Systematic Analysis of Metabolic Pathway Distributions of Bacterial Energy Reserves

Liang Wang[1,2*], Jianye Yang[1], Yue Huang[1], Qinghua Liu[2], Yaping Xu[1], Xue Piao[1,3], Michael J. Wise[4,5]

[1]Department of Bioinformatics, School of Medical Informatics, Xuzhou Medical University, Xuzhou, Jiangsu Province, 221000, China

[2]Jiangsu Key Lab of New Drug Research and Clinical Pharmacy, Xuzhou Medical University, Xuzhou, Jiangsu Province, 221000, China

[3]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China

[4]Department of Computer Science and Software Engineering, University of Western Australia, Perth 6009, WA, Australia

[5]Department of Computer Science and Software Engineering, University of Western Australia, Perth 6009, WA, Australia

[*]For correspondence, please refer to Dr. Liang Wang at leonwang@xzhmu.edu.cn

Abstract

Metabolism of energy reserves are essential for bacterial functions such as pathogenicity, metabolic adaptation, and environmental persistence, *etc*. Previous bioinformatics studies have linked gain or loss of energy reserves such as glycogen and polyphosphate (polyP) with host-pathogen interactions and bacterial virulence based on a comparatively small number of bacterial genomes or proteomes. Thus, understanding the distribution patterns of energy reserves metabolism across bacterial species provides a shortcut route to look into bacterial lifestyle and physiology theoretically. So far, five major energy reserves have been identified in bacteria due to their effective capacity to support bacterial persistence under nutrient deprivation conditions, which include wax ester (WE), triacylglycerol (TAG), polyhydroxyalkanoates (PHA), polyphosphate, and glycogen. Although unknown pathways directly involved in energy reserves keep being discovered with the continuous endeavour of molecular microbiologists and it is currently rather clear about the enzymes related with the metabolism of energy reserves, there is a lack of systematic study of the pathway or key enzyme distributions of the five energy reserves in bacteria from an evolutionary point of

view. With the fast development of sequencing technology, abundant bacterial proteomes are available in public database now. In this study, we sourced 8214 manually reviewed bacterial reference proteomes from UniProt database and used statistical models to search homologous sequences of key enzymes related with energy reserves. The distribution patterns of the pathways for energy reserves metabolism are visualized in taxonomy-based phylogenetic trees. According to the study, it was revealed that specific pathways and enzymes are associated with certain types of bacterial groups, which provides evolutionary insights into the understanding of their origins and functions. In addition, the study also confirmed that loss of energy reserves is correlated with bacterial genome reduction. Through this analysis, a much clearer picture about energy reserves metabolism in bacteria is present, which could serve a guide for further theoretical and experimental analyses of energy reserves metabolism in bacteria.

Keywords

Energy reserve, Hidden Markov model, Evolution, Proteome, Taxonomy

Introduction

Due to the versatility of environmental niches that bacteria colonize through millions of years adaptation and evolution, bacteria have been equipped with specialized sets of metabolic pathways so as to live optimally in these environments, which can be reflected in their characteristic genomes, gene transcription profiles, and also proteomes.[1] Previously, comparative genomics studies have shown that glycogen metabolism loss could serve as an indicator for bacterial parasitic lifestyle and polyP metabolism gain seems to link with free-living lifestyle and stronger bacterial virulence potential.[2,3] In addition, it has been observed that metabolism loss of glycogen or polyP is associated with shrunk genome size, leading to a hint of genome reduction trend.[2,4] Both glycogen and polyphosphate are representative energy reserves in bacteria. Thus, presence or absence of energy reserves in bacteria could be important for *in silico* analysis of bacterial physiology and lifestyle, especially when large number of sequenced bacterial genomes are available and many of those are unculturable through traditional laboratory techniques.

It is well known now that energy reserves play essential roles in bacteria for their regular activities to sense and respond to the changing environments with different types of stresses, such as temperature fluctuation and nutrient deprivation, *etc*.[4] Although there are many different energy-related compounds, not all of them can be classified as energy reserves. According to Wilkinson, three principles should be satisfied for a compound to be considered as an energy reserve, which are: 1) accumulation when energy is over-supplied, 2) utilization when energy is insufficient, and 3) apparent advantages by consuming the compound when compared with those without it.[5] Through physiological and biochemical studies, five energy storage compounds are regarded to meet the criteria, which are WE, TAG, PHAs, polyP and glycogen.[2,4-6]

Although several studies have attempted to investigate the distributions of these energy reserves in bacteria, most of these studies are based on small sets of bacterial genomes or proteomes. No systematic analysis exists that is based on current available reference proteomes from the evolutionary point of view.[2,4,7-10] In this study, we collected 8214 manually-reviewed bacterial proteomes from UniProt database[11] and sourced key enzymes involved directly in the metabolism of energy reserves from public literature and database

(**Table 1**). Full statistical sequence models were constructed *de novo* for these enzymes based on hidden Markov models via HMMER package, which were then used to screen for homologous sequences in all bacterial proteomes.[12] All distribution patterns were presented in **Supplementary Table S1**. In order to gain a clearer view about enzymes distributions of the enzymes along the evolutionary paths, we incorporated phylogenetic trees with our dataset via NCBI taxonomy identifiers.[13] Through a combinational analysis of the pathways in the phylogenetic tree, we identified interesting distribution patterns of energy reserves that are linked with specific groups of bacteria and their lifestyle potentials, which may improve our understanding of the functions of energy reserves in bacteria. In addition, systematic analysis also gives us an overview of enzyme distributions, which could serve a guide for further theoretical and experimental analyses of energy reserves metabolism in bacteria.

Materials and Methods

Proteomes and enzymes collection

Bacterial proteomes were downloaded from UniProt database by using two keywords, *Bacteria* and *Reference Proteomes*, as filters.[11] A total of 8282 bacterial proteomes were collected at the downloaded time in 2018 and only 8214 bacterial proteomes were kept due to outdated taxonomy identifiers that cannot be identified in NCBI taxonomy database when constructing phylogenetic trees.[13] A complete list of all the bacteria with bacterial names, UniProt proteome identifiers, proteome sizes, and distribution patterns of enzymes is available in the **Supplementary Table S1**. As for the metabolism of the five major energy reserves, only key enzymes are considered. For the synthesis of WE and TAG, wax ester synthase/acyl-CoA:diacylglycerol acyltransferase (WS/DGAT) is studied while triacylglycerol lipase (Lip) was screened due to its essential role in the degradation of WE and TAG.[10,14] For metabolism of another neutral lipid polyhydroxyalkanoates, enzymes PhaA, PhaB, PhaC are in the synthesis pathway and PhaZ is in the degradation pathway.[15] There are two different PhaZs, that is, intracellular PhaZ and extracellular PhaZ, both of which are analysed in this study. As for polyP, the key enzyme PPK1 for synthesis and two degradation enzymes, intracellular PPK2 and extracellular PPX, are included.[2] Finally, for glycogen metabolism, two synthesis pathways are considered. The first one involves GlgC, GlgA, and GlgB.[4] The second pathway is TreS, Pep2, GlgE, and GlgB.[16] Key enzyme Rv3032 in another pathway related with glycogen metabolism and capsular glucan is only briefly

mentioned.[16] In addition, archaeal type GlgB was also included for comparative analysis. For details of these enzymes, please refer to **Table 1**.

De novo construction of HMMs

Seed sequences related with metabolism of energy reserves were selected through a comprehensive up-to-date review of literature and were listed in **Table 1**. These seed proteins were used for constructing statistical sequence models based on HMMs via HMMER package.[12] After obtaining sequences for all seed proteins, remote BLAST was performed to collect homologous sequences for each seed protein from the NCBI non-redundant database of protein sequences.[17] Perl script nrdb90.pl was used to remove the homologous sequences with more than 90% similarity from the selected proteins.[18] The standalone command-line version of MUSCLE was used so the MSAs were done automatically.[19] Heads or tails of multiple sequence alignments tend to be more inconsistent.[20] Thus, all MSAs were manually edited to remove heads and tails by using JalView.[21] HMMER was selected for the construction of HMMs through hmmbuild command. Since HMMER only recognizes STOCKHOLM format, all MSAs results were converted from FASTA to STOCKHOLM format. For searching homologs in bacterial proteomes, routine procedures are performed by following HMMER User's Guide eddylab.org/software/hmmer3/3.1b2/Userguide.pdf. Results obtained from HMM screening were present in **Supplementary Table S1**. The presence (copy numbers) or absence of a specific enzyme in a certain bacterial proteome is present.

Data visualization

Phylogenetic trees were first constructed based on NCBI taxonomy identifiers for all bacteria in this study via commercial web server PhyloT https://phylot.biobyte.de/, which were then visualized through the online interactive Tree of Life (iTOL) server https://itol.embl.de/.[22] Distribution patterns of enzymes and their combinations in terms of energy reserves were added to the trees through iTOL pre-defined tol_simple_bar template.[22]

Statistical analysis

Unpaired two-tailed Student's *t*-test was used for statistical analysis. Significant difference was defined as *P-value* < 0.05.

Results

Wax ester and triacylglycerol

The key enzyme that is involved in WE and TAG synthesis in bacteria is WS/DGAT. Through the screening of HMM-based statistical models, 950 out of 8214 bacterial species harbour single copy or multiple copies of WS/DGAT homologs, which are mainly present in phylum *Actinobacteria* and the super-phylum *Proteobacteria*. Only sporadic bacteria in groups such as FCB (a.k.a *Sphingobacteria*) and PVC (a.k.a. *Planctobacteria*), *etc*. have WS/DGAT. No species belonging to phylum *Firmicutes* has WS/DGAT. As for the unclassified bacteria, although no WS/DGAT is identified, they will not be studied due to their unclear nature at current stage. For details, please refer to **Figure 1(A)**. By comparing the proteome sizes of bacteria species with or without WS/DGAT, we found that bacteria with WS/DGAT have average proteome size of 5200 proteins/proteome while those without WS/DGAT have average proteome size of 3047 proteins/proteome (*P-value*<0.001).

[insert Figure 1.]

Within the major phylum of *Proteobacteria*, WS/DGAT is not evenly distributed and γ-*Proteobacteria* have more species with multi-copy WS/DGAT. In addition, two orders *Rhodobacterales* (305 species) and *Enterobacterales* (168 species) that belong to α- and γ-*Proteobacteria* phylum, respectively do not have any WS/DGAT except for one species *Ahrensia* sp. R2A130. As for the phylum *Actinobacteria*, three WS/DGAT abundant regions (R1, R2, and R3) and one WS/DGAT absence region (R4) in the phylogenetic tree are worth of further exploration. R1 includes only one family *Mycobacteriaceae* (115 species) in which bacteria have up to 24 homologs of WS/DGAT. R2 includes families of *Dietziaceae*, *Gordoniaceae*, *Nocardiaceae*, *Tsukamurellaceae*, and *Williamsiaceae*. R3 includes families of *Nocardioidaceae* and *Pseudonocardiales*. R4 is the family *Corynebacteriaceae* (69 species) that have WS/DGAT-free bacteria only. As for the degradation of TAG and probable degradation of WE in bacteria, Lip plays important roles. Only 394 bacterial species have the

enzyme. Distribution analysis showed that this enzyme is almost exclusively present in the super phylum *Proteobacteria*, especially in the β-, γ-, and δ-*Proteobacteria* phylums.

Polyhydroxyalkanoates

Three enzymes (PhaA, PhaB and PhaC) involve in the synthesis of PHAs and two enzymes (intracellular PhaZ and extracellular PhaZ) involve in the utilization of PHAs in bacteria. In this study, we focused on the distribution patterns of PHA synthesis pathway PhaABC and also the two degradation enzymes. Please refer to **Figure 1(B)** for details. Preliminary analysis showed that 836 bacterial species with average proteome size of 4891 proteins/proteome have PhaABC pathway while 536 bacterial species with average proteome size of 718 proteins/proteome do not have the pathway (*P-value*<0.001). In addition, when PhaABC is missing in bacteria, degradation enzymes are also absent except for one species *Chloroflexi bacterium*. Evolutionary analysis based on phylogenetic tree found that complete PHA synthesis and degradation pathway mainly distribute in α- and β-*Proteobacteria*. Some bacteria in phylum *Actinobacteria* and genus *Bacillus* also harbor PhaABC synthesis pathway. However, intracellular PhaZ is rarely present in these species. That is the major difference when comparing with *Proteobacteria* phylums.

Polyphosphate

Three key enzymes are related with polyP metabolism. PPK1 is responsible for polyP synthesis. A total of 5209 bacterial species has PPK1. PPK2 and PPX are used for polyP degradation intracellularly and extracellularly, respectively. 3330 bacterial species have PPK1, PPK2 and PPX enzymes while 2215 bacteria species do not have any of the three enzymes. Their average proteome sizes are 4459 proteins/proteome and 1618 proteins/proteome, respectively (*P-value*<0.001). In our analysis, we independently reviewed the distribution patterns of the three enzymes along phylogenetic trees and the result is displayed in **Figure 1(C)**. The three enzymes are widely distributed across bacterial species. Comparison shows that *Firmicutes* phylum seems to favour PPX more than PPK2 for polyP degradation. In addition, although it was observed that several regions have missing synthesis enzyme or degradation enzyme, only *Mollicutes* class (94 bacterial species) shows apparent lack of the three polyP metabolism enzymes in the phylogenetic tree.

Glycogen

Glycogen metabolism in bacteria has multiple pathways, which includes the classical pathway (GlgC, GlgA, GlgB, GlgP and GlgX), trehalose pathway (TreS, Pep2, GlgE, and GlgB), and Rv3032 pathway. In this study, we only focused on the first two glycogen synthesis pathways and compares their distribution patterns in bacteria. In addition, there are two types of glycogen branching enzymes. One is the common bacterial GlgB belonging to GH13 in CATH database and the other one is known as archaeal GlgB belonging to GH57 in CATH database.[23] We also look into their distribution patterns in bacteria since GlgB is essential in determining glycogen branched structure. Our study showed that 3924 bacteria has classical synthesis pathway (GlgC, GlgA, and GlgB) and their average proteome size is 4163 proteins/proteome while only 489 bacterial species (average proteome size of 1050 proteins/proteome) do not have these enzymes (*P-value*<0.001). Comparison of the two synthesis pathways confirmed that classical synthesis pathways are widely distributed across species. Random loss of the pathways can be inferred from **Figure 1(D)**. In contrast, trehalose pathway is tightly associated with *Actinobacteria* phylum with rarely sporadic presence in other bacterial species. As for the two GlgBs, GH13 GlgB is widely distributed in 4451 bacterial species with a trend of random loss while GH57 GlgB is found in 755 bacteria that are mainly belonging to groups such as *Terrabacteria* and PVC, *etc*.

Discussion

From an evolutionary point of view, if an organism can obtain compounds from other sources, it will tend to discard its own biosynthetic pathway.[24] For example, *Rickettsia* species, *Mycoplasma* species, and *Buchnera*, etc. have extensively reduced genome and abundant energy metabolism pathways are eliminated.[24] In addition, although common belief is that organism will evolve toward complexity, recent analysis supported that reduction and simplification could be the dominant mode of evolution while complexification is just a transit stage based on the neutral genetic loss and streamlining hypothesis.[25] Independent analyses of the distribution patterns of the five energy reserves in bacteria found a consistent and statistically significant correlation between energy reserve loss and reduced proteome size. Previous studies have already reported such a correlation for glycogen and polyP in bacteria.[2,4] In this study, we extend this conclusion by the addition of the reserves WE, TAG, and PHAs, which has not been reported before. It has been confirmed that most of bacteria

losing energy reserve metabolism pathways tend to have a niche-dependent or host-dependent lifestyle, which is the case in our study[2,4,6] Thus, by simply looking into bacterial energy reserve metabolism, we could obtain preliminary views in terms of their lifestyles, though other factors and evidences are required for verification.

WS/DGAT is a bifunctional enzyme and key to the biosynthesis of WE and TAG in bacteria. It was previously thought that WE and TAG are very uncommon lipid storage compounds in bacteria when compared with plants and animals until the novel enzyme was identified.[10] From our analysis, it can be seen that many bacteria belonging to both Gram-positive and Gram-negative categories have the ability to synthesize WE and TAG. However, studies about WS/DGAT are mainly restricted to *Mycobacteria* genus (*Actinobacteria* phylum) and *Acinetobacter* genus (γ-Proteobacteria phylum) due to their clinical significance and industrial use potentials, respectively. It is also apparent to notice that WS/DGAT in phylum *Actinobacteria* tend to have number of homologs far more than other phylums, especially for the bacteria in the three regions (R1, R2, and R3) mentioned above, which include *Mycobacteriaceae*, *Dietziaceae*, *Gordoniaceae*, *Nocardiaceae*, *Tsukamurellaceae*, *Williamsiaceae*, *Nocardioidaceae* and *Pseudonocardiales*. On the other hand, no WS/DGAT is found in the family of *Corynebacteriaceae* (R4), although *Corynebacteriaceae* is closely related with *Mycobacteriacea*.[26] In addition, bacteria in phylum *Firmicutes* do not have any WS/DGAT enzyme, neither. Screen of Phospholipid:diacylglycerol acyltransferase (PDAT), an enzyme that catalyses the acyl-CoA-independent formation of triacylglycerol in yeast and plants, found no homologs in bacteria at all.[27] Thus, this enzyme is exclusively present in eukaryotic organisms.

The family *Corynebacteriaceae* contains the genera *Corynebacterium* and monospecific genus *Turicella*.[28] *Mycobacterium tuberculosis* is the dominant species in *Mycobacteriaceae* (97 species). Mycolic acid (MA) with wax ester is the oxidized form of MA in *Mycobacterium tuberculosis*, which forms an integrated cell wall and plays essential role in host invasion, environmental persistence, and also drug resistance. In addition, *Mycobacterium tuberculosis* also relies on wax ester for dormancy, although WE function in *M. tuberculosis* requires more investigations. Thus, abundance of WS/DGAT in *Mycobacteriacea* has selective advantages in evolution. Considering the abundance of wax ester and its slow degradation, it could also contribute to the long-term survival (more than 360 days) of *M. tuberculosis* in environment.[6] On the other hand, *Corynebacterium* do not

rely on oxidized mycolic acid while *Turicella* does not have mycolic acid at all.[29,30] Thus, there is no need for them to be equipped with the WS/DGAT enzyme. However, how *Mycobacteriaceae* gain WS/DGAT or *Corynebacteriaceae* lose it is not clear and needs more studies. As for *Firmicutes*, the low G+C counterpart of the high G+C *Actinobacteria*, most of its species can form endospores and are resistance to extreme environmental conditions such as desiccation, temperature fluctuation, and nutrient deprivation, *etc*.[31] Thus, they may not need compounds such as WE or TAG for storing energy and dealing with harsh external conditions. How G+C content in the two phylums may impact on the gain or loss of wax ester metabolism is not known.

PHAs are a group of compounds that include but not limited to components such as polyhydroxybutyrate (PHB) and polyhydroxyvalerate (PHV), *etc*., among which PHB is the most common and most prominent member in bacteria.[32,33] Synthesis of PHB involves PhaA, PhaB and PhaC.[32] In addition, there are two types of PHB degradation enzymes, intracellular PhaZ and extracellular PhaZ.[34] Phylogenetic analysis revealed that PHB metabolism is mainly in *Proteobacteria* and *Actinobacteria* (**Figure 1(B)**). Moreover, *Bacillus* genus is also abundant in PHB metabolism. The major difference between *Proteobacteria* and *Actinobacteria* is that *Actinobacteria* rarely has intracellular PhaZ and relies mainly on extracellular PhaZ for PHB degradation. The reason behind the difference could be that *Actinobacteria* has other mechanisms for PHB degradation intracellularly. In addition, *Actinobacteria* are widespread in water and soil and frequently experience nutrient shortage.[35] Thus, they are more focused on storing PHB intracellularly as energy reserve and synthesizing extracellular PhaZ to utilize environmental PHBs released by other organisms.[36] By doing so, they could have great advantage over other organisms in term of viability under harsh conditions. It is also interesting to notice that *Bacillus* genus accumulates PHB intracellularly and has been developed for industrial production of PHB.[37]

PolyP has been known to be ubiquitous in different life domains and claimed to be present in all cells in nature due to their essential roles as energy and phosphate sources.[2] Although a bunch of enzymes are directly linked with polyP metabolism, we only focused on PPK1, PPK2, and PPX in this study because they are most abundant and essential enzymes. **Figure 1(C)** gives an overview of the universally wide distribution of the three enzymes. Although 2215 species across the phylogenetic tree are lack of all three enzymes, an apparent gap was only spotted in the phylum *Tenericutes* and was further confirmed to be *Mollicutes*. A

previous analysis of 944 bacterial proteomes have shown that bacteria with complete loss of polyP metabolism (PPK1, PPK2, PAP, SurE, PPX, PpnK and PpgK) pathway are heavily host-dependent and tend to be obligate intracellular or symbiotic.[2] Consistently, *Mollicutes* is a group of parasitic bacteria that are evolved from a common *Firmicutes* ancestor through reductive evolution.[38] From here, we can infer that not only loss of complete metabolism pathway, but also even loss of key enzymes in an energy reserve metabolism could give a hint about bacterial lifestyle.

For glycogen metabolism, we compared two synthesis pathways, the classical pathway (GlgC, GlgA and GlgB) and the newly identified trehalose-related pathway (TreS, Pep2, GlgE and GlgB).[4,16] Although initial analysis via BLAST search showed that GlgE pathway is represented in 14% of sequenced genomes from diverse bacteria in 2011, our studies showed that, when searching for the complete GlgE pathway by including another three enzymes, it is dominantly restricted to *Actinobacteria* phylum while classical pathway is widely present in the tree as seen in **Figure 1(D)**.[16] In addition, the two types of GlgBs also show interesting patterns. Although GH13 GlgB is widely identified in 54.18% bacteria, GH57 GlgB is only present in 9.19% bacteria with skewed distribution in groups such as *Terrabacteria* phylum and PVC group, *etc*. Another study of 427 archaea proteomes found that the 11 archaea has GH13 GlgB while 18 archaea has GH57 GlgB (unpublished data). Thus, the two GlgBs are rarely present in archaea and mainly exist in bacteria. However, why trehalose-related glycogen metabolism pathway is associated with *Actinobacteria* phylum still needs more experimental exploration.

Conclusions

Distribution patterns of key enzymes and their combined pathways in bacteria provided a comprehensive view of how energy reserves are present and absent during evolutionary process. In general, polyP is most abundant energy reserve in bacteria while polysaccharide glycogen is the second most abundant compound. However, glycogen has multiple synthesis pathways and its metabolism could have more impacts on bacterial physiology due to such flexibility. For the three neutral lipids, there are comparatively minor energy reserves in bacteria and mainly constrained in super phylum *Proteobacteria* and phylum *Actinobacteria*. Within the group, more bacteria have the capacity to accumulate WE and TAG rather than PHB due to the widespread distribution of *wax-dgaT* homologs. polyP only acts as a transit

373 energy reserve while neutral lipids are more sustainable energy provider.[4,39] Thus, neutral
374 lipids could be major player for bacterial persistence under harsh conditions such terrestrial
375 and aquatic environments. As for glycogen, its ability for bacterial environmental viability is
376 still controversial due to the consideration of its interior structures. Its widespread distribution
377 in bacteria also indicated that its metabolism is tightly linked with bacterial essential
378 activities. In sum, through this study, we obtained a much clearer picture about how energy
379 reserves are associated with certain types of bacteria. Further investigation via incorporating
380 bacterial physiology and lifestyle could supply much more feasible explanations to illustrate
381 these linkages, although experimental evidences are indispensable to confirm these
382 theoretical analyses.
383
384 **Acknowledgements**
385
386 The work was supported by Startup Foundation for Excellent Researchers at Xuzhou Medical
387 University [No. D2016007], Nature and Science Foundation for Colleges and Universities of
388 Jiangsu Province [No. 16KJB180028], Innovative and Entrepreneurial Talent Scheme of
389 Jiangsu Province [No. 53053312], and Nature and Science Foundation of Jiangsu Province
390 (2018)
391
392 **Author Contributions**
393
394 LW conceived the core idea of this study. LW, MJW, JY, YH, QL, and YX did all data
395 collection, data visualization, and statistical analysis.
396
397 **Declaration of Conflicting Interests**
398
399 The author declares that there is no conflict of interests.

400 **Table 1** Key enzymes and corresponding UniProt sequences used in this study for statistical modelling via HMMER package.

401

Reference Species	Gene	Enzyme	Length	UniProt ID
Haladaptatus paucihalophilus	*ppk1*	Polyphosphate kinase	707	E7QTB5
Thiomicrospira cyclica	*ppk2*	Polyphosphate kinase 2	264	F6DAB2
Metallosphaera sedula	*gppA*	Ppx/GppA phosphatase	420	A4YFE8
Haloferax massiliensis	*glgC*	Glucose-1-phosphate adenylyltransferase	322	A0A0D6JRD4
Pyrococcus abyssi	*glgA*	Glycogen synthase	437	Q9V2J8
Thermococcus kodakaraensis	*glgB*[*]	1,4-alpha-glucan branching enzyme (GH57)	675	Q5JDJ7
Escherichia coli	*glgB*	1,4-alpha-glucan branching enzyme (GH13)	728	P07762
Mycobacterium tuberculosis	*treS*	Trehalose synthase/amylase	601	P9WQ19
Mycobacterium tuberculosis	*pep2*	Maltokinase	455	Q7DAF6
Picrophilus torridus	*glgE*	Alpha-1,4-glucan: maltose-1-phosphate maltosyltransferase	630	Q6L2Z8
Mycobacterium tuberculosis	*Rv3032*	Glycogen synthase	414	P9WMY9
Haloferax mediterranei	*phaA*	Beta-ketothiolase	383	I3R3D1
Haloarcula hispanica	*phaB*	Acetoacetyl-CoA reductase	247	E1U2R6
Haloarcula hispanica	*phaC*	PHA synthase subunit C	474	G0HQZ6
Burkholderia pseudomallei	*phaZ*[&]	Intracellular polyhydroxyalkanoate depolymerase	423	Q3JIM5
Rhizobium fredii	*phaZ*	Extracellular polyhydroxyalkanoate depolymerase	369	G9AII6
Saccharomyces cerevisiae	*PDAT*[#]	Phospholipid: diacylglycerol acyltransferase	661	P40345
Acinetobacter baylyi	*wax-dgaT*	Wax Ester Synthase/Acyl Coenzyme A: Diacylglycerol Acyltransferase	458	Q8GGG1

Pseudomonas aeruginosa	*lip*	Triacylglycerol lipase	311	P26876

[*]Archaeal type of glycogen branching enzyme (encoded by *glgB* gene) belonging to GH57 class in CATH database. [&]Specialized intracellular PHA depolymerase. [#]TAG synthesis enzyme exclusively present in eukaryotes.

412 **References**

413

414 1. Zhu B, Ibrahim M, Cui Z, et al. Multi-omics analysis of niche specificity provides
415 new insights into ecological adaptation in bacteria. *The ISME Journal.*
416 2016;10(8):2072-2075.
417 2. Wang L, Yan J, Wise MJ, et al. Distribution Patterns of Polyphosphate Metabolism
418 Pathway and Its Relationships With Bacterial Durability and Virulence. *Frontiers in*
419 *Microbiology.* 2018;9.
420 3. Henrissat B, Deleury E, Coutinho PM. Glycogen metabolism loss: a common marker
421 of parasitic behaviour in bacteria? *Trends in Genetics.* 2002;18(9):437-440.
422 4. Wang L, Wise MJ. Glycogen with short average chain length enhances bacterial
423 durability. *Naturwissenschaften.* 2011;98(9):719-729.
424 5. Wilkinson J. The problem of energy-storage compounds in bacteria. *Experimental*
425 *Cell Research.* 1959;7:111-130.
426 6. Wang L, Liu Z, Dai S, Yan J, Wise MJ. The Sit-and-Wait Hypothesis in Bacterial
427 Pathogens: A Theoretical Study of Durability and Virulence. *Frontiers in*
428 *Microbiology.* 2017;8.
429 7. Zhang H, Ishige K, Kornberg A. A polyphosphate kinase (PPK2) widely conserved in
430 bacteria. *Proceedings of the National Academy of Sciences.* 2002;99(26):16678-
431 16683.
432 8. Whitehead MP, Eagles L, Hooley P, Brown MRW. Most bacteria synthesize
433 polyphosphate by unknown mechanisms. *Microbiology.* 2014;160(Pt_5):829-831.
434 9. Achbergerová L, Nahálka J. PPK1 and PPK2 — which polyphosphate kinase is older?
435 *Biologia.* 2014;69(3).
436 10. Kalscheuer R, Steinbüchel A. A Novel Bifunctional Wax Ester Synthase/Acyl-
437 CoA:Diacylglycerol Acyltransferase Mediates Wax Ester and Triacylglycerol
438 Biosynthesis inAcinetobacter calcoaceticusADP1. *Journal of Biological Chemistry.*
439 2003;278(10):8075-8082.
440 11. UniProt: the universal protein knowledgebase. *Nucleic Acids Research.*
441 2017;45(D1):D158-D169.
442 12. Pearson WR, Eddy SR. Accelerated Profile HMM Searches. *PLoS Computational*
443 *Biology*. 2011;7(10).
444 13. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Research.*
445 2011;40(D1):D136-D143.

14. Javed S, Azeem F, Hussain S, et al. Bacterial lipases: A review on purification and characterization. *Progress in Biophysics and Molecular Biology*. 2018;132:23-34.
15. Sagong H-Y, Son HF, Choi SY, Lee SY, Kim K-J. Structural Insights into Polyhydroxyalkanoates Biosynthesis. *Trends in Biochemical Sciences*. 2018.
16. Chandra G, Chater KF, Bornemann S. Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology*. 2011;157(6):1565-1572.
17. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*. 2004;32(Web Server):W20-W25.
18. Holm L, Sander C. Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics*. 1998;14(5):423-429.
19. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004;32(5):1792-1797.
20. Wise MJ. No so HoT - heads or tails is not able to reliably compare multiple sequence alignments. *Cladistics*. 2010;26(4):438-443.
21. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191.
22. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*. 2016;44(W1):W242-W245.
23. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH – a hierarchic classification of protein domain structures. *Structure*. 1997;5(8):1093-1109.
24. Moran NA. Microbial Minimalism. *Cell*. 2002;108(5):583-586.
25. Wolf YI, Koonin EV. Genome reduction as the dominant mode of evolution. *BioEssays*. 2013;35(9):829-837.
26. Pukall R. The Family Dietziaceae. *The Prokaryotes*2014:327-338.
27. Dahlqvist A, Stahl U, Lenman M, et al. Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences*. 2000;97(12):6487-6492.
28. Kämpfer P. Actinobacteria. *Handbook of Hydrocarbon and Lipid Microbiology*2010:1819-1838.

29. Marrakchi H, Lanéelle M-A, Daffé M. Mycolic Acids: Structures, Biosynthesis, and Beyond. *Chemistry & Biology*. 2014;21(1):67-85.

30. Funke G, Stubbs S, Altwegg M, Carlotti A, Collins MD. Turicella otitidis gen. nov., sp. nov., a Coryneform Bacterium Isolated from Patients with Otitis Media. *International Journal of Systematic Bacteriology*. 1994;44(2):270-273.

31. Galperin MY. Genome Diversity of Spore-Forming Firmicutes. *Microbiology Spectrum*. 2013;1(2).

32. Verlinden RAJ, Hill DJ, Kenward MA, Williams CD, Radecka I. Bacterial synthesis of biodegradable polyhydroxyalkanoates. *Journal of Applied Microbiology*. 2007;102(6):1437-1449.

33. Uchino K, Saito T, Gebauer B, Jendrossek D. Isolated Poly(3-Hydroxybutyrate) (PHB) Granules Are Complex Bacterial Organelles Catalyzing Formation of PHB from Acetyl Coenzyme A (CoA) and Degradation of PHB to Acetyl-CoA. *Journal of Bacteriology*. 2007;189(22):8250-8256.

34. Handrick R, Reinhardt S, Kimmig P, Jendrossek D. The "Intracellular" Poly(3-Hydroxybutyrate) (PHB) Depolymerase of Rhodospirillum rubrum Is a Periplasm-Located Protein with Specificity for Native PHB and with Structural Similarity to Extracellular PHB Depolymerases. *Journal of Bacteriology*. 2004;186(21):7243-7253.

35. Ventura M, Canchaya C, Tauch A, et al. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews*. 2007;71(3):495-548.

36. Manganelli R, García-Hidalgo J, Hormigo D, Arroyo M, de la Mata I. Novel Extracellular PHB Depolymerase from Streptomyces ascomycinicus: PHB Copolymers Degradation in Acidic Conditions. *PLoS ONE*. 2013;8(8).

37. Singh G, Kumari A, Mittal A, Yadav A, Aggarwal NK. Polyβ-Hydroxybutyrate Production byBacillus subtilisNG220 Using Sugar Industry Waste Water. *BioMed Research International*. 2013;2013:1-10.

38. Gojobori T, Grosjean H, Breton M, et al. Predicting the Minimal Translation Apparatus: Lessons from the Reductive Evolution of Mollicutes. *PLoS Genetics*. 2014;10(5).

39. Finkelstein DB, Brassell SC, Pratt LM. Microbial biosynthesis of wax esters during desiccation: Adaptation for colonization of the earliest terrestrial environments? *Geology*. 2010;38(3):247-250.

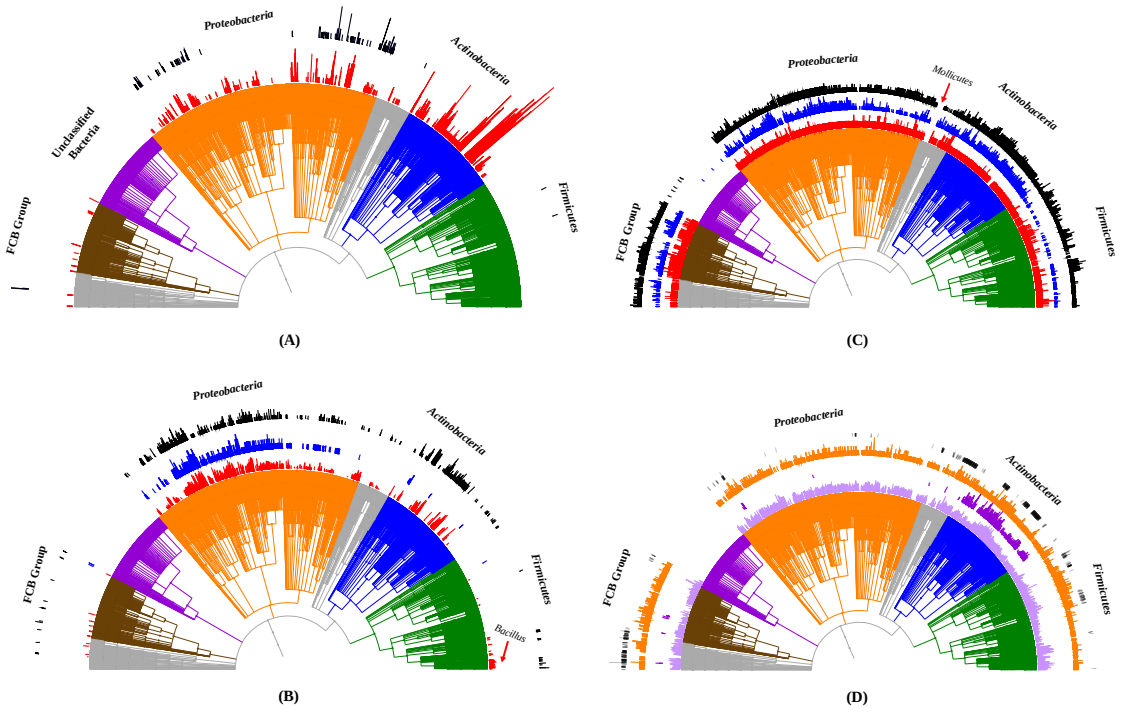

Figure 1 Evolutionary analysis of metabolic pathway distributions of energy reserves in 8214 bacterial species. Four groups of bacteria are highlighted, which are *Firmicutes* (green), *Actinobacteria* (red), *Proteobacteria* (orange), Unclassified Bacteria (violet), and FCB group (brown). **(A)** Distribution patterns of key enzyme WS/DGAT (red bar) and Triacylglycerol lipase Lip (black bar) along the evolutionary tree that is responsible for the synthesis and utilization of neutral lipids in bacteria. WS/DGAT is abundant in *Actinobacteria* with multiple homologs while *Firmicutes* do not have any WS/DGAT-harboring organisms. *Proteobacteria* has moderate number of bacteria with WS/DGAT enzymes. For FCB group and unclassified bacteria, WS/DGAT is only sporadically present. As for the utilization of neutral lipids, the enzyme Lip is almost exclusively restricted to the phylum *Proteobacteria* (β-, γ-, δ, and ε-). Only one organisms belonging to α-proteobacteria has Lip homologous enzyme. **(B)** Distribution patterns of PHB synthesis pathway (PhaA, PhaB and PhaC) in red bar and degradation enzymes intracellular PhaZ (blue bar) and extracellular PhaZ (black bar) along the evolutionary tree. For FCB group and unclassified bacteria, almost no PHB metabolism exists. *Proteobacteria* are abundant in terms of PHB synthesis pathway and the two types of degradative enzymes while *Actinobacteria* have moderate level of synthesis ability and extracellular PhaZ. The intracellular PhaZ is largely missing in *Actinobacteria*. As

for *Firmicutes*, the PHB metabolism mainly distributes in the genus of *Bacillus* (red arrow). **(C)** Distribution patterns of key enzyme PPK1 (red bar), PPK2 (blue bar), and PPX (black bar) along the evolutionary pathway that is responsible for main synthesis and degradation pathways of polyphosphate in bacteria. Polyphosphate metabolism is widely distributed in bacteria. Unclassified bacteria and the class of *Mollicutes* (red arrow) belonging to phylum *Tenericutes* have apparent loss of polyphosphate metabolism in the tree, although data analysis showed 2215 bacterial species do not have PPK1, PPK2 and PPX. **(D)** Distribution patterns of two glycogen synthesis related pathways and two glycogen branching enzymes (GH13 GlgB and GH57 GlgB) along the evolutionary tree. GlgB is responsible for the highly branched structure of glycogen in bacteria. Full glycogen synthesis pathway includes GlgC, GlgA and GlgB (light violet bar), which distributes widely in bacteria. Trehalose-based glycogen synthesis pathway includes TreS, Pep2, GlgE, GlgB (dark violet bar) and is mainly concentrated in the *Actinobacteria* phylum. Distributions of the two GlgB enzymes show that GH57 GlgB (black bar) is mainly present in *Terrabacteria* group. In addition, GH57 GlgB is also identified in PVC group, *Spirochaetes*, *Acidobacteria*, *Fusobacteria*, *Thermotogae*, *Nitrospirae*, *Aquificae*, *Synergistetes*, *Elusimicrobia*, *Nitrospinae/Tectomicrobia* group, *Thermodesulfobacteria*, *Rhodothermaeota*, and *Dictyoglomi*. In contrast, GH13 GlgB is widely distributed in bacterial species.