

1-1-2019

## **Bioinformatics analysis of metabolism pathways of archaeal energy reserves**

Liang Wang

Qinghua Liu

Xiang Wu

Yue Huang

Michael J. Wise

*See next page for additional authors*

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworkspost2013>



Part of the [Biology Commons](#)

---

10.1038/s41598-018-37768-0

Wang, L., Liu, Q., Wu, X., Huang, Y., Wise, M. J., Liu, Z., . . . Wang, C. (2019). Bioinformatics analysis of metabolism pathways of archaeal energy reserves. *Scientific Reports*, 9(1). Available [here](#)

This Journal Article is posted at Research Online.

<https://ro.ecu.edu.au/ecuworkspost2013/5822>

---

**Authors**

Liang Wang, Qinghua Liu, Xiang Wu, Yue Huang, Michael J. Wise, Zhanzhong Liu, Wei Wang, Junfeng Hu, and Chunying Wang

# SCIENTIFIC REPORTS



OPEN

## Bioinformatics Analysis of Metabolism Pathways of Archaeal Energy Reserves

Liang Wang<sup>1,2</sup>, Qinghua Liu<sup>2</sup>, Xiang Wu<sup>1</sup>, Yue Huang<sup>1</sup>, Michael J. Wise<sup>3,4</sup>, Zhanzhong Liu<sup>5</sup>, Wei Wang<sup>1,6,7</sup>, Junfeng Hu<sup>1,8</sup> & Chunying Wang<sup>5</sup>

Energy storage compounds play crucial roles in prokaryotic physiology. Five chemical compounds have been identified in prokaryotes as energy reserves: polyphosphate (polyP), polyhydroxyalkanoates (PHAs), glycogen, wax ester (WE) and triacylglycerol (TAG). Currently, no systematic study of archaeal energy storage metabolism exists. In this study, we collected 427 archaeal reference sequences from UniProt database. A thorough pathway screening of energy reserves led to an overview of distribution patterns of energy metabolism in archaea. We also explored how energy metabolism might have impact on archaeal extremophilic phenotypes. Based on the systematic analyses of archaeal proteomes, we confirmed that metabolism pathways of polyP, PHAs and glycogen are present in archaea, but TAG and WE are completely absent. It was also confirmed that PHAs are tightly related to halophilic archaea with larger proteome size and higher GC contents, while polyP is mainly present in methanogens. In sum, this study systematically investigates energy storage metabolism in archaea and provides a clear correlation between energy metabolism and the ability to survive in extreme environments. With more genomic editing tools developed for archaea and molecular mechanisms unravelled for energy storage metabolisms (ESMs), there will be a better understanding of the unique lifestyle of archaea in extreme environments.

Energy storage compounds are important for prokaryotic physiological activities such as regulatory signalling, intracellular persistence, pathogenicity and environmental long-term survival<sup>1</sup>. Wilkinson<sup>2</sup> first proposed three pre-requisites for defining an energy storage compound, according to which, five energy storage compounds have been confirmed to exist, that is, polyphosphate (polyP), glycogen, and storage lipids such as polyhydroxyalkanoates (PHAs), triacylglycerol (TAG), and wax ester (WE)<sup>3</sup>. These are considered as energy storage mechanisms (ESMs)<sup>1</sup>. Distribution patterns of metabolism pathways of the five energy storage compounds were previously investigated in bacteria<sup>1</sup>. However, currently there is no study focusing on how pathways of energy storage compounds are distributed in archaea. Considering that many archaea are extremophiles, theoretical study of ESMs via pathway distributions would be beneficial to dissect archaeal extremophilic physiology, hence a better understanding of life's adaptations to extreme environments<sup>4</sup>. An introduction to the five energy storage compounds in prokaryotes is briefly summarized below.

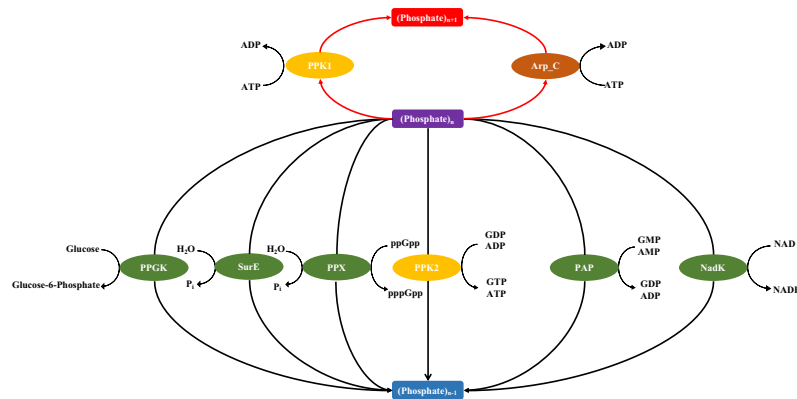
### Polyphosphate

Although it is known that polyP is ubiquitous from bacteria to mammals, its metabolism pathway has not been fully elucidated so far and its accumulation is only identified sporadically in archaea through experimental study<sup>5-7</sup>. Beyond its roles as a phosphorus resource and energy reserve, polyP is currently also linked to a set of physiological functions in extremophilic archaea, such as metal resistance, salt tolerance, oxidative stress adaptation,

<sup>1</sup>Department of Bioinformatics, School of Medical Informatics, Xuzhou Medical University, Xuzhou, Jiangsu, China.

<sup>2</sup>Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, School of Pharmacy, Xuzhou Medical University, Xuzhou, Jiangsu, China. <sup>3</sup>The Marshall Centre for Infectious Diseases Research and Training, University of Western Australia, Perth, Western Australia, Australia. <sup>4</sup>Department of Computer Science and Software Engineering, School of Physics, Mathematics and Computing, University of Western Australia, Perth, Western Australia, Australia.

<sup>5</sup>Xuzhou Infectious Diseases Hospital, Xuzhou, Jiangsu, China. <sup>6</sup>School of Public Health, Capital Medical University, Beijing, China. <sup>7</sup>School of Medical Sciences, Edith Cowan University, Perth, WA, Australia. <sup>8</sup>Department of Computer Science, School of Medical Informatics, Xuzhou Medical University, Xuzhou, Jiangsu, China. Correspondence and requests for materials should be addressed to L.W. (email: [leonwang@xzhmu.edu.cn](mailto:leonwang@xzhmu.edu.cn))



**Figure 1.** Illustration of polyP metabolism and pathway distributions in archaea. (a) Red arrows represent synthesis pathways (PPK1 and Arp\_C) that convert (phosphate)<sub>n</sub> to (phosphate)<sub>n+1</sub> while deep blue arrows represent independent degradation pathways (PPGK, SurE, PPX, PAP, PPK2 and NadK) that break down (phosphate)<sub>n</sub> into (phosphate)<sub>n-1</sub>. Arp\_C is an abbreviation of actin protein complex of ArpABCEFGH that shares the same Pfam domain actin (PF00022)<sup>8</sup>. (b) Eight polyP metabolism pathways and nine archaeal physiology types are presented. Full\_Path (7 archaea) encompasses polyP-related six enzymes while No\_Path (64 archaea) have lost all six enzymes from archaeal proteomes. Functional\_Path (99 archaea) means archaea having PPK1 and at least one polyP-degrading enzyme. PPK1\_PPK2 (18 archaea), PPK1\_PAP (30 archaea), PPK1\_SurE (88 archaea), PPK1\_PPX (39 archaea), and PPK2\_NADK (78 archaea) means archaea having PPK1 and the designated polyP-degrading enzyme. For each pathway, average proteome size (APS) and GC content were given at the bottom of each boxplot.

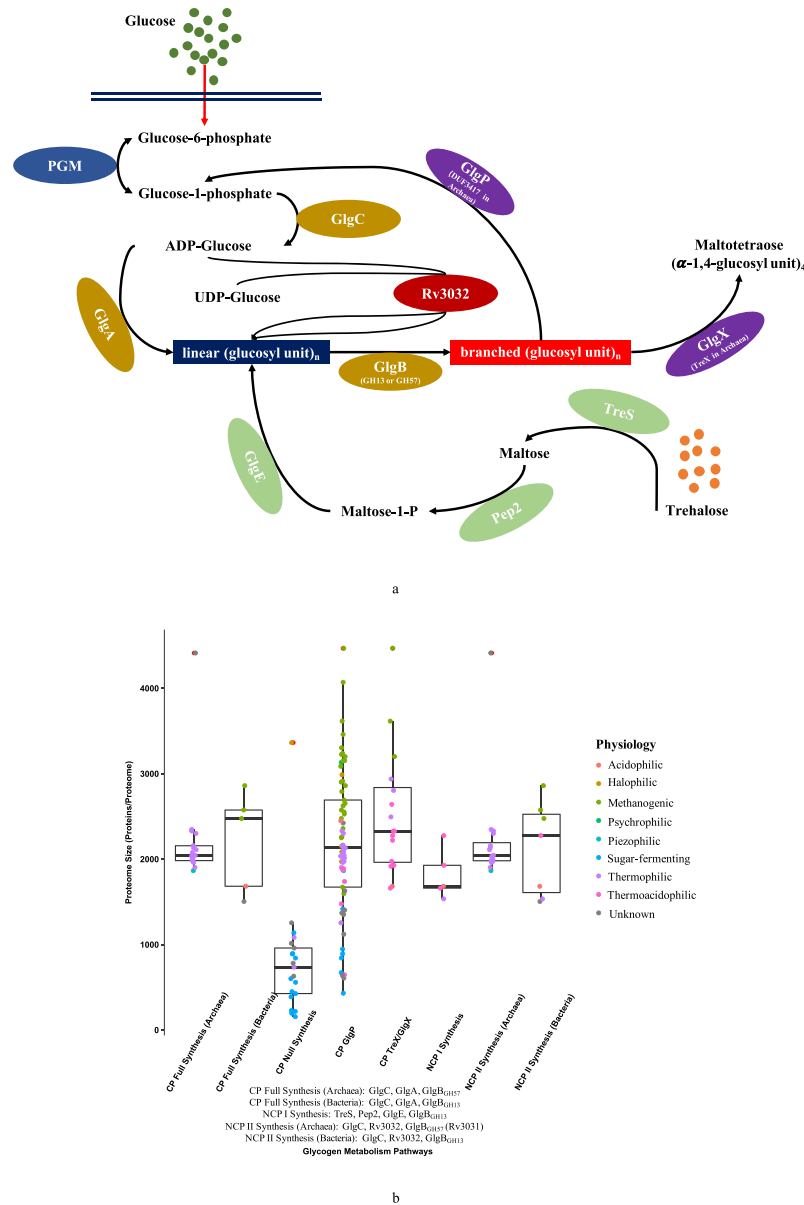
temperature tolerance, and other environmental stresses<sup>6</sup>. Thus, polyP plays important roles for archaeal physiological adaption to environmental changes and stress conditions where archaea reside. In bacteria, enzymes directly linked with polyP metabolism include polyphosphate kinase (PPK1), polyphosphate kinase 2 (PPK2), polyphosphate:AMP phosphotransferase (PAP), 5'/3'-Nucleotidase (SurE), exopolyphosphatase (PPX), NAD kinase and polyphosphate glucokinase<sup>8</sup>. Detailed functions for each enzyme is listed in a previous study by Wang *et al.*<sup>8</sup> A recent phylogenetic study revealed that PPK1 and PPK2 are less common in bacteria and other unknown enzymes may involve in polyP metabolism<sup>7</sup>. In addition, evolutionary analysis showed that PPK2 evolves earlier than PPK1, matching with Arthur Kornberg's theory that polyP utilization is much older than synthesis<sup>7</sup>. There is also an actin-related protein complex encoded by *arpABCEFGH* that may fulfil the role of the unknown mechanism for polyP synthesis in bacteria<sup>9</sup>. Although it was proposed that enzymes involved in polyP metabolism show structure conservation among bacteria and archaea, only two enzymes PPX and PPK were analysed in archaea through comparative genomics and there is no overview of a complete polyP metabolism in archaea (Fig. 1)<sup>6</sup>.

## Glycogen

Carbohydrate polymers and oligomers are stored by living cells for a variety of purposes, such as energy storage and stress resistance. Glycogen is one of the most widely spread carbohydrates that have been identified in archaea, bacteria and eukaryotes<sup>10</sup>. The classical pathway (CP) of bacterial glycogen metabolism includes five essential enzymes that are ADP-glucose pyrophosphorylase (GlgC), glycogen synthase (GlgA), glycogen branching enzyme (GlgB), glycogen phosphorylase (GlgP), and glycogen debranching enzyme (GlgX)<sup>1</sup>. A second glycogen synthesis pathway (non-classical pathway I, abbreviated as NCP I) was recently identified. It involves four enzymes TreS, Pep2, GlgE and GlgB, which correlates trehalose metabolism with glycogen metabolism<sup>11</sup>. Another Rv3032-centric pathway (non-classical pathway II, abbreviated as NCP II) is responsible for lipopolysaccharide and putative glycogen biosynthesis in *Mycobacteria* spp<sup>12</sup>. A brief summary of glycogen metabolism pathways is illustrated in Fig. 2. Previous analysis showed that most archaeal species lack the genes for the classical glycogen pathway, and there is only one archaeal species *Picrophilus torridus* DSM 9790 with all four enzymes in NCP I pathway through BLAST search<sup>11</sup>.

## Storage Lipids

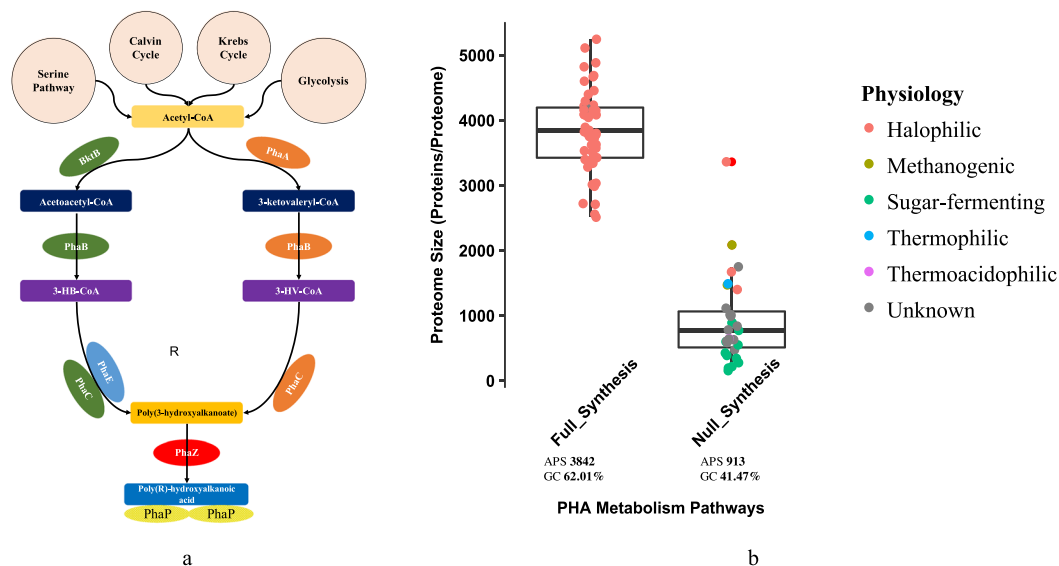
There are three common storage lipids in bacteria and archaea: polyhydroxyalkanoates (PHAs) (Fig. 3), wax ester (WE) (Fig. 4), and triacylglycerol (TAG) (Fig. 5). PHAs are a group of naturally-occurring bio-polyesters accumulating in prokaryotes as major energy and carbon sources. They are composed of (*R*)-hydroxy fatty acids and are classified into three categories according to the number of units: short chain length (SCL) PHA, medium chain length (MCL) PHA and long chain length (LCL) PHA. According to the functional alkyl *R* group, PHAs are assigned with specific names and carbon numbers, such as poly-β-hydroxybutyrate (C4) and poly-β-hydroxyvalerate (C5)<sup>13</sup>. Among the family of PHAs, the best known is poly-β-hydroxybutyrate (PHB). Recent studies supported that archaea mainly store PHB and PHV intracellularly<sup>14</sup>. In addition, a mixed PHA consisting of PHB and PHV is also reported in archaea<sup>15</sup>. As for chain length, archaea normally accumulate SCL PHA<sup>16</sup>. A variety of studies confirmed that PHA metabolism involves PhaA, PhaB and PhaC in the synthesis pathway and PhaZ in the degradation pathway (Fig. 3). In addition, synthase subunit PhaE and PHA granule associated protein PhaP are also a part of the synthesis mechanism<sup>14,17</sup>. A novel bifunctional enzyme wax ester



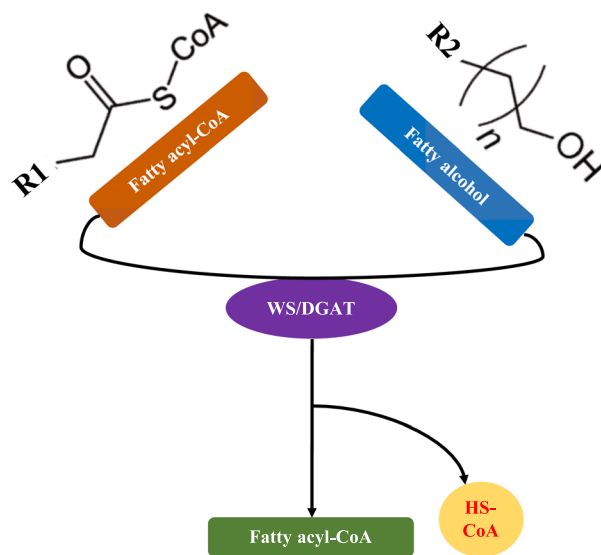
**Figure 2.** Illustration of glycogen metabolism and pathway distributions in archaea. **(a)** Three pathways are present: (1) classical pathway (CP) of glycogen metabolism (synthesis enzymes GlgC, GlgA and GlgB in dark yellow and degradation enzymes GlgP and GlgX in purple) that is widely distributed in bacterial species; (2) non-classical pathway I (NCP I) including TreS, Pep2, GlgE and GlgB, which links trehalose and maltose metabolism with glycogen; (3) non-classical pathway II (NCP II) including GlgC, Rv3032, and GlgB, which is responsible for glycogen formation and synthesis of 6-O-methylglucosyl-containing lipopolysaccharides (MGLP), a pathogenic factor. PGM is phosphoglucomutase transforming glucose-6-phosphate into glucose-1-phosphate for glycogenesis. **(b)** Distribution of glycogen metabolism pathways in 427 archaeal species. Eight groups of enzymes are formulated, which are CP Full Synthesis (Archaea), CP Full Synthesis (Bacteria), CP Null Synthesis, CP GlgP, CP TreX/GlgX, NCP I Synthesis, NCP I Synthesis I (Archaea), NCP II Synthesis (Archaea), NCP II Synthesis (Bacteria). Enzymes in five major groups were detailed at the bottom of the figure.

synthase/acyl-CoA: diacylglycerol acyltransferase (WS/DGAT) discovered initially in *Acinetobacter calcoaceticus* ADP1 was shown to be present in numerous Gram-positive and Gram-negative bacteria, hence confirmation of wide spread distribution of WE and TAG storage lipids in bacteria<sup>18,19</sup>. It was previously proposed that lipid droplets in archaea accumulated PHA exclusively, while the other two storage lipids, wax ester and triacylglycerol, only arose in bacterial lineages<sup>4</sup>. Initial BLAST search found out 18 homologous WS/DGAT enzymes in archaeal genomes<sup>18</sup>. However, a subsequent analysis only saw a WS/DGAT homologue in one archaeon, which raised the question about how WS/DGAT is distributed in archaea<sup>19</sup>.

In this study, we collected as seed sequences all the enzymes directly linked with metabolism of five energy storage compounds in archaea. Hidden Markov models of all seed proteins were constructed *ab initio* by following

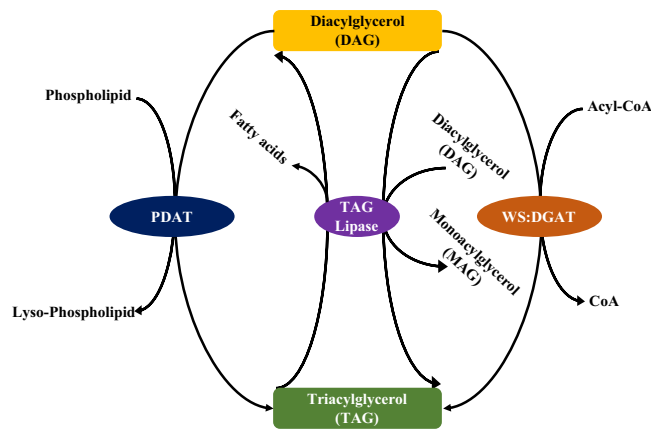


**Figure 3.** Illustration of classical PHAs metabolism and pathway distributions. **(a)** There are currently four types of PHA synthase system identified. Type I, III, and IV are related with SCL PHAs. Type I has a single PhaC, Type III has a complex of PhaC and PhaE while Type IV is recently identified that consists of PhaC and PhaR. In this study, Type I and Type III synthase systems are considered. The full metabolism pathway includes enzymes PhaA, PhaB, PhaC, PhaE, PhaP, and PhaZ. Since BktB and PhaA share the same protein domain organizations (PF00108 and PF02803) that consist of Thiolase-N and Thiolase-C, only PhaA was studied in this study in order to avoid redundancy. **(b)** Distribution of PHA metabolism pathways in 427 archaeal species. Only synthesis pathways with complete and null enzymes are analysed. Full\_Synthesis (51 archaea) group includes PhaA, PhaB, PhaC, PhaE, and PhaP. No\_Synthesis (31) group does not have any of these enzymes. For each pathway, average proteome size (APS) and GC content were given at the bottom of each boxplot.



**Figure 4.** Illustration of wax ester synthesis pathway in prokaryotes. The bifunctional enzyme wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase (WS/DGAT) that is responsible for the synthesis of both TAG and WE, is emphasized in this study.

standard procedures<sup>3,8</sup>. A total of 427 archaeal reference proteomes were analysed for pathway distribution. A variety of correlational analyses in terms of pathway distribution patterns, proteome size, and archaeal physiology were performed. This study is the first comparative genomic study of energy storage compounds in archaea. The significance of the study resides in identification of the linkages between energy storage compounds and archaeal extremophilic phenotypes, which will promote our understanding of archaeal adaptation to extreme environments. Finally, a set of archaeal enzymes involving ESMs are also discovered, which may have potential benefits for industrial and clinical applications in future studies.



**Figure 5.** Illustration of TAG metabolism pathway. Phospholipid:diacylglycerol acyltransferase (PDAT) is only present in yeast and plants. Thus, it is not considered in this study. Triacylglycerol lipase is sourced from *Pseudomonas aeruginosa* that reversibly converts Triacylglycerol and Diacylglycerol. WS/DGAT is the same enzyme as in WE synthesis.

## Results and Discussion

HMM-based screening of archaeal proteomes based on selected seed enzymes that are directly linked to ESMs generated a complete map of pathway distributions (Supplementary Table S1). In addition, literature and database mining (UniProt<sup>20</sup> and PATRIC<sup>21</sup>) classified 427 archaeal species into three predominant sub-categories, that is, halophiles, thermophiles and methanogens, *etc.*, which is consistent with previous reports (Supplementary Table S2)<sup>22</sup>. Distribution patterns of pathways of all five ESMs were analysed in the 427 archaeal species. Phylogenetic analysis was also incorporated into the study in order to have a grouped overview of energy storage metabolisms in archaea.

**Bioinformatics Analysis of PolyP Distribution in Archaea.** PolyP has been correlated with prokaryotic adaptation to extreme environments, such as heavy metal and high salinity resistances, *etc.*<sup>6,23–25</sup>. A number of archaeal species, such as *Sulfolobus metallicus* and *Metallosphaera sedula*, were reported to store polyP intracellularly<sup>26</sup>. In this study, a set of eight enzymes that are directly involved in polyP metabolism in prokaryotes was scrutinized in archaeal proteomes. Polyphosphate glucokinase (PPGK) and actin-related protein complex (Arp\_C) genes were excluded because only a couple of species have them (data not shown).

As for the six polyP-metabolism related enzymes, PPK1 is responsible for polyP synthesis while the other five enzymes are in the degradation pathway (Fig. 1). Previous study found that PPK1 is widely conserved in bacteria and archaea<sup>27</sup>. In this archaeal study, we observed the co-occurrence of PPK1 and PPK2 (also PAP, a homolog of PPK2 with duplicated domains PF03976) (Supplementary Table S1)<sup>28</sup>, which is consistent with previous bacterial analysis<sup>8,9</sup>. However, PPK1 is only moderately present (99 out of 427 archaeal species) and the combined presence of PPK1 and PPK2 are even more rare in archaea (18 out of 427 archaeal species compared with 504 out of 944 bacterial species)<sup>8</sup>. Technically, coupling PPK1 and any one of the five enzymes (PPK2, PAP, SurE, PPX, and NADK) in archaea could be able to synthesize and degrade polyP (Fig. 1). Thus, a total of eight polyP metabolism groups were analysed, which include Full\_Path group (7 archaea) encompassing polyP-related six enzymes (PPK1, PPK2, PAP, SurE, PPX, and NADK), No\_Path group (64 archaea) involving species with none of the six enzymes. Functional\_Path group (99 archaea) means archaea having PPK1 and at least one polyP-degrading enzyme (PPK2, PAP, SurE, PPX, or NADK). PPK1\_PPK2 (18 archaea), PPK1\_PAP (30 archaea), PPK1\_SurE (88 archaea), PPK1\_PPX (39 archaea), and PPK2\_NADK (78 archaea) means archaea having PPK1 and the designated polyP-degrading enzyme. Proteome size and GC content were also incorporated into polyP pathway distribution analyses (Fig. 1).

According to the results, presence of Full\_Path group may suggest an ancient origin of polyphosphate. In the Full\_Path group, the three archaeal genera, *Methanobolus*, *Methanoregula*, and *Methanosarcina*, have complete polyP pathways. It was reported that cells grown on methanol accumulated more phosphate than autotrophically grown cells<sup>29</sup>. Polyphosphate has been frequently observed in *Methanosarcina*, and high concentrations of orthophosphates are associated with acetotrophic methanogenesis<sup>29–31</sup>. In addition, air-adapted *Methanosarcina acetivorans* has both enhanced methane production and polyP accumulation<sup>32</sup>. However, there are no definitive physiological roles for polyP in archaea for methane production<sup>29</sup>. On the other hand, much experimental evidence links polyphosphate with adaptation to extreme environments, such as hyperthermal, heavy metal concentration and hyper-salinity, *etc.*, which suggests potential roles of polyP in methanogens as a contributor to multiple extremophilic phenotypes<sup>6,24,33,34</sup>. As for high salinity resistance, phosphate transporter operon (*pst*) was found to be highly induced under high salt concentration in *Methanosarcina mazei*<sup>35</sup>. Together with polyphosphate, they might serve as counterions for potassium ions, hence increasing resistance to high salinity conditions<sup>35</sup>. In the presence of copper or other heavy metals, polyP can be rapidly degraded into phosphate in species like *Sulfolobus metallicus* by exopolyphosphatase, the first polyP enzyme identified in archaea<sup>26</sup>. The phosphate then forms metal-phosphate complexes and is excreted through inorganic phosphate transport system (Pit)<sup>34,36</sup>. In addition, NAD kinase and PPX may also serve a similar role for heavy metal resistance in *Metallosphaera sedula*<sup>36</sup>.

64 out of 427 archaea (No\_Path group) were found having none of the six enzymes. For the previous bacteria study, polyP metabolism loss was linked with reduced proteome size and could be an indicator for obligate symbiotic, parasitic or commensal (SPC) lifestyle<sup>8,37</sup>. Reduced proteome was also observed in archaeal No\_Path group (Fig. 1). However, archaea are almost exclusively free-living species and SPC archaea are rarely reported<sup>38</sup>. Thus, their proteome reduction could not be interpreted by genetic drift with increased non-coding genes and loss of biosynthetic pathways<sup>37</sup>. Loss of polyP enzymes or the whole pathway could be caused by minimization of cell complexity under extreme environments, where polyP metabolism might not be necessary<sup>37</sup>. As for the functional pathway groups, it is interesting to notice that all species with PPK1 are coupled with at least one polyP degrading enzyme. In addition, proteome sizes and GC contents of the PPK1\_PPK2, PPK2\_SurE and PPK1\_NADK groups are similar (Fig. 1). In contrast, PPK1\_PAP and PPK1\_PPX are distinct due to their comparatively smaller genome size and lower GC content (Fig. 1). In addition, PAP (polyphosphate: AMP Phosphotransferase) and PPX (exopolyphosphatase) seem to be more closely associated with methanogenic archaea than other enzymes.

Intriguingly, some archaeal species that are methanogenic, halophilic or thermophilic such as *Metallosphaera sedula* and *Sulfolobus sp.* have been reported to accumulate polyP experimentally<sup>26</sup>. However, enzyme screening via HMM-based sequence models did not identify polyP synthesis enzyme PPK1 (Supplementary Table S1). In addition, it was also reported that polyphosphate synthesis enzymes have not yet been described for *Crenarchaeota*, neither, though polyP granules were found in the cytoplasm and could be rapidly hydrolysed to inorganic phosphate<sup>39</sup>. It has been suggested that there might be an unknown archaeal PPK enzyme<sup>40</sup>. This is probably the reason that HMM-based sequence search cannot find homologs. Thus, further experimental studies should be performed to identify the novel enzyme(s). In addition, polyP synthesis pathways seem to be mainly restricted to halophiles and methanogens. Considering that horizontal gene transfer (HGT) from bacteria to halophiles and methanogens is well known<sup>41</sup>, it could be possible that these archaea acquired the respective genes related to polyP metabolism from bacteria whereas other archaea might use so far unknown pathways.

**Bioinformatics Analysis of Glycogen Distribution in Archaea.** Presence of glycogen has been reported decades ago in some archaeal genera as *Thermococcus*, *Sulfolobus*, *Thermoproteus*, and *Desulfurococcus*<sup>42</sup>. In addition, *Methanogens* such as *Methanosarcina thermophila* were also found to accumulate glycogen<sup>43</sup>. Although some species belonging to the above-mentioned archaeal genera were reported to accumulate glycogen, it doesn't mean that all species in these archaeal groups should have the same feature. In addition, we only studied 427 reference archaeal proteomes, it is not supposed to cover all species in archaea. So far, three representative glycogen metabolism pathways have been proposed, which are classical pathway (CP)<sup>44</sup>, non-classical (NCP I) trehalose-related pathway<sup>45</sup> and non-classical (NCP II) liposaccharide-related pathway (Fig. 2)<sup>46</sup>. For CP, there are three archaea-specific enzymes corresponding to their bacterial counterparts, which are archaeal GlgB, GlgP and GlgX (or TreX)<sup>47</sup>. Archaeal GlgB and GlgP are vastly different from bacterial GlgB and GlgP according to HMM-based Pfam analysis (Table 1)<sup>47,48</sup>. Although GlgX and TreX are homologous with high similarity, their actual functions are rather divergent<sup>49</sup>. HMM-based sequence search showed that they share an identical distribution pattern in selected archaeal proteomes (Supplementary Table S1). Thus, at sequence level we cannot distinguish TreX and GlgX. As for the NCP I, only one archaeal species *Picrophilus torridus* DSM 9790 was previously identified to harbour all four enzyme TreS, Pep2, GlgE and GlgB, which was also confirmed not to be caused by horizontal gene transfer<sup>11</sup>. Finally, Rv3032, the essential enzyme in the other non-classical alpha-glucan metabolism pathway, is responsible for the synthesis of linear alpha-1,4-glycosidic chains<sup>50</sup>.

Glycogen metabolism distribution in archaea is vastly different from that in bacteria. Among 1202 studied bacterial proteomes, we observed that 402 strains (245 species) have a complete set of essential enzymes (GlgC, GlgA, GlgB, GlgP, and GlgX) for glycogen metabolism<sup>1</sup>. Our analysis showed that no archaeon harbours the complete classic pathway (CP) of glycogen metabolism with all five essential enzymes. Thus, we focused only on glycogen synthesis pathways. In this study, we divided glycogen metabolism into eight groups, which are CP Full Synthesis (Archaea), CP Full Synthesis (Bacteria), CP Null Synthesis, CP GlgP, CP TreX/GlgX, NCP I Synthesis, NCP II Synthesis (Archaea), NCP II Synthesis (Bacteria) (Fig. 2). CP Full Synthesis is a pathway with GlgC, GlgA and GlgB in an archaeal proteome. However, because there are two types of GlgB, we used archaea and bacteria to distinguish the two pathways. From our study, archaeal GlgB (Q5JDJ7) and bacterial GlgB (P07762) do not occur in the same archaeal proteome, and CP Full Synthesis (Archaea) pathway (average proteome size 2214, GC content 47.24%) is closely linked with thermophiles. As for the 17 archaea in the CP Full Synthesis (Archaea) group, most of which belongs to *Thermococci* class. As for physiological traits, they all are thermophilic archaea except for Candidatus *Lokiarchaeota* archaeon CR\_4 whose lifestyle is not clear. On the other hand, only five species possess CP Full Synthesis (Bacteria) pathway (average proteome size 2220, GC content 48.36%), three of which are methanogenic. In fact, both methanogenic and thermophilic archaea have been reported to accumulate glycogen<sup>51</sup>. Recently, central carbon mechanism in *Methanosarcina acetivorans* was systematically investigated in order to optimize methane production<sup>52</sup>. Glycogen was experimentally identified and its biological function was considered to be an environmental advantage for *Methanosarcinales* when carbon sources are scarce<sup>52</sup>. However, no halophilic archaea were reported to store glycogen by far. On the other hand, most of the archaea in the CP Null Synthesis group possess the sugar-fermenting trait, though a couple of thermophilic archaea are also present. All the sugar-fermenting archaea in this study belong to the candidate division MSBL1 from the unclassified Euryarchaeota. It was suggested that MSBL1 archaea could ferment glucose via the Embden–Meyerhof–Parnas pathway and were also capable of autotrophic growth when glucose and other fermentable sugars are not available<sup>53</sup>. Thus, there might be no need for these archaea to store glycogen as energy reserve.

There are another two non-classical pathways (NCP) for glycogen synthesis. Although only one archaeon was previously reported to have the complete NCP I pathway, in this study, more species were identified to use the pathway to synthesize alpha-glucan, such as *Acidiplasma cupricumulans* BH2 and *Thermoplasmatales* archaeon, etc. Considering that large number of archaeal genomes are yet to be discovered, it could be possible



	Reference Species	D <sup>#</sup>	Gene	Protein	Enzyme	L <sup>*</sup>	UPID <sup>*</sup>	PID <sup>*</sup>
Polyphosphate (polyP)	<i>Haladaptatus paucihalophilus</i>	A	<i>ppk1</i>	PPK1	Polyphosphate kinase	707	E7QTB5	PF13089 PF02503 PF13090
	<i>Thiomicrospira cyclica</i>	A	<i>ppk2</i>	PPK2	Polyphosphate kinase 2	264	F6DAB2	PF03976
	<i>Methanohalophilus mahii</i>	A	<i>pap</i>	PAP	Polyphosphate: AMP phosphotransferase	503	D5EAS0	PF03976 PF03976
	<i>Haloarcula marismortui</i>	A	<i>surE</i>	SurE	5'-nucleotidase	269	Q5V4L7	PF01975
	<i>Metallosphaera sedula</i>	A	<i>gppA</i>	GPPA	Ppx/GppA phosphatase	420	A4YFE8	PF02541
	<i>Pyrococcus horikoshii</i>	A	<i>ppnK</i>	PPNK	NAD kinase	277	O58801	PF01513
	<i>Mycobacterium tuberculosis</i>	B	<i>ppgK</i>	PPGK	Polyphosphate glucokinase	265	P9WIN1	PF00480
	<i>Dictyostelium discoideum</i>	E	<i>arp</i>	Arp_C	Actin-related protein (Complex)	/	/	PF00022
Glycogen	<i>Haloferax massiliensis</i>	A	<i>glgC</i>	GlgC	Glucose-1-phosphate adenylyltransferase	322	A0A0D6JRD4	PF00483
	<i>Pyrococcus abyssi</i>	A	<i>glgA</i>	GlgA	Glycogen synthase	437	Q9V2J8	PF08323 PF00534
	<i>Thermococcus kodakaraensis</i>	A	<i>glgB</i>	GlgB	1,4-alpha-glucan branching enzyme (GH57)	675	Q5JDJ7	PF03065 PF09210 PF14520
	<i>Escherichia coli</i>	B	<i>glgB</i>	GlgB	1,4-alpha-glucan branching enzyme (GH13)	728	P07762	PF02922 PF00128 PF02806
	<i>Thermococcus gammatolerans</i>	A	<i>glgP</i>	GlgP	Alpha-glucan phosphorylase (DUF3417)	826	C5A1K5	PF11897 PF00343
	<i>Escherichia coli</i>	B	<i>glgP</i>	GlgP	Alpha-glucan phosphorylase	815	P0AC86	PF00343
	<i>Sulfolobus solfataricus</i>	A	<i>treX</i>	TreX	Glycogen debranching enzyme	718	Q7LX99	PF02922 PF00128 PF02806
	<i>Escherichia coli</i>	B	<i>glgX</i>	GlgX	Glycogen debranching enzyme	657	P15067	PF02922 PF00128 PF02806
	<i>Mycobacterium tuberculosis</i>	B	<i>treS</i>	TreS	Trehalose synthase/amylase	601	P9WQ19	PF00128 PF16657
	<i>Mycobacterium tuberculosis</i>	B	<i>pep2</i>	Pep2	Maltokinase	455	Q7DAF6	/
	<i>Picrophilus torridus</i>	A	<i>glgE</i>	GlgE	Alpha-1,4-glucan: maltose-1-phosphate maltosyltransferase	630	Q6L2Z8	PF11896 PF00128
	<i>Mycobacterium tuberculosis</i>	B	<i>Rv3032</i>	Rv3032	Glycogen synthase	414	P9WMY9	PF13439 PF00534
Polyhydroxyalkanoates (PHAs)	<i>Haloferax mediterranei</i>	A	<i>phaA</i>	PhaA	Beta-ketothiolase	383	I3R3D1	PF00108 PF02803
	<i>Haloarcula hispanica</i>	A	<i>phaB</i>	PhaB	Acetoacetyl-CoA reductase	247	E1U2R6	PF00106 PF08659 PF13561
	<i>Haloarcula hispanica</i>	A	<i>phaC</i>	PhaC	PHA synthase subunit C	474	G0HQZ6	PF07167 PF00561 PF06850 PF14520
	<i>Haloarcula hispanica</i>	A	<i>phaE</i>	PhaE	PHA synthase subunit E	181	G0HQZ5	PF09712
	<i>Haloferax mediterranei</i>	A	<i>phaP</i>	PhaP	PHA granule-associated protein	154	I3R9Z2	/
	<i>Burkholderia pseudomallei</i>	B	<i>phaZ</i>	PhaZ	Intracellular polyhydroxyalkanoate depolymerase	423	Q3JIM5	PF06850
	<i>Rhizobium fredii</i>	B	<i>phaZ</i>	PhaZ	Polyhydroxyalkanoate depolymerase	369	G9AII6	PF10503
Triacylglycerol (TAG)	<i>Saccharomyces cerevisiae</i>	E	<i>PDAT</i>	PDAT	Phospholipid: diacylglycerol acyltransferase	661	P40345	PF02450
	<i>Acinetobacter baylyi</i>	B	<i>wax-dgaT</i>	WS/ DGAT	Wax Ester Synthase/Acyl Coenzyme A: Diacylglycerol Acyltransferase	458	Q8GGG1	PF03007 PF06974
	<i>Pseudomonas aeruginosa</i>	B	<i>lip</i>	Lip	Triacylglycerol lipase	311	P26876	PF00561
Wax Ester (WE)	<i>Acinetobacter baylyi</i>	B	<i>wax-dgaT</i>	WS/ DGAT	Wax Ester Synthase/Acyl Coenzyme A: Diacylglycerol Acyltransferase	458	Q8GGG1	PF03007PF06974

**Table 1.** Enzymes directly associated with the metabolism of five energy storage compounds (polyphosphate, glycogen, polyhydroxyalkanoate, triacylglycerol and wax ester) in microorganisms. <sup>#</sup>D: three life domains: Archaea (A), Bacteria (B) and Eukaryote (E). L: Protein length. UPID: UniProt Identifier. PID: Pfam Identifier.

that Tres-Pep2-GlgE-GlgB pathway might be widespread in archaeal domain, contradictory to current view that the pathway is mainly present in bacteria<sup>11</sup>. As for the NCP II synthesis pathway, it has three enzymes GlgC, Rv3032 and GlgB (GH57) and is mainly associated with thermophilic lifestyle (Fig. 2). As for all glycogen-related enzymes, Rv3032 is most widely distributed and 290 archaea has homologs of this enzyme in their proteomes, which suggested that this enzyme might be centric to glycogen formation in archaea.

Degrading enzymes exist in classical glycogen metabolism pathway, which are glycogen phosphorylase (GlgP) and glycogen debranching enzyme (TreX or GlgX). 69 archaea have GlgP enzyme while only 16 archaea have TreX or GlgX enzyme. Since TreX and GlgX are highly similar, HMM-based sequence search cannot distinguish the two proteins. According to Fig. 2, it becomes obvious that many archaea harbour glycogen degrading enzymes but much fewer the biosynthesis pathways. On our data, GlgP is more widespread in archaea than GlgX or TreX. For detailed illustration, please refer to Fig. 2. As for the CP Null Synthesis pathway, there are 21 archaea in this group. These have a small average proteome size of 800 proteins per proteome. Thus, consistent with other two storage compounds, loss of energy reserve ability may lead to a reduced genome size, although the specific mechanisms could be different<sup>37</sup>.

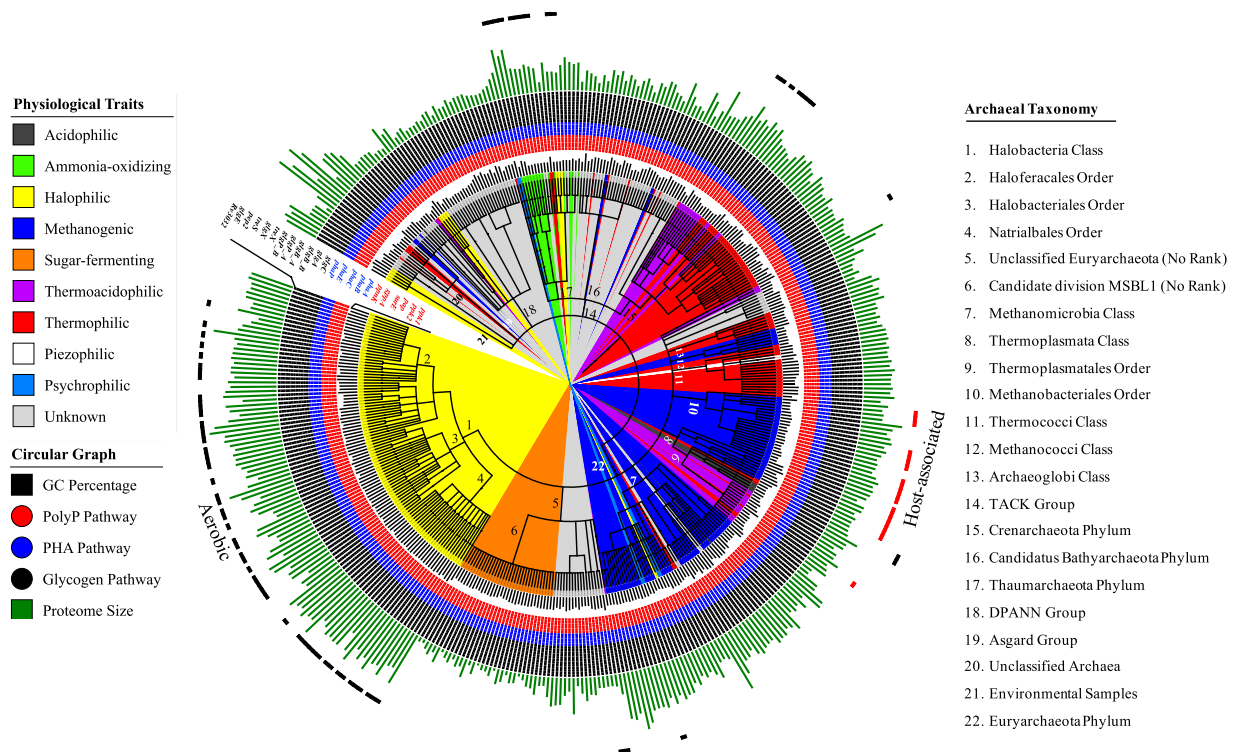
**Bioinformatics Analysis of PHA Distribution in Archaea.** Polyhydroxyalkanoic acids (PHAs) are a complex class of biodegradable polyesters found in a wide range of microorganisms, among which short-chain-length (SCL) polyhydroxybutyrate (PHB), poly-3-hydroxyvalerate (PHV) and poly-3-hydroxybutyrate-co-3-hydroxyvalerate (PHBV) are the three main components synthesized in archaea<sup>16</sup>. Currently, researchers are mainly interested in the industrial production and biotechnological applications of the biodegradable PHAs from archaea<sup>54</sup>. Biological functions of PHAs in archaea are less focused. It has been known that PHAs are normally stored in archaea as intracellular energy storage compound and have many potentials in medical applications<sup>55</sup>. Key enzymes involved in PHA biosynthesis include PhaA, PhaB, PhaC and PhaE (Fig. 3)<sup>16</sup>. In addition, PhaP, a PHA-related phasin, was also found to be an essential protein in PHA formation, which is responsible for PHA accumulation and granule morphology<sup>17</sup>.

Pathway analysis in 427 archaea showed that 51 species have complete synthesis pathway for PHA accumulation while 31 species have none of the enzymes (Supplementary Table S1). All of the 51 species with complete PHA synthesis pathway were halophiles, which indicates that PHA accumulation is associated with halophilic phenotype. A recent review also stated that PHA production in archaea has been limited to *Haloarchaeal* species<sup>13</sup>. Although PHA was widely studied in halophilic archaea for industrial purposes, its contributions to archaeal physiology were rarely investigated. Several bacterial studies provided evidences to support the protective roles that PHA might play for archaea under stress conditions, such as UV irradiation, heat, osmotic shock, desiccation and oxidative stress. For example, it has been reported that disruption of the polyhydroxyalkanoate synthase (PhaC) gene in *Aeromonas hydrophila* reduces its survival ability under stress conditions including high osmotic pressure<sup>56</sup>. The study suggested that RNA polymerase sigma S (RpoS) played an important regulatory role in the enhanced resistance of the species to stresses conferred by PHBHHx, a copolymer consisting of 3-hydroxybutyrate and 3-hydroxyhexanoate<sup>56</sup>. However, molecular mechanisms for preferred accumulation of PHAs in halophilic archaea and how PHA contributes to archaeal adaptation in high salinity environments are still lacking and require further experimental exploration. As for the 31 species with no PHA synthesis enzymes, they are associated with different types of archaeal physiological phenotypes and no specific correlational pattern is identified (Fig. 3). Consistently, their proteome sizes are significantly reduced on average ( $P$ -value < 0.05).

**TAG and WE Metabolisms in Archaea.** Currently, TAG is mainly distributed in members of the prokaryotic *Actinomycetes*, while WE is more closely associated with marine or aquatic organisms such as *Marinobacter* or *Acinetobacter*, etc.<sup>57–59</sup>. TAG and WE distributions in archaea are less investigated systematically due to experimental limitations and genome availability. Initial scanning for 18 archaeal genome sequences failed to identify any homologues of bifunctional wax ester synthase/acyl-CoA: Diacylglycerol acyltransferase<sup>18</sup>. Until now, TAG accumulation in archaea has not been reported<sup>60</sup>. In this study, all essential enzymes for direct metabolism of both wax ester (*wax-dgaT*) and triacylglycerol (*PDAT*, *lip* and *wax-dgaT*) are not identified in 427 archaeal proteomes through different bioinformatic methods, such as hmmscan, phmmer, and jackhmmer, which confirmed previous conclusion that archaea do not accumulate TAG and WE. However, our study cannot exclude the possibility that other unrecognized pathways might exist, which might be responsible for synthesizing or utilizing the two neutral lipids.

**Phylogenetic Analysis of ESM in Archaea.** Archaeal phylogenetic tree constructed from NCBI taxonomy identifiers provided a clustered overview of species with a variety of physiological from halophiles, to methanogens and to thermophiles (Fig. 6). Halophiles are correlated with higher GC content and larger proteome size, while thermophilic archaea normally have small genome sizes and lower GC content. Thus, it is consistent with previous reports that high temperature leads to small genome size, but high GC content is not correlated with thermal stability<sup>61</sup>. In contrast, genomes with high GC content may be more stable in high salinity environments. From the distribution patterns of the five energy storage metabolism pathways, we can find that polyP metabolism is important in methanogens while PHA metabolism plays important roles in halophiles. As for glycogen metabolism, it is possible that it is correlated with thermophilic and thermoacidophilic archaea. Another feature presented in Fig. 6 that is worth of mentioning is that halophiles are mainly associated with aerobic lifestyles (outermost black circle)<sup>62</sup>, while host-associated archaea are mainly found among the methanogens (outermost red circle)<sup>63</sup>.

At this point, archaea are still ecological dark matter due to being unculturable and living in inhospitable environments. Evolutionary study of archaea is still at the infant stage due to the limited number of completely sequenced and assembled archaeal genomes. Insights from our analyses of the 427 archaeal proteomes in this study are helpful for a better understanding of the common features in archaeal genomes and their physiology. With more archaeal genomes available in near future, there will be a much clearer picture about archaeal metabolism, physiology, and adaptation to extreme niches.



**Figure 6.** Overview of distribution patterns of all enzymes directly linked with PolyP (red dot), PHA (blue dot), and glycogen (black dot) metabolism. The taxonomy identifier based phylogenetic tree is coloured based on archaeal physiology such as acidophilic (dark grey), halophilic (yellow) and thermophilic (red), etc. For the circular graph from inside out, they are GC percentage (black bars), PolyP metabolism (*ppk1*, *ppk2*, *pap*, *surE*, *gppA*, *ppnK*), PHA metabolism enzymes (*phaA*, *phaB*, *phaC*, *phaE*, *phaP*), glycogen metabolism enzymes (*glgC*, *glgA*, *glgB\_B*, *glgB\_A*, *glgP\_A*, *glgP\_B*, *treX*, *glgX*, *treS*, *pep2*, *glgE*, *Rv3032*) and proteome sizes (green bars). Blank and filled dots mean loss and gain of corresponding enzymes. For each dot circle, corresponding gene name was given. The outmost two dashed circles represent two phenotypes, aerobic lifestyle (black) and host-associated lifestyle (red).

## Methods and Materials

**Collection of Archaeal and Bacterial Proteomes.** A total of 428 archaeal reference proteomes were downloaded from UniProt database via website scraping<sup>20</sup>. Proteome UP000053961 (corresponding to taxon 301375, *Methanosaeta harundinacea*) was removed from the analysis list due to its inconsistency during construction of phylogenetic tree through phyloT (<https://phylot.biobyte.de/>). Thus, 427 archaea were included in this study. All proteome file names were renamed with their unique UniProt proteome identifiers. Associated information such as corresponding names, NCBI taxonomy identifiers<sup>64</sup>, proteome sizes (proteins per proteome) and enzyme distributions related to five energy storage compounds were all included in Supplementary Table S1 for further analysis.

**Collection of Metabolism Pathways.** Pathways of five energy storage compounds were selected for thorough screening in archaeal and bacterial proteomes: polyP (involving genes *ppk1*, *ppk2*, *pap*, *surE*, *gppA*, *ppnK*, *ppgK*, and *arp*), glycogen (*glgC*, *glgA*, *glgB*, *glgP*, *treX*, *glgX*, *treS*, *pep2*, *glgE* and *Rv3032*), PHAs (*phaA*, *phaB*, *phaC*, *phaE*, *phaP* and *phaZ*), WE (*wax-dgaT*) and TAG (*PDAT*, *wax-dgaT* and *lip*). All enzymes directly linked with the metabolism of the five energy storage compounds were collected and are summarized in Table 1, together with gene names, protein names, original source, enzyme length, UniProt ID and Pfam ID.

**Ab initio Construction of Hidden Markov Model (HMM).** Seed sequences of proteins directly linked to the metabolism of energy storage compounds were selected by mining literature and searching UniProt database, focusing on archaeal species<sup>20</sup>. For those enzymes with no homologous sequences in archaea, bacterial counterparts were used. When archaea and bacteria have non-homologous enzymes for the same function, both enzymes were included as seed proteins for HMM constructions, such as GH57 and GH13 GlgB. A complete list of the collected proteins is presented in Table 1. After obtaining sequences for all seed proteins, position specific iterated BLAST (PSI-BLAST) was performed to collect homologous sequences for each seed protein from the NCBI non-redundant (nr) database of protein sequences (Max target sequence number = 1000, E-value < 0.001) except for GlgB, GlgP, GlgX, PhaP and PhaE<sup>65,66</sup>. 5000 homologues of the five enzymes were collected due to their high conservation. Perl script nrdb90.pl was used to remove 90% or higher similar sequences<sup>67</sup>. The standalone command-line version of MUSCLE was used so the MSAs were done automatically<sup>68</sup>. Heads or tails of multiple sequence alignments tend to be more inconsistent<sup>69</sup>. Thus, all MSAs were manually edited to remove heads and tails by using JalView<sup>70</sup>. The HMMER package was used for *ab initio* construction of HMMs through the

hmmbuild command by using multiple sequence alignments<sup>71</sup>. All MSAs results were converted from FASTA to STOCKHOLM format before construction by using Biopython command *SeqIO.convert*<sup>72</sup>. A total of 427 archaeal reference proteomes were scanned for the presence of the collected seed proteins. All distribution patterns of enzymes across species were recorded in Supplementary Table S1.

**Proteome Screening for Enzyme Distributions.** In order to make sure that all homologous sequences of corresponding enzyme for energy storage compounds were identified, several different methods were initially used to compare the results of archaeal and bacterial proteome searching: full-length hidden Markov models, phmmer (similar to BLASTP), jackhmmmer (similar to PSI-BLAST), and concatenated Pfam domains. Only full-length HMM results were analysed in this study. The other three methods turned out to be overly restrictive or overly loose when compared with HMM methods. For the screening results, E-value was set to 0.001 and minimal length of hit sequences was 60% of query sequences in order to get rid of fragmentary sequences. Pfam server (<http://pfam.xfam.org/>) was used to double-check domain organizations of hit sequences whenever suitable.

**Evolutionary Analysis.** NCBI taxonomy identifiers were obtained from UniProt database, which were later used for constructing a phylogenetic tree via phyloT (<https://phyloT.biobyte.de/about.cgi>) in order to investigate distribution patterns of energy storage metabolism in evolutionary settings. phyloT is an online server, using NCBI taxonomy identifier to generate a pruned tree automatically in user-defined output format. Tree visualization and annotation were undertaken through the freely available online editing tool iTOL, with pre-defined *tol\_binary* and *tol\_simple\_bar* templates, *etc.*<sup>73</sup>. Distribution patterns of ESMs are added into the phylogenetic tree as circular dot plots in order to identify how energy reserves are distributed in evolution.

**Statistical Analysis.** Python scripts and R programming, especially the ggplot2 package, were used throughout the study for data visualization (R Core Team, 2015). Unpaired two-tailed Student's t-test was used for statistical analysis. Significant difference was defined as p-value less than 0.05.

## Data Availability Statement

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

## References

- Wang, L. & Wise, M. J. Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften* **98**, 719–729, <https://doi.org/10.1007/s00114-011-0832-x> (2011).
- Wilkinson, J. F. The problem of energy-storage compounds in bacteria. *Experimental Cell Research*, **20** (1959).
- Wang, L., Liu, Z., Dai, S., Yan, J. & Wise, M. J. The Sit-and-Wait Hypothesis in Bacterial Pathogens: A Theoretical Study of Durability and Virulence. *Front Microbiol* **8**, 2167, <https://doi.org/10.3389/fmicb.2017.02167> (2017).
- Murphy, D. J. The dynamic roles of intracellular lipid droplets: from archaea to mammals. *Protoplasma* **249**, 541–585, <https://doi.org/10.1007/s00709-011-0329-7> (2012).
- Docampo, R. In *Inclusions in Prokaryotes* (ed. Shiverly, J. M.) (Springer-Verlag, 2006).
- Orell, A., Navarro, C. A., Rivero, M., Aguilar, J. S. & Jerez, C. A. Inorganic polyphosphates in extremophiles and their possible functions. *Extremophiles* **16**, 573–583, <https://doi.org/10.1007/s00792-012-0457-9> (2012).
- Achbergerová, L. & Nahálka, J. PPK1 and PPK2—which polyphosphate kinase is older? *Biologia* **69**, 7 (2014).
- Wang, L. *et al.* Distribution Patterns of Polyphosphate Metabolism Pathway and Its Relationships With Bacterial Durability and Virulence. *Frontiers in Microbiology*, **12** (2018).
- Whitehead, M. P., Eagles, L., Hooley, P. & Brown, M. R. Most bacteria synthesize polyphosphate by unknown mechanisms. *Microbiology* **160**, 829–831, <https://doi.org/10.1099/mic.0.075366-0> (2014).
- Ball, S., Colleoni, C., Cenci, U., Raj, J. N. & Tirtiaux, C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot* **62**, 1775–1801, <https://doi.org/10.1093/jxb/erq411> (2011).
- Chandra, G., Chater, K. F. & Bornemann, S. Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology* **157**, 1565–1572, <https://doi.org/10.1099/mic.0.044263-0> (2011).
- Stadthagen, G. *et al.* Genetic Basis for the Biosynthesis of Methylglucose Lipopolysaccharides in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **282**, 27270–27276, <https://doi.org/10.1074/jbc.M702676200> (2007).
- Tan, G. A. *et al.* Start a Research on Biopolymer Polyhydroxyalkanoate (PHA): A Review. *Polymers* **6**, 49 (2014).
- Han, J. *et al.* Wide Distribution among Halophilic Archaea of a Novel Polyhydroxyalkanoate Synthase Subtype with Homology to Bacterial Type III Synthases. *Applied and Environmental Microbiology* **76**, 7811–7819, <https://doi.org/10.1128/aem.01117-10> (2010).
- Kourmentza, C. *et al.* Recent Advances and Challenges towards Sustainable Polyhydroxyalkanoate (PHA) Production. *Bioengineering* **4**, <https://doi.org/10.3390/bioengineering4020055> (2017).
- Poli, A., Di Donato, P., Abbamondi, G. R. & Nicolaus, B. Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by archaea. *Archaea* **2011**, 13, <https://doi.org/10.1155/2011/693253> (2011).
- Cai, S. *et al.* Identification of the Haloarchaeal Phasin (PhaP) That Functions in Polyhydroxyalkanoate Accumulation and Granule Formation in *Haloferax mediterranei*. *Applied and Environmental Microbiology* **78**, 1946–1952, <https://doi.org/10.1128/aem.07114-11> (2012).
- Kalscheuer, R. & Steinbuechel, A. A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. *J Biol Chem* **278**, 8075–8082, <https://doi.org/10.1074/jbc.M210533200> (2003).
- R., K. In *Handbook of Hydrocarbon and lipid microbiology* (ed. Timmis, K. N.) (Springer, 2010).
- UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699, <https://doi.org/10.1093/nar/gky092> (2018).
- Wattam, A. R. *et al.* Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* **45**, D535–D542, <https://doi.org/10.1093/nar/gkw1017> (2017).
- Hedderich, R. & Whitman, W. B. In *The Prokaryotes* (eds Rosenberg, E. *et al.*) (Springer, 2013).
- Remonsellez, F., Orell, A. & Jerez, C. A. Copper tolerance of the thermoacidophilic archaeon *Sulfolobus metallicus*: possible role of polyphosphate metabolism. *Microbiology* **152**, 59–66, <https://doi.org/10.1099/mic.0.28241-0> (2006).
- Voica, D. M., Bartha, L., Banciu, H. L. & Oren, A. Heavy metal resistance in halophilic Bacteria and Archaea. *FEMS Microbiol Lett* **363**, <https://doi.org/10.1093/femsle/fnw146> (2016).

25. Soto, D. F. *et al.* Global effect of the lack of inorganic polyphosphate in the extremophilic archaeon *Sulfolobus solfataricus*: A proteomic approach. *J Proteomics*, <https://doi.org/10.1016/j.jprot.2018.02.024> (2018).
26. Orell, A., Remonsellez, F., Arancibia, R. & Jerez, C. A. Molecular Characterization of Copper and Cadmium Resistance Determinants in the Biomining Thermoacidophilic Archaeon *Sulfolobus metallicus*. *Archaea* **2013**, 1–16, <https://doi.org/10.1155/2013/289236> (2013).
27. Rao, N. N., Gomez-Garcia, M. R. & Kornberg, A. Inorganic polyphosphate: essential for growth and survival. *Annu Rev Biochem* **78**, 605–647, <https://doi.org/10.1146/annurev.biochem.77.083007.093039> (2009).
28. Nocek, B. *et al.* Polyphosphate-dependent synthesis of ATP and ADP by the family-2 polyphosphate kinases in bacteria. *Proc Natl Acad Sci USA* **105**, 17730–17735, <https://doi.org/10.1073/pnas.0807563105> (2008).
29. Rudnick, H., Hendrich, S., Pilatus, U. & Blotevogel, K. Phosphate accumulation and the occurrence of polyphosphates and cyclic 2,3-diphosphoglycerate in *Methanosarcina frisia*. *Archives of Microbiology* **154**, 5 (1990).
30. Ferry, J. G. *Methanogenesis* (Springer, 1993).
31. Wang, R., Li, Y., Wang, Y., Chen, Y. & Vanrolleghem, P. A. Effect of high orthophosphate concentration on mesophilic anaerobic sludge digestion and its modeling. *Chemical Engineering Journal* **260**, 10 (2015).
32. Jasso-Chavez, R. *et al.* Air-adapted *Methanosarcina acetivorans* shows high methane production and develops resistance against oxygen stress. *PLoS One* **10**, e0117331, <https://doi.org/10.1371/journal.pone.0117331> (2015).
33. Toso, D. B., Javed, M. M., Czornyj, E., Gunsalus, R. P. & Zhou, Z. H. Discovery and Characterization of Iron Sulfide and Polyphosphate Bodies Coexisting in *Archaeoglobus fulgidus* Cells. *Archaea* **2016**, 4706532, <https://doi.org/10.1155/2016/4706532> (2016).
34. Cardona, S. T., Chavez, F. P. & Jerez, C. A. The exopolyphosphatase gene from *Sulfolobus solfataricus*: characterization of the first gene found to be involved in polyphosphate metabolism in archaea. *Appl Environ Microbiol* **68**, 4812–4819 (2002).
35. Pfluger, K. *et al.* Identification of genes involved in salt adaptation in the archaeon *Methanosarcina mazei* Go1 using genome-wide gene expression profiling. *FEMS Microbiol Lett* **277**, 79–89, <https://doi.org/10.1111/j.1574-6968.2007.00941.x> (2007).
36. Bini, E. Archaeal transformation of metals in the environment. *FEMS Microbiol Ecol* **73**, 1–16, <https://doi.org/10.1111/j.1574-6941.2010.00876.x> (2010).
37. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J* **8**, 1553–1565, <https://doi.org/10.1038/ismej.2014.60> (2014).
38. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *Bioessays* **35**, 829–837, <https://doi.org/10.1002/bies.201300037> (2013).
39. Martinez-Bussenius, C., Navarro, C. A. & Jerez, C. A. Microbial copper resistance: importance in biohydrometallurgy. *Microb Biotechnol* **10**, 279–295, <https://doi.org/10.1111/1751-7915.12450> (2017).
40. Rivero, M. *et al.* Inorganic Polyphosphate, Exopolyphosphatase, and Pho84-Like Transporters May Be Involved in Copper Resistance in *Metallosphaera sedula* DSM 5348(T). *Archaea* **2018**, 5251061, <https://doi.org/10.1155/2018/5251061> (2018).
41. Nelson-Sathi, S. *et al.* Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc Natl Acad Sci USA* **109**, 20537–20542, <https://doi.org/10.1073/pnas.1209119109> (2012).
42. Knig, H., Skorko, R., Zillig, W. & Reiter, W.-D. Glycogen in thermoacidophilic archaeobacteria of the genera *Sulfolobus*, *Thermoproteus*, *Desulfurococcus* and *Thermococcus*. *Archives of Microbiology* **132**, 297–303, <https://doi.org/10.1007/bf00413378> (1982).
43. Murray, P. A. & Zinder, S. H. Polysaccharide reserve material in the acetotrophic methanogen, *Methanosarcina thermophila* strain TM-I: accumulation and mobilization. *Archives of Microbiology* **147**, 8 (1987).
44. Montero, M. *et al.* *Escherichia coli* glycogen genes are organized in a single glgBXCAP transcriptional unit possessing an alternative suboperonic promoter within glgC that directs glgAP expression. *Biochem J* **433**, 107–117, <https://doi.org/10.1042/BJ20101186> (2011).
45. Koliwer-Brandl, H. *et al.* Metabolic Network for the Biosynthesis of Intra- and Extracellular alpha-Glucans Required for Virulence of *Mycobacterium tuberculosis*. *PLoS Pathog* **12**, e1005768, <https://doi.org/10.1371/journal.ppat.1005768> (2016).
46. Sambou, T. *et al.* Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Mol Microbiol* **70**, 762–774, <https://doi.org/10.1111/j.1365-2958.2008.06445.x> (2008).
47. Murakami, T., Kanai, T., Takata, H., Kuriki, T. & Imanaka, T. A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* **188**, 5915–5924, <https://doi.org/10.1128/JB.00390-06> (2006).
48. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–285, <https://doi.org/10.1093/nar/gkv1344> (2016).
49. Woo, E. J. *et al.* Structural insight into the bifunctional mechanism of the glycogen-debranching enzyme TreX from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* **283**, 28641–28648, <https://doi.org/10.1074/jbc.M802560200> (2008).
50. Stadthagen, G. *et al.* Genetic basis for the biosynthesis of methylglucose lipopolysaccharides in *Mycobacterium tuberculosis*. *J Biol Chem* **282**, 27270–27276, <https://doi.org/10.1074/jbc.M702676200> (2007).
51. Brasen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate Metabolism in Archaea: Current Insights into Unusual Enzymes and Pathways and Their Regulation. *Microbiology and Molecular Biology Reviews* **78**, 89–175, <https://doi.org/10.1128/mmb.00041-13> (2014).
52. Santiago-Martínez, M. G. *et al.* The nutritional status of *Methanosarcina acetivorans* regulates glycogen metabolism and gluconeogenesis and glycolysis fluxes. *The FEBS Journal* **283**, 1979–1999, <https://doi.org/10.1111/febs.13717> (2016).
53. Mwirichia, R. *et al.* Metabolic traits of an uncultured archaeal lineage—MSBL1—from brine pools of the Red Sea. *Sci Rep* **6**, 19181, <https://doi.org/10.1038/srep19181> (2016).
54. Quillaguamán, J., Guzmán, H., Van-Thuoc, D. & Hatti-Kaul, R. Synthesis and production of polyhydroxyalkanoates by halophiles: current potential and future prospects. *Applied Microbiology and Biotechnology* **85**, 1687–1696, <https://doi.org/10.1007/s00253-009-2397-6> (2009).
55. Luef, K. P., Stelzer, F. & Wiesbrock, F. Poly(hydroxy alkanate)s in Medical Applications. *Chem Biochem Eng Q* **29**, 287–297, <https://doi.org/10.15255/CABEQ.2014.2261> (2015).
56. Zhao, Y. H., Li, H. M., Qin, L. F., Wang, H. H. & Chen, G. Q. Disruption of the polyhydroxyalkanoate synthase gene in *Aeromonas hydrophila* reduces its survival ability under stress conditions. *FEMS Microbiol Lett* **276**, 34–41, <https://doi.org/10.1111/j.1574-6968.2007.00904.x> (2007).
57. Rontani, J. F., Bonin, P. C. & Volkman, J. K. Production of wax esters during aerobic growth of marine bacteria on isoprenoid compounds. *Appl Environ Microbiol* **65**, 221–230 (1999).
58. Alvarez, H. M. & Steinbuechel, A. Triacylglycerols in prokaryotic microorganisms. *Appl Microbiol Biotechnol* **60**, 367–376, <https://doi.org/10.1007/s00253-002-1135-0> (2002).
59. Ishige, T., Tani, A., Sakai, Y. & Kato, N. Wax ester production by bacteria. *Curr Opin Microbiol* **6**, 244–250 (2003).
60. Garay, L. A., Boundy-Mills, K. L. & German, J. B. Accumulation of high-value lipids in single-cell microorganisms: a mechanistic approach and future perspectives. *J Agric Food Chem* **62**, 2709–2727, <https://doi.org/10.1021/jf4042134> (2014).
61. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol Evol* **5**, 966–977, <https://doi.org/10.1093/gbe/evt050> (2013).

62. Andrei, A.-Ş., Banciu, H. L. & Oren, A. Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiology Letters* **330**, 1–9, <https://doi.org/10.1111/j.1574-6968.2012.02526.x> (2012).
63. Heitman, J., Lurie-Weinberger, M. N. & Gophna, U. Archaea in and on the Human Body: Health Implications and Future Directions. *PLOS Pathogens* **11**, <https://doi.org/10.1371/journal.ppat.1004833> (2015).
64. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res* **40**, D136–143, <https://doi.org/10.1093/nar/gkr1178> (2012).
65. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
66. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**, D501–504, <https://doi.org/10.1093/nar/gki025> (2005).
67. Holm, L. & Sander, C. Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* **14**, 423–429 (1998).
68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340> (2004).
69. Wise, M. J. No so HoT – heads or tails is not able to reliably compare multiple sequence alignments. *Cladistics* **26**, 6 (2010).
70. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191, <https://doi.org/10.1093/bioinformatics/btp033> (2009).
71. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30–38, <https://doi.org/10.1093/nar/gkv397> (2015).
72. Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423, <https://doi.org/10.1093/bioinformatics/btp163> (2009).
73. 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 Res* **44**, W242–245, <https://doi.org/10.1093/nar/gkw290> (2016).

## Acknowledgements

We thank Dr. Pamela D. Mancha-Agresti from Universidade Federal de Minas Gerais, Brazil for inspiring us to conduct this study. The work was supported by Startup Foundation for Excellent Researchers at Xuzhou Medical University [D2016007], Nature and Science Foundation for Colleges and Universities of Jiangsu Province [16KJB180028], Innovative and Entrepreneurial Talent Scheme of Jiangsu Province (2017), and Natural Science Foundation of Jiangsu Province [BK20180997].

## Author Contributions

L.W., M.J.W., W.W. and C.W. wrote the main manuscript. L.W., Q.L., X.W., Y.H., Z.L. and J.H. conducted proteome collection, data collection, and data visualization. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37768-0>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2019