine
558 as well as similar proteins being present within the *Alteromonadaceae* and the
559 expanded search identified additional matches within the *Pseudoaltermonadaceae*.
560 PMSF was shown to significantly inhibit the activity of Est22. Est22 was identified as
561 belonging to the ESTHER family 'Homoserine_transacetylase'.
562
563 **3.2.27. Family XXXV**
564   The acetyl xylan esterase AcXE2 (25 kDa) was isolated from *Caldicellulosiruptor*
565 *bescii,* a highly thermophilic species able to degrade cellulose (61). As an acetyl xylan
566 esterase, AcXE2 removes the acetyl group on acetylated xylobiose and glucose.
567 Sequence analysis identified the catalytic serine within the pentapeptide GDSIT. Whilst
568 highly similar enzymes were only identified within the *Caldicelulosiruptor*, potential
569 members of this family were also identified within the *Bacillaceae*, *Chthonomonadaceae*
570 and *Paenibacillaceae*, suggesting a diverse host range. Maximum activity was identified
571 against C2 and C4 and optimum temperature of 70°C and pH 7.5. AcXE2 had no match
572 to any family within the ESTHER classification system.
573
574

575 # 4. Conclusions
576   In this paper, we have significantly expanded the classification of lipolytic
577 enzymes to 35 families and 11 true lipase subfamilies. Defining the biochemical
578 features of each family along with providing type proteins allows for a greater number of
579 potentially novel lipases to be assigned to pre-described families, preventing
580 inconsistency within the literature. The lack of multiple members for each family with
581 proven activity prevented comparative analysis of motifs and biochemical features.
582
583   N-terminal single transmembrane alpha-helices were detected within type
584 proteins for 4 families/sub-families. Experimental evidence gained via deletion of the N-
585 terminal transmembrane alpha-helix of LipA identified that this helix lies within the 3D
586 structure of the protein allowing for opening of the lipase cap (62). It is likely that the
587 transmembrane alpha-helices detected within the newly described families play a
588 similar role in the structure and functionality of the proteins and do not represent
589 membrane-bound lipases.
590

591   Previous research has shown that sequence identity does not guarantee shared
592 functionality (63). The average pairwise similarity between the lipolytic enzymes was
593 31.3% with the minimum being 14.2%. Whilst these enzymes do not share identical
594 functionality, they have demonstrated the ability to hydrolyse the release of pNP-bound
595 fatty acids. The low level of sequence similarity between lipolytic enzymes suggests that
596 *in vitro* examination of proteins sharing low homology with known lipolytic enzymes may
597 lead to identification of additional families.
598
599   Cultivation-independent investigation of environmental samples led to the
600 identification and characterisation of multiple lipolytic families suggesting that a greater
601 number of enzymes that have yet to be characterised likely exist. The inability of many
602 families to be assigned to a host via genome search suggests that not only do further
603 lipolytic enzymes remain uncharacterised, but their host species have yet to be cultured.
604 We hope that with the recent increased effort in cultivation of novel species from
605 environmental samples additional enzymes of biotechnological importance will be
606 identified.
607
608
609

5. Acknowledgements

613

6. Conflict of interest

615   The authors declare no conflict of interest.
616

7. Funding information

620

8. Author contribution

622   TCAH performed the analysis and writing with advice and editing provided by TC.
623
624

9. References

626 1.   Drouault S, Juste C, Marteau P, Renault P, Corthier G. Oral Treatment with
627   Lactococcus lactis Expressing Staphylococcus hyicus Lipase Enhances Lipid
628   Digestion in Pigs with Induced Pancreatic Insufficiency. 2002;68(6):3166–8.
629 2.   Canaan S, Maurin D, Chahinian H, Pouilly B, Durousseau C, Frassinetti F, et al.
630   Expression and characterization of the protein Rv1399c from Mycobacterium
631   tuberculosis: A novel carboxyl esterase structurally related to the HSL family. Eur
632   J Biochem. 2004;271(19):3953–61.
633 3.   Jaeger KE, Eggert T. Lipases for biotechnology. Curr Opin Biotechnol.
634   2002;13(4):390–7.

635 4.   Guo J, Zheng X, Xu L, Liu Z, Xu K, Li S, et al. Characterization of a Novel
636   Esterase Rv0045c from Mycobacterium tuberculosis. PLoS One. 2010;5(10).
637 5.   Rao L, Xue Y, Zhou C, Tao J, Li G, Lu JR, et al. A thermostable esterase from
638   Thermoanaerobacter tengcongensis opening up a new family of bacterial lipolytic
639   enzymes. Biochim Biophys Acta - Proteins Proteomics. Elsevier B.V.;
640   2011;1814(12):1695–702.
641 6.   Levisson M, van der Oost J, Kengen SWM. Characterization and structural
642   modeling of a new type of thermostable esterase from Thermotoga maritima.
643   FEBS J. 2007;274(11):2832–42.
644 7.   Glogauer A, Martini VP, Faoro H, Couto GH, Müller-Santos M, Monteiro RA, et al.
645   Identification and characterization of a new true lipase isolated through
646   metagenomic approach. Microb Cell Fact. BioMed Central Ltd; 2011;10(1):54.
647 8.   Yu S, Yu S, Han W, Wang H, Zheng B, Feng Y. A novel thermophilic lipase from
648   Fervidobacterium nodosum Rt17-B1 representing a new subfamily of bacterial
649   lipases. J Mol Catal B Enzym. Elsevier B.V.; 2010;66(1–2):81–9.
650 9.   Masomian M, Abd Rahman RNZR, Salleh AB, Basri M. Analysis of comparative
651   sequence and genomic data to verify phylogenetic relationship and explore a new
652   subfamily of bacterial lipases. PLoS One. 2016;11(3):1–20.
653 10.   Sorokin DY, Jones BE. Improved method for direct screening of true lipase-
654   producing microorganisms with particular emphasis on alkaline conditions.
655   Microbiology. 2009;78(1):125–30.
656 11.   Arpigny JL, Jaeger K-E. Bacterial lipolytic enzymes: classification and properties.
657   Biochem J. 1999;343(1):177.
658 12.   Handrick R, Reinhardt S, Focarete ML, Scandola M, Adamus G, Kowalczuk M, et
659   al. A New Type of Thermoalkalophilic Hydrolase of Paucimonas lemoignei with
660   High Specificity for Amorphous Polyesters of Short Chain-length Hydroxyalkanoic
661   Acids. J Biol Chem. 2001;276(39):36215–24.
662 13.   Ewis HE, Abdelal AT, Lu CD. Molecular cloning and characterization of two
663   thermostable carboxyl esterases from Geobacillus stearothermophilus. Gene.
664   2004;329(1–2):187–95.
665 14.   Lee MH, Lee CH, Oh TK, Song JK, Yoon JH. Isolation and characterization of a
666   novel lipase from a metagenomic library of tidal flat sediments: Evidence for a
667   new family of bacterial lipases. Appl Environ Microbiol. 2006;72(11):7406–9.
668 15.   Bender J, Rydzewski K, Broich M, Schunder E, Heuner K, Flieger A.
669   Phospholipase PlaB of Legionella pneumophila represents a novel lipase family.
670   Protein residues essential for lipolytic activity, substrate specificity and hemolysis.
671   J Biol Chem. 2009;284(40):27185–94.
672 16.   Kim EY, Oh KH, Lee MH, Kang CH, Oh TK, Yoon JH. Novel cold-adapted alkaline
673   lipase from an intertidal flat metagenome and proposal for a new family of
674   bacterial lipases. Appl Environ Microbiol. 2009;75(1):257–60.
675 17.   Lee MH, Hong KS, Malhotra S, Park JH, Hwang EC, Choi HK, et al. A new
676   esterase EstD2 isolated from plant rhizosphere soil metagenome. Appl Microbiol
677   Biotechnol. 2010;88(5):1125–34.
678 18.   Hu Y, Fu C, Huang Y, Yin Y, Cheng G, Lei F, et al. Novel lipolytic genes from the
679   microbial metagenomic library of the South China Sea marine sediment. FEMS
680   Microbiol Ecol. 2010;72(2):228–37.

681 19. Jeon JH, Kim JT, Lee HS, Kim SJ, Kang SG, Choi SH, et al. Novel lipolytic
682   enzymes identified from metagenomic library of deep-sea sediment. Evidence-
683   based Complement Altern Med. 2011;2011.
684 20. Bassegoda A, Pastor FIJ, Diaz P. Rhodococcus sp. strain CR-53 lipr, the first
685   member of a new bacterial lipase family (Family X) displaying an unusual Y-type
686   oxyanion hole, similar to the Candida antarctica lipase clan. Appl Environ
687   Microbiol. 2012;78(6):1724–32.
688 21. Charbonneau DM, Beauregard M. Role of Key Salt Bridges in Thermostability of
689   G. thermodenitrificans EstGtA2: Distinctive Patterns within the New Bacterial
690   Lipolytic Enzyme Family XV. PLoS One. 2013;8(10):1–18.
691 22. Zarafeta D, Moschidi D, Ladoukakis E, Gavrilov S, Chrysina ED, Chatziioannou A,
692   et al. Metagenomic mining for thermostable esterolytic enzymes uncovers a new
693   family of bacterial esterases. Sci Rep. Nature Publishing Group;
694   2016;6(November):1–16.
695 23. Castilla A, Panizza P, Rodríguez D, Bonino L, Díaz P, Irazoqui G, et al. A novel
696   thermophilic and halophilic esterase from Janibacter sp. R02, the first member of
697   a new lipase family (Family XVII). Enzyme Microb Technol. 2017;98(Family
698   XVII):86–95.
699 24. Parapouli M, Foukis A, Stergiou PY, Koukouritaki M, Magklaras P, Gkini OA, et al.
700   Molecular, biochemical and kinetic analysis of a novel, thermostable lipase
701   (LipSm) from Stenotrophomonas maltophilia Psi-1, the first member of a new
702   bacterial lipase family (XVIII). J Biol Res. BioMed Central; 2018;25(1):1–12.
703 25. Kovacic F, Babic N, Krauss U, Jaeger K. Classification of Lipolytic Enzymes from
704   Bacteria. In: Aerobic Utilization of Hydrocarbons, Oils and Lipids. 2019. p.
705   Chapter 24.
706 26. Chu X, He H, Guo C, Sun B. Identification of two novel esterases from a marine
707   metagenomic library derived from South China Sea. Appl Microbiol Biotechnol.
708   2008;80(4):615–25.
709 27. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. Identification of
710   novel lipolytic genes and gene families by screening of metagenomic libraries
711   derived from soil samples of the German Biodiversity Exploratories. FEMS
712   Microbiol Ecol. 2011;78(1):188–201.
713 28. Kang HY, Kim JF, Kim MH, Park SH, Oh TK, Hur CG. MELDB: A database for
714   microbial esterases and lipases. FEBS Lett. 2006;580(11):2736–40.
715 29. Messaoudi A, Belguith H, Ghram I, Hamida J Ben. LIPABASE: a database for
716   "true" lipase family enzymes. Int J Bioinform Res Appl. 2011;7(4):390.
717 30. Pleiss J, Fischer M, Peiker M, Thiele C, Schmid RD. Lipase engineering
718   database: Understanding and exploiting sequence-structure-function
719   relationships. J Mol Catal - B Enzym. 2000;10(5):491–508.
720 31. Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P, Chatonnet A. ESTHER, the
721   database of the α/β-hydrolase fold superfamily of proteins: Tools to explore
722   diversity of functions. Nucleic Acids Res. 2013;41(D1):423–9.
723 32. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of
724   progressive multiple sequence alignment through sequence weighting, position-
725   specific gap penalties and weight matrix choice. Nucleic Acids Res.
726   1994;22(22):4673–80.

727   33.   Stothard P. The Sequence Manipulation Suite: javascript programs for analysing
728   and formating protein and DNA sequences. Biotechniques. 2000;28(6):1102–4.
729   34.   Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane
730   protein topology with a hidden Markov model: Application to complete genomes. J
731   Mol Biol. 2001;305(3):567–80.
732   35.   Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank.
733   Nucleic Acids Res. 2008;36(SUPPL. 1):25–30.
734   36.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
735   search tool. J Mol Biol. 1990 Oct;215(3):403–10.
736   37.   Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A
737   large-scale evaluation of computational protein function prediction. Nat Methods.
738   2013;10(3):221–7.
739   38.   Veeramachaneni V, Makałowski W. Visualizing sequence similarity of protein
740   families. Genome Res. 2004;14(6):1160–9.
741   39.   Rubino F, Mccabe MS, Newbold J, Veneman JB, Creevey CJ. MGkit :
742   Metagenomic Framework For The Study Of Microbial Communities communities
743   from a taxonomic and functional perspective , little has Rationale analysis of
744   metagenomic data that includes fully automated analysis and carry out statistical
745   tests to iden.
746   https://figshare.com/articles/MGkit_Metagenomic_Framework_For_The_Study_Of
747   _Microbial_Communities/1269288. 2014.
748   40.   Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P, et al.
749   A standardized bacterial taxonomy based on genome phylogeny substantially
750   revises the tree of life. bioRxiv. Nature Publishing Group; 2018;(August):256800.
751   41.   Samoylova Y V., Sorokina KN, Romanenko M V., Parmon VN. Cloning,
752   expression and characterization of the esterase estUT1 from Ureibacillus
753   thermosphaericus which belongs to a new lipase family XVIII. Extremophiles.
754   Springer Japan; 2018;22(2):271–85.
755   42.   Rao L, Xue Y, Zheng Y, Lu JR, Ma Y. A Novel Alkaliphilic Bacillus Esterase
756   Belongs to the 13th Bacterial Lipolytic Enzyme Family. PLoS One. 2013;8(4).
757   43.   Charbonneau DM, Meddeb-Mouelhi F, Beauregard M. A novel thermostable
758   carboxylesterase from Geobacillus thermodenitrificans: Evidence for a new
759   carboxylesterase family. J Biochem. 2010;148(3):299–308.
760   44.   Kitaura S, Suzuki K, Imamura S. Monoacylglycerol lipase from moderately
761   thermophilic Bacillus sp. Strain H-257: Molecular cloning, sequencing, and
762   expression in Escherichia coli of the gene. J Biochem. 2001;129(3):397–402.
763   45.   Chow J, Kovacic F, Dall Antonia Y, Krauss U, Fersini F, Schmeisser C, et al. The
764   Metagenome-Derived Enzymes LipS and LipT Increase the Diversity of Known
765   Lipases. PLoS One. 2012;7(10).
766   46.   Liu K, Wang J, Bu D, Zhao S, McSweeney C, Yu P, et al. Isolation and
767   biochemical characterization of two lipases from a metagenomic library of China
768   Holstein cow rumen. Biochem Biophys Res Commun. Elsevier Inc.;
769   2009;385(4):605–11.
770   47.   Lee MH, Oh KH, Kang CH, Kim JH, Oh TK, Ryu CM, et al. Novel metagenome-
771   derived, cold-adapted alkaline phospholipase with superior lipase activity as an
772   intermediate between phospholipase and lipase. Appl Environ Microbiol.

773   2012;78(14):4959–66.

774 48. Latif H, Lerman JA, Portnoy VA, Tarasova Y, Nagarajan H, Schrimpe-Rutledge
775   AC, et al. The Genome Organization of Thermotoga maritima Reflects Its
776   Lifestyle. PLoS Genet. 2013;9(4).

777 49. Raddadi N, Giacomucci L, Totaro G, Fava F. Marinobacter sp. from marine
778   sediments produce highly stable surface-active agents for combatting marine oil
779   spills. Microb Cell Fact. BioMed Central; 2017;16(1):1–13.

780 50. Yan YW, Jiang QY, Wang JG, Zhu T, Zou B, Qiu QF, et al. Microbial communities
781   and diversities in mudflat sediments analyzed using a modified
782   metatranscriptomic method. Front Microbiol. 2018;9(JAN):1–15.

783 51. Golyshin PN, Chernikova TaN, Abraham WR, Lünsdorf H, Timmis KN, Yakimov
784   MM. Oleiphilaceae fam. nov., to include Oleiphilus messinensis gen. nov., sp.
785   nov., a novel marine bacterium that obligately utilizes hydrocarbons. Int J Syst
786   Evol Microbiol. 2002;52(3):901–11.

787 52. Li M, Yang LR, Xu G, Wu JP. Cloning and characterization of a novel lipase from
788   Stenotrophomonas maltophilia GS11: The first member of a new bacterial lipase
789   family XVI. J Biotechnol. 2016;228:30–6.

790 53. Couto GH, Glogauer A, Faoro H, Chubatsu, L.S. SEM, Pedrosa, et al. Isolation of
791   a novel lipase from a metagenomic library derived from mangrove sediment from
792   the south Brazilian coast. Genet Mol Res. 2010;9(1):514–23.

793 54. Zheng X, Guo J, Xu L, Li H, Zhang D, Zhang K, et al. Crystal structure of a novel
794   esterase RV0045c from mycobacterium tuberculosis. PLoS One. 2011;6(5).

795 55. Ortega C, Anderson LN, Frando A, Sadler NC, Brown RW, Smith RD, et al.
796   Systematic Survey of Serine Hydrolase Activity in Mycobacterium tuberculosis
797   Defines Changes Associated with Persistence. Cell Chem Biol. Elsevier Ltd;
798   2016;23(2):290–8.

799 56. Kim Y-E, Yoon H, Kim M, Nam Y-J, Kim H, Seo Y, et al. Metagenomic analysis of
800   bacterial communities on Dokdo Island. J Gen Appl Microbiol. 2014;60(2):65–74.

801 57. Jiménez DJ, Montaña JS, Álvarez D, Baena S. A novel cold active esterase
802   derived from Colombian high Andean forest soil metagenome. World J Microbiol
803   Biotechnol. 2012;28(1):361–70.

804 58. Ouyang LM, Liu JY, Qiao M, Xu JH. Isolation and biochemical characterization of
805   two novel metagenome-derived esterases. Appl Biochem Biotechnol.
806   2013;169(1):15–28.

807 59. Wu G, Zhang S, Zhang H, Zhang S, Liu Z. A novel esterase from a psychrotrophic
808   bacterium Psychrobacter celer 3Pb1 showed cold-adaptation and salt-tolerance. J
809   Mol Catal B Enzym. Elsevier B.V.; 2013;98:119–26.

810 60. Fang Z, Li J, Wang Q, Fang W, Peng H, Zhang X, et al. A Novel Esterase from a
811   Marine Metagenomic Library Exhibiting Salt Tolerance Ability. J Microbiol
812   Biotechnol. 2014;24(246):771–80.

813 61. Soni S, Sathe SS, Odaneth AA, Lali AM, Chandrayan SK. SGNH hydrolase-type
814   esterase domain containing Cbes-AcXE2: a novel and thermostable acetyl xylan
815   esterase from Caldicellulosiruptor bescii. Extremophiles. Springer Japan;
816   2017;21(4):687–97.

817 62. Zha D, Zhang H, Zhang H, Xu L, Yan Y. N-terminal transmembrane domain of
818   lipase LipA from Pseudomonas protegens Pf-5: A must for its efficient folding into

819   an active conformation. Biochimie. 2014;105:165–71.

820   63.   Punta M, Ofran Y. The rough guide to in silico function prediction, or how to use
821   sequence and structure information to predict protein function. PLoS Comput Biol.
822   2008;4(10).

823

824

825

826

827

828

Table 1(on next page)

Bacterial lipolytic enzyme families.

Each family is provided with a type protein for which the accession number and name is provided along with its proposed function.

Family	Sub-family	Type protein	Accession	Function
Family I	1		D50587	True lipase
	2		X70354	True lipase
	3		D11455	True lipase
	4		M74010	True lipase
	5		U78785	True lipase
	6		M12715	True lipase / Phospholipase
	7		X99255	True lipase
	8	Lip1	Q3IF07	True lipase
	9	LipG	DQ458963	True lipase
	10	RlipE1	FJ529693	-
	11	MPlaG	EU285670	True lipase / Phospholipase
Family II			P10480	Secreted acyltranferase
Family III			M86351	Extracellular lipase
Family IV			X62835	Esterase
Family V			M58445	PHA-depolymerase
Family VI			D90904	Carboxylesterases
Family VII			Q01470	Carbamate hydrolase
Family VIII			AAA99492	Stereoselective esterase
Family IX		phaZ7	AY026355	Poly(3-hydroxylbutyrate) depolymerase
Family X		EstD	NP_228147	Carboxylesterase
Family XI		EstA	ABY60416	Esterase
Family XII		LipEH166	EU515239	Lipase
Family XIII		Est30	AY186197	Carboxylesterase
Family XIV		EstA3	NP_623858	Carboxylesterase
Family XV		EstGtA2	AEN92268	Carboxylesterase
Family XVI		LipSM54	AGF29555	Lipase
Family XVII		LipJ2	KX096709	Lipase
Family XVIII		plaB	EF408871	Phosphatidylcholine-specific phospholipase
Family XIX		LipSM	KX353755	Lipase
Family XX		FLS18D	ACL67852	Esterase
Family XXI		FnL	ABS61180	Lipase
Family XXII		EstD2	GQ866023	Secreted esterase
Family XXIII		LipA	ACJ13070	Lipase
Family XXIV		Rv0045c	NP_214559	Esterase
Family XXV		EM3L4	GQ340926	Lipase
Family XXVI		EstGS	HQ156900	Secreted esterase
Family XXVII		EstGH	HQ156914	Esterase
Family XXVIII		LipR	YP_345621	Lipase
Family XXIX		EstGX1	HQ262952	Lipase
Family XXX		EstP2K	JN001203	Esterase
Family XXXI		Est12	KF313138	Esterase
Family XXXII		Est9x	AFR79233	Lipase
Family XXXIII		EstDZ2	KX301277	Secreted carboxylesterase

Family XXXIV	Est22	GAC12013	Homoserine transacetylase
Family XXXV	AcXE2	ACM59679	Acetyl xylan esterase

1