1   **Microbes use encapsulin protein organelles to sequester toxic reactions**

2   *Tobias W. Giessen[†,‡] and Pamela A. Silver*[†,‡]*

3   **Author affiliations:**

4   [†]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

5   [‡]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA

6   **Corresponding author:**

7   Pamela A. Silver, Harvard Medical School, Department of Systems Biology 200 Longwood Avenue,
8   Warren Alpert Building, Boston, MA 02115, TEL: 617-432-6401, FAX: 617-432-5012, Email:
9   pamela_silver@hms.harvard.edu

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27   **Abstract**

28   Cells organize and regulate their metabolism via membrane- or protein-bound organelles. In this way,

29   incompatible processes can be spatially separated and controlled. In prokaryotes, protein-based

30   compartments are used to sequester harmful reactions and store useful compounds. These protein

31   compartments play key roles in various metabolic and ecological processes ranging from iron

32   homeostasis to carbon fixation. We identified more than 900 proteinaceous encapsulin

33   nanocompartment systems in bacterial and archaeal genomes. Encapsulins can be found in 15 bacterial

34   and 2 archaeal phyla. Our analysis reveals 1 new capsid type and 9 previously unknown cargo proteins

35   targeted to the interior of encapsulins. We experimentally characterize 3 newly identified encapsulin

36   systems and illustrate their involvement in iron mineralization, oxidative and nitrosative stress

37   resistance and anaerobic ammonium oxidation, a process responsible for 30% of the N lost from the

38   oceans. We propose that encapsulins represent a widespread strategy for toxic reaction sequestration in

39   prokaryotes.

40

41

42

43

44

45

46

47

48   Cells employ compartmentalization to overcome many difficult metabolic and physiological challenges[1].

49   Eukaryotes mainly use membrane-bound organelles to sequester and control the flow of metabolites,

50   store genetic information and segregate protein processing and export[2, 3]. In contrast, the majority of

51   prokaryotes do not possess intracytoplasmic membrane systems and instead rely on alternative

52   approaches to achieve spatial control[4, 5]. A number of different protein-based compartmentalization

53   systems have been discovered, most prominently bacterial microcompartments (BMCs) like the CO_2-

54   fixing carboxysome[6, 7] and ferritins involved in iron homeostasis and maintaining redox balance[8].

55   Encapsulating enzymes or biosynthetic pathways in semi-permeable protein organelles can increase the

56   local concentrations of metabolites and enzymes, prevent the loss of toxic or volatile intermediates [9]

57   and create unique microenvironments necessary for the proper functioning of specialized enzymes [10]. In

58   addition, encapsulation allows for incompatible reactions and processes to take place in a single cell at

59   the same time.

60   A new class of proteinaceous compartments referred to as encapsulin nanocompartments has recently

61   been discovered[11, 12]. They have been reported in a number of bacterial and archaeal genomes and

62   assemble into $T = 1$ (60 subunits, 20-24 nm) or $T = 3$ (180 subunits, 30-32 nm) icosahedral hollow

63   capsids. The key feature of encapsulin systems is that they specifically encapsulate cargo proteins, which

64   are targeted to the encapsulin capsid interior via small C-terminal peptide tags referred to as targeting

65   peptides (TPs)[13]. So far, 7 encapsulin systems have been investigated. Encapsulins internalizing ferritin-

66   like proteins (Flps) were discovered in *Thermotoga maritima*[12], *Myxococcus xanthus*[14] and

67   *Rhodospirillum rubrum*[15] while an Flp-encapsulin fusion protein was found in *Pyrococcus furiosus*[16]. The

68   *M. xanthus* encapsulin system is involved in starvation-induced oxidative stress response, protecting

69   cells from the iron-dependent generation of reactive oxygen species (ROS). A second type of encapsulin

70   system encapsulates DyP-type peroxidases (Peroxi) in *Rhodococcus erythropolis*[17], *Brevibacterium*

71   *linens*[18] and *Mycobacterium tuberculosis*[19]. A role in oxidative stress response was proposed based on

3

72   cargo protein identity. These examples from diverse microbes raise the question of whether

73   compartmentalization using protein-based organelles is much more widespread than previously

74   assumed.

75   Here, we systematically explore the distribution, diversity and function of encapsulin gene clusters in

76   bacterial and archaeal genomes. Our approach is the first comprehensive analysis of encapsulin systems

77   in microbial genomes. By surveying and analyzing the newly identified encapsulin gene clusters, we can

78   assess the potential for functional nanocompartment production in organisms occupying diverse

79   habitats ranging from anaerobic hyperthermophilic environments to the human host, and predict their

80   functions at the molecular level. This enables us to formulate informed hypotheses about the roles

81   encapsulin nanocompartments play in survival, pathogenesis and in shaping global ecological processes.

82   We use a combination of bioinformatics, phylogenetic analysis and biochemistry to reveal and

83   characterize 3 new encapsulation systems involved in key cellular processes - iron mineralization,

84   oxidative and nitrosative stress resistance and anaerobic ammonium oxidation (anammox).

85   **Results**

86   ***Encapsulin systems are widespread in bacteria and archaea***

87   The widespread distribution of encapsulin nanocompartments makes them one of the most

88   predominant classes of protein organelles. Our systematic approach identified 909 encapsulin systems

89   distributed in 15 bacterial and 2 archaeal phyla spanning a remarkable breadth of microbial diversity and

90   habitats. Initially, position-specific iterative (PSI)-BLAST searches using all experimentally confirmed

91   encapsulin capsid proteins were carried out against the NCBI database of non-redundant protein

92   sequences. After consolidating the hits resulting from different searches, the 10 kb up- and downstream

93   regions of the identified encapsulin homologs were searched for the presence of phage-related genes

94   using individual protein BLAST searches. When genes related to any known phage morphotype were

4

95   detected, hits were excluded from subsequent analysis. This step was necessary to ensure that all

96   identified systems are indeed encapsulins and not unknown HK97-related prophages. This resulted in a

97   list of 909 encapsulins distributed in 15 bacterial and 2 archaeal phyla (Fig. 1A and Supplementary Data

98   1). Encapsulins can be found in anaerobic thermophilic microbes of the genera *Thermotoga*, *Clostridium*

99   (Bacteria) and *Pyrococcus* (Archaea) as well as aerobic mesophilic soil bacteria (e.g. *Streptomyces* and

100   *Bacillus*), cyanobacteria (e.g. *Acaryochloris*) and host-associated microbes, both commensal (e.g.

101   *Bacteroidetes* and *Propionibacterium*) and pathogenic (e.g. *Mycobacterium*, *Escherichia* and

102   *Pseudomonas*). Actinobacteria and Proteobacteria harbor the largest number of encapsulin systems (47

103   and 27%, respectively) while encapsulins can also be found in many less explored phyla like the

104   Planctomycetes and Tectomicrobia. Interestingly, 14 of the identified species contain 2 distinct

105   encapsulin systems (core operons, Data S1). In 8 cases, the systems were of the same core cargo type

106   and likely originated from a duplication of the core operon locus, while in 6 cases, 2 different core cargo

107   types were present (consistently Heme and Peroxi), suggesting independent acquisition via horizontal

108   gene transfer. The presence of multiple encapsulin systems within a single organism argues that each of

109   them conveys a distinct evolutionary benefit.

110   ***A diversity of cargo types***

111   We identify new cargo types and other encapsulin-associated components illustrating that encapsulin

112   systems are more complex than previously assumed. An encapsulin system is defined by the core

113   operon, composed of the encapsulin capsid and core cargo genes (Fig. 1B). All cargo proteins are

114   characterized by the presence of a targeting peptide (TP) at one of the termini. Additional cargo proteins

115   can be located outside the core operon and are referred to as 'secondary cargo'. Other associated

116   components that are not targeted for encapsulation, but might play important functional roles can be

117   encoded close to the encapsulin core operon.

118   Our search reveals 7 previously unknown core cargo types. The set of encapsulin-containing genomes

119   was searched for the presence of core cargo proteins by inspecting the neighborhood of a given

120   encapsulin gene for genes encoding proteins with C- or N-termini similar to known encapsulin TPs using

121   MultiGeneBlast[20]. The 7 newly identified core cargo proteins are generally located directly upstream of a

122   given encapsulin gene (Fig. 1B and C). They include hemerythrins (Heme), di-iron proteins known to

123   coordinate small gaseous compounds like O_2 and NO^{21}, a new type of 4 helix bundle protein (Geobac)

124   prevalent in Firmicutes and 4 distinct Flp-family proteins[22]. In addition, anammox bacteria of the phylum

125   Planctomycetes, capable of carrying out anaerobic ammonium oxidation[23], exhibit a divergent operon

126   organization with a conserved putative cargo gene located downstream of an unusual fusion-encapsulin.

127   We discovered 4 new secondary cargo proteins and show that up to 4 different cargo types are present

128   in a single genome. One approach to study the physiological functions of encapsulins is to correlate core

129   cargo type with the identity of additional secondary cargo proteins and non-encapsulated genetically

130   associated components[24]. BLAST searches in the set of encapsulin-encoding genomes were carried out

131   using 3 different TP consensus sequences as queries. The first consensus sequence used was based on

132   the TPs of all core cargo types identified in this study. Using the core consensus we identified 76

133   secondary cargo proteins of 4 different types (Flp: 27, Peroxi: 8, Heme: 14 and Ferredoxin (Fer): 27), 3 of

134   them previously unknown (Peroxi, Heme and Fer) (Fig. 2A). Interestingly, the newly discovered

135   Ferredoxins were present exclusively in Geobac-systems found either directly downstream of the

136   encapsulin gene or separated by one other gene. Additionally, TPs present in Ferredoxins were not

137   located at the C-terminus, as is the case for all other identified cargo, but at the N-terminus. Our search

138   using the FolB-based consensus sequence yielded 76 instances of FolB-type secondary cargo proteins

139   and 8 examples of a new putative cargo type, namely BphB (biphenyl dehydrogenase)-like short chain

140   dehydrogenases. Searches using the BfrB-consensus sequence resulted in the identification of 172 BfrB-

141   type secondary cargo proteins. More than one third (363) of all identified systems contain more than

142   one cargo type that is tagged for encapsulation (Supplementary Fig. 1A). We identified 65 instances of 3

143   cargo types being present in one genome and one example of 4 co-occurring cargo proteins (core:

144   Peroxi, secondary: Heme, FolB and BfrB; *Mycobacterium simiae* ATCC 25275). Besides core and

145   secondary cargo, non-targeted associated proteins are often found in the vicinity of encapsulin core

146   operons and are confined to specific genera and core cargo types. They include cupin-type oxygenases

147   found in close association with many Peroxi systems in Mycobacteria (Supplementary Fig. 1B), non-

148   targeted Flp proteins found close to Flp or Flp-fusion systems in archaea (Supplementary Fig. 1C) and

149   different types of electron-transfer proteins clustered around Flp core operons (Supplementary Fig. 1D).

150   The most diverse class of identified cargo are Flp-type proteins. They represent a large superfamily with

151   varied functions unified by the presence of a 4 helix bundle domain harboring a dinuclear iron center[22].

152   A sequence similarity network (SSN) analysis[25] of all (core, fusion and secondary) Flp-type cargo proteins

153   identified in this study allowed us to visualize distinct subfamilies and showed that 7 distinct classes of

154   Flp proteins are targeted for encapsulation by encapsulin compartments (Fig. 2B). Two (Flp-1 and Flp-2)

155   have previously been identified to be part of encapsulin systems either as distinct cargo proteins (*T.*

156   *maritima*, *M. xanthus* and *R. rubrum*)[12, 14, 15] or as an N-terminal fusion domain (*P. furiosus*)[16]. Flp-3,

157   exclusively identified in Archaea, has not been characterized and shows similarity to proteins found in

158   ferric iron- and arsenic-reducing species[26]. Two types of bacterioferritin-like proteins (BfrA and BfrB) are

159   usually involved in maintaining iron homeostasis in various bacteria representing an analogous system

160   to the eukaryotic Ferritins (Ftn)[22]. Additionally, proteins with similarity to DNA-binding proteins from

161   starved cells (DPSL) were identified in a number of archaeal species. Finally, a previously unknown family

162   of double Flp fusion proteins (Flp+Flp) was shown to be present in a number of thermophilic bacteria

163   (Supplementary Fig. 1E).

164   ***Encapsulins can be divided into four families***

7

165   Encapsulin nanocompartments were known to assemble into either small (ca. 22 nm in diameter) or

166   large (ca. 32 nm in diameter) capsids with triangulation numbers of T=1 (60 subunits) or T=3 (180

167   subunits), respectively[11]. However, it was unclear what factors determine their assembly into small or

168   large compartments or how capsid size correlates with cargo type and function. To address these issues,

169   we initially carried out a SSN analysis of all identified encapsulin capsid proteins (Fig. 3A). We found that

170   encapsulins cluster into 4 distinct groups (edge % identity: 24). It was immediately apparent that all Flp-

171   fusion encapsulins cluster together. By mapping the characterized systems onto our network we found

172   that small (60mer) encapsulins cluster together in one of the 2 large sequence groups while the single

173   known large (180mer) system (*M. xanthus*) was found in the second large sequence cluster

174   (Supplementary Fig. 2A). None of the known systems mapped onto the fourth sequence cluster from

175   now on referred to as 'New'.

176   There are 2 main determinants at the sequence level predictive of capsid size (Supplementary Fig. 2B). A

177   T3-motif (GPxGxG) present only in large capsids occurs at the N-terminal end of the E-loop while an

178   additional 8 to 10 residue E-loop insertion is present in most T=3 capsids. The fourth sequence cluster

179   (New) could not be classified as T=1 or T=3 due to large sequence insertions and no clear presence of

180   T3-motifs or E-loop insertions. Interestingly, all of the mentioned sequence insertions are present in

181   loops exposed on the outer surface of the encapsulin compartment (Supplementary Fig. 2C). The distinct

182   character of New capsids is additionally supported by the fact that all of them are found in similar gene

183   clusters that are distinct from other encapsulin loci. These clusters often contain genes encoding

184   proteins related to lipid/isoprenoid metabolism and a likely hexameric P-loop NTPase (Supplementary

185   Fig. 3).

186   All characterized encapsulins were able to assemble into nanocompartments when heterologously

187   produced in *E. coli*, highlighting the inherent robustness and assembly efficiency of encapsulins.

188   Furthermore, these assembled encapsulin nanocompartments could be isolated for biophysical and

189 biochemical analyses. We characterized 3 examples each of the T=1, T=3 and New clusters. We picked

190 sequences that are diverse both in terms of phylogenetic origin and core cargo type. After heterologous

191 expression and purification (Supplementary Fig. 2D), negative-stain transmission electron microscopy

192 (TEM) was used to visualize encapsulins. As predicted, all T=1 type encapsulins assemble into small

193 capsids of 22 (*Burkholderia xenovorans* and *Escherichia coli*) or 23 nm (*Streptomyces* sp. AA4) in average

194 diameter (Fig. 3B and C). The representatives chosen from the T=3 cluster were clearly different and as

195 hypothesized resulted in larger capsids. Interestingly, 2 of the 3 capsids were larger than expected with

196 average diameters of 37 (*Bacillaceae bacterium* MTCC 10057) and 38 nm (*Clostridium thermocellum*)

197 while *Kuenenia stuttgartiensis* capsids were 33 nm in size, similar to the previously reported *M. xanthus*

198 encapsulin capsids (32 nm). The representatives chosen from the New sequence cluster stemmed from

199 3 different phyla and assembled into larger capsids of 35 (Cyanobacteria: *Acaryochloris marina*), 36

200 (Proteobacteria: *Methylocapsa acidiphila*) and 37 nm (Actinobacteria: *Geodermatophilaceae*

201 URHB0062), similar in appearance to T=3 capsids. Overall, all investigated encapsulin capsids showed

202 very uniform size distributions with standard deviations between 0.9 (*Streptomyces* sp. AA4) and 2.1 nm

203 (*K. stuttgartiensis*) (Fig. 3C).

204 Certain core cargo types are confined to specific sequence clusters. We mapped the core cargo types of

205 all identified systems onto the SSN (Fig. 3A). Except for Flp-type systems, which were present in both the

206 T=1 and T=3 sequence clusters, all other core cargo types were confined to a specific capsid type. Peroxi

207 and Heme systems could only be found in the T=1 cluster while Geobac and NIR-HAO systems were only

208 present in the T=3 cluster. The match of core cargo types with a specific capsid size implies that different

209 capsid sizes have been adapted to serve particular functions.

210 ***A new hemerythrin cargo protects cells from oxidative and nitrosative stress in an encapsulation-***

211 ***dependent manner***

9

212   Hemerythrins are 4 helix bundle proteins that often form homo- or heterooligomeric complexes with

213   one dinuclear iron center per subunit. In prokaryotes, hemerythrin-like domains have recently been

214   proposed to be involved in O_2 and NO sensing and in the regulation of oxidative stress response[27].

215   Hemerythrins, predicted to fold into a 4 helix bundle (Supplementary Fig. 4A), are a new core cargo type

216   (Heme) in encapsulin systems. Heme core operons generally consist of the encapsulin capsid gene and a

217   hemerythrin containing a C-terminal TP encoded directly upstream (Fig. 4A). In one species (*Sporichthya*

218   *polymorpha* DSM 43042) a second hemerythrin can be found directly downstream of the encapsulin.

219   Most organisms containing a Heme system encode BfrB as an additional secondary cargo at a separate

220   locus in their genomes while in *Pseudonocardia dioxanivorans* CB1190 a Peroxi secondary cargo is

221   present.

222   Hemerythrin cargo is encapsulated by and co-purifies with encapsulin particles in a TP-dependent

223   manner. We expressed the *Streptomyces* sp. AA4 Heme core operon in *E. coli*. After capsid purification,

224   negative-stain TEM was used for particle visualization (Fig. 4B). Compared with capsids resulting from

225   the expression of the encapsulin gene alone, Heme core operon particles contain additional protein

226   density in the compartment interior. In many cases, cargo was organized in a 5- or 6-fold symmetrical

227   manner appearing as puncta associated with the interior surface of encapsulin particles (Fig. 4B insets

228   and Supplementary Fig. 4B). To confirm that the observed cargo was indeed hemerythrin, we compared

229   SDS-PAGE gels of purified encapsulin and Heme core operon capsids (Fig. 4C). As a control, particles of a

230   system where the C-terminal 12 amino acids of the hemerythrin cargo were removed was used. Only in

231   the wild-type system did hemerythrin co-purify with encapsulin particles confirming the dependence of

232   hemerythrin encapsulation on a C-terminal TP. Based on gel densitometry analysis ca. 20 molecules of

233   hemerythrin were encapsulated per 60 subunit capsid. To investigate the oligomerization state of

234   hemerythrin cargo in the absence of encapsulin we performed analytical size exclusion chromatography

235   (Fig. S4C). We found that hemerythrin cargo was present as a stable dimer in solution.

10

236   UV-Vis spectroscopy reveals that the diiron center of encapsulated Heme is protected from strong

237   reducing agents by the encapsulin capsid. Purified Heme core operon capsids showed 2 main absorption

238   peaks at 348 and 404 nm (as-isolated) dependent on the presence of hemerythrin, similar to previously

239   reported spectra of the met-form of hemerythrins where both iron centers are present in their ferric

240   (Fe^{3+}) but oxygen-free form (Fig. 4D)[28]. A clear change in the absorption spectrum upon dithionite

241   reduction occurs with the appearance of a new absorption peak at 427 nm. Following dithionite

242   treatment dinuclear iron proteins usually lose all absorption features in this region resulting from the

243   destruction of their metal center. Together, these results indicate that encapsulation protects the metal

244   center from being destroyed by high concentrations of strong reducing agents. Further UV-Vis analysis

245   of non-encapsulated hemerythrin cargo confirmed that without encapsulation all features in the UV-Vis

246   spectrum are lost upon dithionite incubation (Supplementary Fig. 4D).

247   The Heme containing encapsulin compartment protects *E. coli* from oxidative and nitrosative stress. We

248   carried out experiments investigating the influence of 2 stressors on the survival of *E. coli* expressing the

249   Heme core cargo system, namely hydrogen peroxide (H_2O_2) and nitric oxide (NO); hemerythrins detoxify

250   H_2O_2[29] and have been implicated in NO-sensing. Cells expressing either hemerythrin alone, encapsulin

251   alone or the complete Heme core cargo system were exposed to H_2O_2 or the NO-generator NOC-18.

252   Samples were taken after different time intervals and plated as serial dilutions for CFU-counting. A

253   strong protective effect against H_2O_2 was observed only when both components (encapsulin and

254   hemerythrin) were present while no effect could be detected for hemerythrin alone (Fig. 4E, upper

255   panel). Similarly, the Heme core cargo system was able to completely rescue *E. coli* growth when

256   exposed to NO (Fig. 4E, lower panel). In contrast to H_2O_2-exposure, expressing non-encapsulated

257   hemerythrin before applying NO stress led to ca. 50% survival compared with only ca. 20% for the

258   control (only encapsulin capsid present). Additionally, direct interaction of NO with hemerythrin was

259   confirmed via UV-Vis spectroscopy (Supplementary Fig. 4D).

11

260   *A new 4 helix bundle protein found in Firmicutes shows encapsulation-dependent iron mineralization*

261   *and oxidative stress protection*

262   Systematic analysis revealed the presence of a new core cargo type in a subset of Firmicutes annotated

263   as 'Geobacter domain of unknown function' (Geobac). In 61% of cases an additional new cargo type,

264   Ferredoxin, is closely associated with Geobac core operons (Fig. 5A). Fer is either located directly

265   downstream of the encapsulin gene, in most cases overlapping with it, or separated by one additional

266   gene whose function is unknown. No other secondary cargo types can be found in genomes encoding

267   Geobac systems. We turned to secondary and 3D structure prediction which revealed that Geobac likely

268   possesses a 4 helix bundle fold with 2 additional shorter C-terminal helices preceding the canonical TP

269   present in all core cargo types (Supplementary Fig. 4A and 5A). Surprisingly, none of the residues usually

270   associated with iron-coordinating 4 helix bundle proteins are present (Supplementary Fig. 5A).

271   Geobac cargo is encapsulated by and co-purifies with encapsulin particles in a TP-dependent fashion. To

272   further characterize the Geobac core cargo system we focused on the encapsulin, Geobac protein and

273   Fer cargo found in *Bacillaceae bacterium* MTCC 10057. *E. coli*-expression followed by particle

274   purification yielded assembled capsids (Fig. 5B). SDS-PAGE analysis showed the TP-dependent co-

275   purification of Geobac cargo with encapsulin compartments with ca. 150 molecules of Geobac per 180

276   subunit capsid (Fig. 5C). Analytical size-exclusion chromatography indicated that Geobac cargo exists in

277   an equilibrium between dimer and monomer in solution (Supplementary Fig. 5B). Taken together, the

278   above results suggest that Geobac binds to single encapsulin capsid proteins in a 1:1 or 2:1 ratio, leading

279   to a compartment where the interior surface of the majority of capsid proteins interacts with 1 or 2

280   cargo molecules.

281   Fer represents the first encapsulin cargo with an N-terminal TP and co-purifies with encapsulin particles.

282   When expressing the 3-gene operon (Geobac-Enc-Fer), no additional co-purifying cargo could be

12

283   observed (Fig. 5C) while the 2-gene operon (Enc-Fer) showed a band with the correct apparent

284   molecular weight expected for Fer (Supplementary Fig. 5C and D). Based on gel densitometry, 10

285   molecules of Fer co-purify per 180 subunit capsid. The identity of this protein band was confirmed to be

286   Fer via in-gel tryptic digest followed by mass spectrometry (Supplementary Fig. 5E). This suggests that

287   the N-terminal TP found in Fer is a weaker targeting signal compared with other C-terminal TPs

288   explaining why initially no co-purifying Fer cargo could be observed.

289   Geobac is an encapsulation dependent iron binding protein and is able to mineralize iron *in vivo*. UV-Vis

290   spectroscopy of purified capsids showed an absorption peak at 407 nm only when both Geobac core

291   cargo components were present, which shifted to 427 nm upon dithionite exposure indicating the

292   presence of bound iron (Fig. 5D). Since many 4 helix bundle proteins are involved in iron metabolism [22],

293   we carried out experiments where Geobac-expressing cells were exposed to elevated concentrations of

294   iron. Washed cells were either fixed for TEM or subjected to inductively coupled plasma mass

295   spectrometry (ICP-MS). When cells were expressing both Geobac cargo and encapsulin electron-dense

296   puncta were observed while no such puncta where present when only encapsulin was being expressed

297   (Fig. 5E). Electron-dense material was observed in about two thirds of cells and often organized into

298   regular paracrystalline arrangements (Supplementary Fig. 6A and B). The average particle diameter was

299   23 nm (Supplementary Fig. 6C) suggesting that particles mineralize inside encapsulin compartments

300   (outer diameter: 37 nm, shell thickness: 4-5 nm). ICP-MS analysis was used to compare the iron content

301   of cells expressing different combinations of Geobac system genes. Values adjusted for the protein

302   levels of the Geobac cargo protein (Supplementary Fig. 6D) showed that only when both complete

303   Geobac cargo (with TP) and encapsulin were present could significantly elevated iron levels be observed

304   (Fig. 5F). All other strains showed only background or slightly above background levels of retained iron.

305   Interestingly, cells expressing the 3-gene operon including Fer were able to store the largest amount of

306   iron suggesting that Fer is not necessary for iron mineralization but might improve its efficiency.

13

307   The Geobac encapsulin system protects *E. coli* from oxidative stress in the presence of elevated levels of

308   iron and likely represents the main iron storage system in Geobac-encoding organisms. We exposed *E.*

309   *coli* cells expressing Enc, Geobac cargo or both to H_2O_2 and monitored the CFU-count over time (Fig. 5G).

310   No increase in survival could be observed when only H_2O_2 was present. Only when cells were exposed to

311   both H_2O_2 and elevated levels of iron could a clear difference be recorded.

312   ***Encapsulins in anammox bacteria***

313   Anaerobic ammonium oxidation (anammox, converting NH_4^+ and NO_2^- to N_2) carried out by anammox

314   bacteria of the phylum Planctomycetes was proposed over 50 years ago based on thermodynamic

315   calculations[30] and is now known to be responsible for over 30% of the nitrogen lost from the oceans[31].

316   The anammox process is confined to a membrane-bound organelle referred to as the anammoxosome[32].

317   This organelle is proposed to help sequester highly toxic anammox intermediates (NO, hydrazine and

318   potentially hydroxylamine) thus protecting the rest of the cell[23]. However, it is unclear how the many

319   components inside the anammoxosome are protected from for example NO generated as an

320   intermediate.

321   We identified encapsulin systems in 4 anammox genomes (*K. stuttgartiensis, B. sinica, J. caeni* and *B.*

322   *fulgida*, Fig. 6A) encoding a large fusion protein NIR-HAO and an encapsulin N-terminally fused to a

323   cytochrome domain. Anammox encapsulins harbor an additional N-terminal diheme cytochrome

324   domain connected to the encapsulin capsid protein by a ca. 35 amino acid linker rich in proline, alanine

325   and glycine (cEnc, Supplementary Fig. 7A). Based on the presence of CxxCH-motifs we identified the

326   fusion domain as a c-type cytochrome. The 2 heme binding sites both contain additional conserved but

327   different residues (1: CxxxxDCxxCH, 2: ExxxxHxxCxYxxxxCxxCH) that likely influence their coordination

328   sphere and thus redox potential (Supplementary Fig. 7A). The predicted topology of encapsulin capsids

14

329   results in the N-terminus of capsid proteins protruding toward the interior of the assembled

330   compartment, leading to a particle whose inner surface is lined with diheme cytochrome domains.

331   In all 4 encapsulin loci identified in anammox bacteria a gene encoding a 2-domain fusion protein

332   (annotated as NIR: nitrite reductase (2 cupredoxin motifs) and HAO: hydroxylamine oxidoreductase (8

333   heme binding sites)) is located directly downstream of the encapsulin gene (Fig. 6A). Although no TP was

334   present in NIR-HAO, the close association of both genes, their conservation in all 4 genomes and the fact

335   that the associated encapsulin was an unusual fusion protein, led us to hypothesize that NIR-HAO might

336   represent a new type of cargo.

337   Anammox encapsulins assemble into stable capsids in the presence and absence of their N-terminal c-

338   type cytochrome fusion domain when produced in *E. coli*. We expressed the fusion encapsulin cEnc (*K.*

339   *stuttgartiensis*) as well as a mutant missing the N-terminal cytochrome domain (Enc) under anaerobic

340   conditions in *E. coli* followed by purification and TEM as well as SDS-PAGE analysis (Fig. 6B and C). Both

341   proteins are able to assemble into stable compartments, but a size difference was observed with Enc

342   forming 29 nm capsids while cEnc formed 33 nm particles. This shows that the N-terminal cytochrome

343   domain influences capsid assembly and thus size. UV-Vis spectroscopy of purified cEnc showed the

344   presence of heme that could be reduced using dithionite resulting in a shift of the 418 nm peak to

345   422 nm and the appearance of 2 new characteristic peaks at 527 and 554 nm confirming the presence of

346   c-type heme (Fig. 6D).

347   The NIR domain of the NIR-HAO cargo protein can be produced in *E. coli* under anaerobic conditions and

348   co-purifies with cEnc and Enc. First, we focused on the expression of the 2-gene operon cEnc-NIR-HAO

349   (*K. stuttgartiensis*). However, NIR-HAO was insoluble when expressed under anaerobic conditions in *E.*

350   *coli*. Difficulties in heterologously producing soluble proteins from anammox bacteria is a well-known

351   phenomenon [33]. We therefore focused on the N-terminal cupredoxin domain NIR (Supplementary Fig.

15

352   7C and D). We hypothesized that cargo targeting of the NIR-HAO cargo is mediated by the interaction of

353   NIR and the diheme cytochrome domain of cEnc based on studies showing direct interaction of

354   homologous proteins [34]. Anaerobically expressing NIR and the 2-gene operons consisting of cEnc/Enc

355   and NIR in *E. coli* indeed yielded soluble NIR (Fig. 6E). NIR co-purified with both cEnc and Enc with ca. 30

356   (cEnc) and ca. 15 (Enc) copies of NIR per 180mer capsid (Supplementary Fig. 7E). The decrease in the

357   amount of co-purifying NIR for the Enc system cannot be explained by the difference in size between

358   cEnc and Enc alone indicating that the presence of the diheme cytochrome domain improves cargo

359   loading. Based on these results, we propose that NIR-HAO is encapsulated by the fusion encapsulin cEnc.

360   **Discussion**

361   Here, we systematically identify and analyze nanocompartment gene clusters in bacterial and archaeal

362   genomes and classify them into core cargo groups. We establish a predictive method to organize

363   encapsulin capsid proteins into 4 distinct families, reflecting their size, cargo association and function.

364   We further experimentally characterize 3 newly discovered core cargo systems for assembly, cargo

365   targeting and physiological function establishing their involvement in oxidative and nitrosative stress

366   resistance, iron homeostasis and the anammox process.

367   The newly identified Heme core cargo system assembles into capsids of the T=1 class. The expression of

368   the core operon in *E. coli* confers H_2O_2 and NO resistance indicating a protective capacity of the

369   encapsulin system. Previous reports illustrated that hemerythrin model complexes and mutants are able

370   to detoxify H_2O_2 via its disproportionation into O_2 and H_2O [29, 35]. A similar mechanism in an encapsulation-

371   dependent manner, is likely responsible for the observed increase in survival upon H_2O_2 exposure when

372   the complete Heme core cargo system is present. Alternatively, increased survival could stem from

373   efficient sequestration inside the encapsulin compartment or a combination of both active

374   encapsulation-dependent detoxification and sequestration. The partial rescue from NO toxicity observed

16

375   for cells expressing Heme alone could indicate that resistance is mainly based on NO-binding via

376   hemerythrin and not active detoxification. Encapsulation would then lead to a more efficient

377   sequestration of NO, preventing toxic downstream reactions with O_2, iron or thiols more thoroughly

378   than free hemerythrin.

379   The newly identified encapsulin cargo protein Geobac, a 4 helix bundle protein found in Firmicutes, is

380   able to mineralize iron. In 61% of identified Geobac-containing gene clusters, a third gene encoding the

381   new Fer cargo type was identified. This arrangement is reminiscent of bacterioferritin-associated

382   ferredoxins which have been shown to improve mobilization of mineralized iron from bacterioferritins[36].

383   Fer in Geobac systems could similarly act as an electron-shuttle involved in mineralization or

384   mobilization of iron stored inside the much larger encapsulin compartments. In the presence of elevated

385   iron levels, cells expressing the Geobac core cargo system exhibited improved viability upon H_2O_2

386   exposure. This can be explained by the efficient sequestration of intracellular free iron when the

387   complete Geobac system is present leading to a decrease of toxic ROS-generating Fenton reactions[37]. In

388   addition, iron mineralization associated with ferroxidase activity is known to be able to actively remove

389   H_2O_2 in Flp-type proteins[38]. Based on the above results and the fact that all Firmicutes harboring Geobac

390   systems do not encode classical ferritins (Ftn) or bacterioferritins (BfrA or BfrB), we propose that Geobac

391   encapsulin operons represent the main strategy for iron storage in these organism. The fact that 2

392   conceptually similar but evolutionarily not closely related systems (ferritins and Geobac encapsulins)

393   exist to address the same essential function (iron homeostasis) highlights the power of convergent

394   evolution.

395   The presence of encapsulins in anammox bacteria is of particular significance given their role in the

396   global nitrogen cycle. Anammox encapsulins are fusion proteins where the encapsulin capsid protein

397   carries a unique N-terminal cytochrome domain. The capsid was shown to assemble into smaller

398   (33 nm) capsids than the other 2 T=3 encapsulins analyzed in this study (37 and 38 nm). The N-terminus

17

399   of encapsulin capsid proteins protrudes towards the interior of the assembled compartment. Thus, the

400   presence of a unique N-terminal domain likely influences capsid assembly and could explain the

401   differing capsid size observed for *K. stuttgartiensis*. All encapsulin loci identified in anammox genomes

402   also possess a conserved TP-less cargo protein NIR-HAO located directly downstream of cEnc. Anammox

403   genomes can contain up to 10 homologs annotated as HAOs (hydroxylamine oxidoreductases) with most

404   of them not having been assigned a physiological function[23]. The HAO domain present in the unique

405   fusion protein NIR-HAO is likely involved in reductive catalysis based on the absence of a heme-

406   coordinating tyrosine residue empirically found to be predictive of oxidative or reductive modes of HAO

407   function[39]. Encapsulation of 3 domains (diheme c-type cytochrome, NIR and HAO) inside a protein

408   compartment suggests that co-localization and sequestration play key parts in this protein organelles'

409   function. A protein shell would act as a barrier confining toxic intermediates sequestered or generated

410   inside the encapsulin to a dedicated cellular space. Such a system could function in sequestering toxic

411   reactions involved in the anammox process, as a detoxification system inside the anammoxosome and in

412   recycling intermediates lost through diffusion.

413   The widespread occurrence of protein-based organelles in prokaryotes demonstrates that many

414   bacteria and archaea employ sophisticated compartmentalization mechanisms to spatially organize and

415   regulate their metabolism and adapt to dynamic and hostile environments. Encapsulins are generally

416   found within a subset of a given lineage, often occupying unusual habitats, which argues that they were

417   acquired via horizontal gene transfer rather than vertical inheritance. Encapsulin function is based on

418   the targeted encapsulation of functional protein components and enzymatic functions often involved in

419   carrying out toxic reactions. Thus, we propose that encapsulin systems are a general strategy found in

420   microbes to sequester toxic transformations and molecules and serve as specific storage compartments.

421   **Methods**

422   **Computational analysis of microbial genomes and encapsulin gene clusters**

18

423   Genome-mining for encapsulin capsids was initiated by carrying out PSI-BLAST[40] searches using the

424   following experimentally confirmed capsid proteins as queries: WP_010865203.1 (*T. maritima*),

425   WP_009884820.1 (*B. linens*), WP_020909834.1 (*R. erythropolis*), ABF87797.1 (*M. xanthus*),

426   WP_048059055.1 (*P. furiosus*) and NP_215313.1 (*M. tuberculosis*). Five iterations where run using the

427   following parameters: Expect threshold: 1, word size: 3, matrix: BLOSUM62, gap costs: Existence: 11

428   Extension: 1, compositional adjustments: conditional compositional score matrix adjustment, PSI-

429   threshold: 0.0005. All hits were combined and the surroundings (10 kb) of the encapsulin gene inspected

430   for the presence of genes annotated as phage-related using BLASTp with default settings. If the case,

431   these hits were deleted. This resulted in a list of 909 encapsulin-encoding loci.

432   Phylogenetic analysis was based on a Clustal Omega (ClustalO)[41] alignment carried out using the default

433   settings of the Multiple Sequence Alignment online tool of the European Molecular Biology Laboratory's

434   European Bioinformatics Institute (EMBL-EBI). A nearest-neighbor phylogenetic tree based on the

435   ClustalO alignment was generated using Geneious 9.1.4 with the HK97 bacteriophage capsid protein as

436   the outgroup (NP_037666.1) and then annotated and visualized using the Interactive Tree of Life (iTOL

437   v3) online server[42].

438   To identify core cargo proteins present in the set of encapsulin-encoding genomes MultiGeneBlast

439   (MGB)[20] searches using the respective encapsulin proteins and a TP consensus sequence based on

440   characterized core cargo proteins (NP_228595.1 WP_044480584.1 NP_215314.1 WP_009884819.1

441   ABF88760.1, SDGSLGIGSLKRS) as dual queries were carried out. The consensus was generated using

442   Geneious 9.1.4 and was based on ClustalO alignments. MGB was run in the 'Architecture search' mode

443   with minimal sequence coverage at 20%, minimal identitiy at 20%, the maximum distance between

444   genes in locus at 4 kb and the weight of synteny conservation at 1. A custom database of all encapsulin-

445   encoding genomes generated with MGB was used. Only hits where the central residue of the hit string

446   was located closer than 15 residues from the N- or C-terminus were considered valid.

19

447 Secondary cargo proteins were discovered using BLASTp searches of encapsulin-encoding genomes with

448 three different TP-consensus sequences as queries. BLASTp parameters were automatically adjusted for

449 short input sequences. The first consensus was generated from all identified core cargo proteins

450 (DGSLGTIGSLKGE). The remaining consensus queries were based on BLASTp hit lists of FolB

451 (RGSVVPAGGAAV) and BfrB (DGAPPAAGGAL) previously identified in *M. tuberculosis*[19]. Alignments were

452 generated using ClustalO and the consensus obtained via Geneious 9.1.4.

453 Sequence similarity network (SSN) analysis was carried out using the Enzyme Similarity Tool of the

454 Enzyme Function Initiative[25] with input fasta files containing all protein sequences to be analyzed. After

455 the initial datasets were created we used an alignment score (based on the alignment score vs percent

456 identity plot) that would correspond to a percent identity of 20 for initial outputting and interpretation

457 of the SSN. The resulting xgmml file was imported into Cytoscape 3.3.0. Each node in the network

458 corresponds to a single protein sequence while each edge represents a pairwise connection between

459 two proteins based on a specified % identity cutoff. The cutoffs used for final outputting were 24

460 (encapsulin capsid proteins) and 39 (Flp-type proteins). The networks were visualized using the yFiles

461 organic layout option of Cytoscape 3.3.0.

462 **Molecular biology and cloning**

463 All constructs used in this study were ordered as gBlock Gene Fragments from Integrated DNA

464 Technologies (IDT). Codon usage was optimized for *E. coli* with the IDT Codon Optimization Tool with the

465 amino acid sequences of the respective proteins of interest as input. For operons that contained

466 overlapping and/or multiple genes, intergenic regions and overlapping sequences were not changed.

467 Heme, Geobac and NIR cargo constructs for characterization in the absence of encapsulin were ordered

468 containing C-terminal His$_6$ tags. For TP-less Heme and Geobac cargo proteins, the 12 (Heme) and 10

469 (Geobac) C-terminal residues were omitted, thus removing the TP. The native *K. stuttgartiensis* proteins

470 cEnc and NIR-HAO contain N-terminal signal peptides reminiscent of twin-arginine peptides. They were

20

471   predicted using the SignalIP 4.1 Server[43]. For cEnc, the N-terminal 36 amino acids

472   (MVMGILNTFKKVYAVTGFFALLAVFSLSQVGSSAFA) and for NIR-HAO the N-terminal 27 amino acids

473   (MLNKSAALVPVVLAFLFLFLCFQCLYA) were omitted when ordering the respective constructs for cloning.

474   The dissected NIR domain of the didomain protein NIR-HAO (CAJ73224.1) was based on the 300 N-

475   terminal residues of the signal peptide-less protein based on domain annotations using Pfam, NCBI-CDD

476   and PROSITE. The N-terminal diheme cytochrome domain of cEnc was removed by omitting the 50

477   amino acids at the start of the signal peptide-less cEnc protein resulting in Enc. Gibson Assembly®

478   Master Mix was obtained from New England BioLabs (NEB). DNA sequencing was carried out by

479   GENEWIZ. NEB Turbo Competent *E. coli* cells were used for all cloning procedures while BL21 (DE3) Star

480   Competent *E. coli* (NEB) cells were used for protein production and all other experiments. The vectors

481   pETDuet1 and pCDFDuet1 were used for all cloning procedures. For the construction of expression

482   constructs, Gibson Assembly was employed. The respective gBlock Gene Fragments containing 20 bp

483   overlaps for direct assembly were combined with NdeI and PacI digested pETDuet1 or pCDFDuet1

484   yielding the constructs/strains listed in Table S1. Chemically competent *E. coli* Turbo cells were

485   transformed and the resulting plasmids confirmed by sequencing.

486   **Expression and purification of proteins and protein compartments**

487   All expression experiments were carried out in lysogeny broth (LB) supplemented with ampicillin (100

488   µg/mL), spectinomycin (100 µg/mL) or both. Size exclusion chromatography/gelfiltration for capsid

489   purification was performed with an ÄKTA Explorer 10 (GE Healthcare Life Sciences) equipped with a

490   HiPrep 16/60 Sephacryl S-500 HR column (GE Healthcare Life Sciences). For analytical size exclusion to

491   determine oligomerization state in solution a Superdex 200 10/300 GL column (GE Healthcare Life

492   Sciences) was employed. Protein samples were desalted and concentrated with Amicon Ultra Filters

493   from Millipore. For SDS-PAGE analysis, Novex Tris-Glycine Gels (ThermoFisher Scientific) were used. DNA

494   and protein concentrations were measured using a Nanodrop ND-1000 instrument (PEQLab) and UV-Vis

495   spectra were recorded with the same instrument.

496   Assembled and sequence-confirmed plasmids were used to transform *E. coli* BL21 (DE3) Star cells (ca.

497   0.5 ng total DNA). For co-transformations of two constructs ca. 15 ng of each plasmid was used.

498   For standard aerobic protein expressions 50 mL LB was inoculated (1:50) using an over-night culture,

499   grown at 37°C and 200 rpm to an OD_{600} of 0.5 and then induced with IPTG (final concentration: 0.1 mM).

500   Cultures were grown at 30°C for 18 h, harvested via centrifugation (4000 rpm, 15 min, 4°C) and the

501   pellets either immediately used or flash frozen in N_2(l) and stored at -20°C.

502   For anaerobic expressions of the *K. stuttgartiensis* constructs cEnc, Enc and NIR, LB medium was

503   inoculated (1:50) and initially grown aerobically to an OD_{600} of 0.5. The culture was then made anoxic

504   through vigorous purging using argon. For all cEnc containing expressions, hemin (stock: 500 µM in 10

505   mM KOH, final concentration: 10 µM) and freshly made $FeSO_4$ (stock: 200 mM in water, final

506   concentration: 10 µM) were added. For NIR expressions $CuCl_2$ (stock: 200 mM in water, final

507   concentration: 2 µM) was added. After adding IPTG to a final concentration of 0.1 mM, the cultures

508   were transferred to Vacu-Quik jars (Almore Internationl, inc.). Two BD GasPak EZ satchets were added.

509   This was followed by two rounds of air removal and argon flushing of the jar. Then cultures were

510   incubated at 30°C for 18 h under an argon atmosphere. Cells were centrifuged (4000 rpm, 15 min, 4°C)

511   and the pellets either immediately used or frozen in N_2(l) and stored at -20°C.

512   For encapsulin and His-tagged protein purifications, pellets were thawed, resuspended in 5 mL PBS (pH

513   7.4) buffer, then lysozyme (final concentration: 1 mg/mL) and DNaseI (final concentration: 1 µg/mL)

514   were added and the cells incubated for 20 min on ice. Cell suspensions were subjected to sonication

515   using a 550 Sonic Dismembrator (FisherScientific). Power level 3.5 was used with a pulse time of 8 sec

516   and an interval of 10 sec. Total pulse time was 3 min. Cell debris was subsequently removed through

22

517   centrifugation (8000 rpm, 15 min, 4°C). The cleared supernatant was then used either for protein affinity

518   or encapsulin particle purification.

519   His-tagged proteins were purified using Ni-NTA agarose resin (Qiagen) via batch Ni-NTA affinity

520   procedure following the supplier's instructions. Buffer A (50 mM Tris, pH 8, 300 mM NaCl, 2 mM β-

521   mercaptoethanol, 5% glycerol, 20 mM imidazol) was used to wash resin after protein binding and buffer

522   B (50 mM Tris, pH 8, 300 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, 250 mM imidazol) was used

523   to elute bound protein. Samples were concentrated and dialyzed using Amicon filters and PBS (pH 7.4)

524   buffer and evaluated using SDS-PAGE. Further analyses were carried out directly or the next day with

525   protein being stored on ice.

526   To cleared lysates (5 mL) containing encapsulin particles 0.1 g NaCl and 0.5 g of PEG-8000 was added

527   (ca. 10% w/v final concentration), followed by incubation on ice for 30 min. The precipitate was

528   collected through centrifugation (8000 rpm, 15 min, 4°C), suspended in 3 mL PBS buffer and filtered

529   using a 0.2 µm syringe filter. The samples were then subjected to size exclusion chromatography using

530   PBS buffer and a flow rate of 1 mL/min. Fractions were evaluated using SDS-PAGE analysis and protein

531   compartment containing fractions were combined and concentrated using Amicon filters. The samples

532   were either directly subjected to additional analysis or stored on ice overnight. The relative ratios

533   adjusted for molecular mass of encapsulin capsid proteins and co-purifying cargo was determined using

534   the Gel Analyzer of the open source image processing package Fiji based on ImageJ 1.51f.

535   **Transmission electron microscopy (TEM)**

536   200 Mesh Gold Grids (FCF-200-Au, EMS) were used for all TEM experiments. TEM experiments of fixed

537   cells or negatively stained protein samples were carried out at the HMS Electron Microscopy Facility

538   using a Tecnai G2 Spirit BioTWIN instrument.

539   For negative-staining TEM, encapsulin samples were diluted to 1-10 µM using Hepes buffer (10 mM

540   Hepes, pH 7.4) and subsequently adsorbed onto formvar/carbon coated gold grids. Grids were glow-

23

541   discharged using a 100x glow discharge unit (EMS) to increase their hydrophilicity (10 sec, 25 mA) before

542   applying 5 µL of diluted sample. After 1 min adsorption time, excess liquid was blotted off using

543   Whatman #1 filterpaper, washed with H_2O and floated on a 10 µL drop of staining solution (0.75%

544   uranyl formate in H_2O) for 35 sec. After removal of excess liquid samples were ready for TEM analysis at

545   80 keV.

546   For TEM analysis of fixed cells, 0.5 mL of bacterial culture was fixed by adding fixative (1:1 v/v, 1.25%

547   formaldehyde, 2.5% glutaraldehyde, 0.03% picric acid in 0.1 M sodium cacodylate buffer, pH 7.4). The

548   sample was then incubated at 25°C for 1 h and centrifuged for 3 min at 3000 rpm. The sample

549   containing pellet and supernatant in the same tube was further incubated for 6-18 h at 4°C. Cells were

550   subsequently washed three times in cacodylate buffer, 4 times with maleate buffer pH 5.15 followed by

551   staining with 1% uranyl acetate for 30 min. Then, the sample was dehydrated (15 min 70% ethanol, 15

552   min 90% ethanol, 2 x 15 min 100% ethanol) and exposed to propyleneoxide for 1 h. For infiltration, a

553   mixture of Epon resin and proylenoxide (1:1) was incubated for 2 h at 25°C before moving it to an

554   embedding mold filled with freshly mixed Epon. The sample was allowed to sink and subsequently

555   moved to a polymerization oven (24 h, 60°C). Ultrathin sections (60-90 nm) were then cut at -120°C

556   using a cryo-diamond knife (Reichert cryo-ultramicrotome) and transferred to formvar/carbon coated

557   grids.

558   **Determination of average particle diameter**

559   To determine the average diameters of encapsulin particles and electron-dense puncta resulting from

560   Geobac-expression in the presence of elevated levels of iron, TEM micrographs were analyzed using the

561   open source image processing package Fiji based on ImageJ 1.51f. Micrographs were converted to 8-bit

562   binary images, thresholded and processed using the particle analyzer plugin. For encapsulin capsids at

563   least 500 particles were analyzed, while 141 electron-dense puncta (Geobac) were recorded. The

564   diameters reported are based on Fiji Feret diameter output values.

24

565   **UV-Vis analysis of Heme, Geobac and cEnc**

566   After purification, as-isolated (oxidized) samples of Heme, Geobac, cEnc and loaded and empty

567   encapsulin capsids were directly subjected to UV-Vis analysis (Nanodrop ND-1000). To study the

568   interaction of Heme, Geobac and cEnc with different small molecule partners, samples were made

569   anoxic through repeated removal of air and flushing with argon using glass vials sealed with rubber

570   septa. Samples in PBS (100 µM) were incubated with sodium dithionite (excess: ca. 50 mM), sodium

571   azide (500 µM), sodium cyanide (500 µM), H_2O_2 (500 µM) or diethylenetriamine/nitric oxide (NOC-18,

572   6 mM) for 5 min on ice before carrying out measurements.

573   **Oxidative and nitrosative stress survival assays**

574   Overnight cultures of strains expressing different components of the Heme and Geobac encapsulin

575   systems were grown at 37°C and 200 rpm. They were used to inoculate 50 mL LB cultures (1:25) which

576   were grown at 37°C to an OD_{600} of 0.5, then induced using IPTG (0.1 mM) and incubated at 30°C and

577   200 rpm for 2 h. Cultures were then diluted to an OD_{600} of 0.3 using fresh LB. H_2O_2 (final concentration:

578   0.5 mM) or NOC-18 (final concentration: 6 mM) were then added and cultures incubated at 37°C and

579   200 rpm in the dark. For Geobac, cultures with and without additional $FeSO_4$ (250 µM) were prepared.

580   Controls did not contain H_2O_2/NOC-18. Samples were taken after 0, 1, 2 and 4 h and diluted (1:10-

581   1:1000) using fresh LB. 100 µL of each dilution was then spread on antibiotic-containing LB agar plates

582   which were incubated at 37°C for 18 h. Colonies on plates were manually counted the next day. All

583   experiments were carried out at least in triplicate. Counts were normalized using non-H_2O_2/NOC-18

584   treated cultures and expressed as survival percentages.

585   **Inductively coupled plasma mass spectrometry (ICP-MS)**

586   ICP-MS measurements were carried out at the Trace Metals Lab at Harvard T.H. Chan School of Public

587   Health. For the analysis of iron content of cells expressing different Geobac components, 1 mL (OD_{600} =

588   2.4) of induced over-night culture was pelleted and then washed four times with H_2O. Pellets were then

589   prepared for ICP-MS analysis using nitric acid resulting in a final concentration of 2% (w/v) and

590   measured with a Perkin Elmer Elan DRC-II Inductively Coupled Plasma-Mass Spectrometer. Reported

591   results are based on three independent measurements per sample.

592   **In vivo iron mineralization assays**

593   Overnight cultures of cells expressing Geobac constructs were used to inoculate 50 mL LB medium and

594   grown to an OD_{600} of 0.5. Before induction with 0.1 mM IPTG, 250 µM freshly prepared $FeSO_4$ (in water)

595   was added. The cultures were then incubated at 30°C for 18 h. Then cells were fixed for TEM analysis.

596   **Protein identification**

597   Protein identification of SDS-PAGE gel bands was carried out at the Harvard Medical School Taplin Mass

598   Spectrometry Facility. Coomassie Blue stained gels were destained to clear background. Single bands

599   were excised with as little excess empty gel as possible. Excised gel bands were cut into approximately 1

600   mm^3 pieces. Gel pieces were washed and dehydrated with acetonitrile for 10 min. Pieces were then

601   completely dried in a speed-vac. Rehydration of the gel pieces was done with 50 mM NH_4HCO_3 solution

602   containing 12.5 ng/µl modified sequencing-grade trypsin (Promega, Madison, WI) at 4°C. After 45 min,

603   excess trypsin solution was removed and replaced with 50 mM NH_4HCO_3 solution to just cover the gel

604   pieces. Samples were incubated at 37°C room overnight. Peptides were extracted by removing the

605   NH_4HCO_3 solution, followed by one wash with 50% acetonitrile and 1% formic acid. The extracts were

606   then dried in a speed-vac (1 h). The samples were stored at 4°C until analysis. On the day of analysis the

607   samples were reconstituted in 5-10 µL of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-

608   scale reverse-phase HPLC capillary column was created by packing 2.6 µm C18 spherical silica beads into

609   a fused silica capillary (100 µm inner diameter, 30 cm length) with a flame-drawn tip. After equilibrating

610   the column each sample was loaded via a Famos auto sampler (LC Packings, San Francisco, CA) onto the

611   column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B

612   (97.5% acetonitrile, 0.1% formic acid). As peptides eluted they were subjected to electrospray ionization

613   and then entered into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific,

614   Waltham, MA). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum

615   of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were

616   determined by matching protein databases with the acquired fragmentation pattern using Sequest

617   (Thermo Fisher Scientific, Waltham, MA). All databases include a reversed version of all the sequences

618   and the data was filtered to between a one and two percent peptide false discovery rate.

619   **References**

620   1.   Diekmann, Y. & Pereira-Leal, J.B. Evolution of intracellular compartmentalization. *Biochem. J.*
621   **449**, 319-331 (2013).
622   2.   Martin, W. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philosophical*
623   *transactions of the Royal Society of London. Series B, Biological sciences* **365**, 847-855 (2010).
624   3.   Schrader, M., Godinho, L.F., Costello, J.L. & Islinger, M. The different facets of organelle
625   interplay-an overview of organelle interactions. *Frontiers in cell and developmental biology* **3**, 56
626   (2015).
627   4.   Cornejo, E., Abreu, N. & Komeili, A. Compartmentalization and organelle formation in bacteria.
628   *Curr. Opin. Cell Biol.* **26**, 132-138 (2014).
629   5.   Giessen, T.W. & Silver, P.A. Encapsulation as a Strategy for the Design of Biological
630   Compartmentalization. *J. Mol. Biol.* (2015).
631   6.   Kerfeld, C.A. & Erbilgin, O. Bacterial microcompartments and the modular construction of
632   microbial metabolism. *Trends Microbiol.* **23**, 22-34 (2015).
633   7.   Kerfeld, C.A. & Melnicki, M.R. Assembly, function and evolution of cyanobacterial
634   carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66-75 (2016).
635   8.   Hintze, K.J. & Theil, E.C. Cellular regulation and molecular interactions of the ferritins. *Cell. Mol.*
636   *Life Sci.* **63**, 591-600 (2006).
637   9.   Pfeifer, F. Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* **10**, 705-715
638   (2012).
639   10.   Chen, A.H., Robinson-Mosher, A., Savage, D.F., Silver, P.A. & Polka, J.K. The bacterial carbon-
640   fixing organelle is formed by shell envelopment of preassembled cargo. *PloS one* **8**, e76127
641   (2013).
642   11.   Giessen, T.W. Encapsulins: microbial nanocompartments with applications in biomedicine,
643   nanobiotechnology and materials science. *Curr. Opin. Chem. Biol.* **34**, 1-10 (2016).
644   12.   Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment.
645   *Nat. Struct. Mol. Biol.* **15**, 939-947 (2008).
646   13.   Rurup, W.F., Snijder, J., Koay, M.S., Heck, A.J. & Cornelissen, J.J. Self-sorting of foreign proteins
647   in a bacterial nanocompartment. *J. Am. Chem. Soc.* **136**, 3828-3832 (2014).
648   14.   McHugh, C.A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria
649   from oxidative stress. *EMBO J.* **33**, 1896-1911 (2014).
650   15.   He, D. *et al.* Structural characterization of encapsulated ferritin provides insight into iron storage
651   in bacterial nanocompartments. *eLife* **5** (2016).

27

652   16.   Akita, F. *et al.* The crystal structure of a virus-like particle from the hyperthermophilic archaeon
653   Pyrococcus furiosus provides insight into the evolution of viruses. *J. Mol. Biol.* **368**, 1469-1483
654   (2007).
655   17.   Rahmanpour, R. & Bugg, T.D. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and
656   peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097-2104 (2013).
657   18.   Valdes-Stauber, N. & Scherer, S. Isolation and characterization of Linocin M18, a bacteriocin
658   produced by Brevibacterium linens. *Appl. Environ. Microbiol.* **60**, 3809-3814 (1994).
659   19.   Contreras, H. *et al.* Characterization of a Mycobacterium tuberculosis nanocompartment and its
660   potential cargo proteins. *J. Biol. Chem.* **289**, 18279-18289 (2014).
661   20.   Medema, M.H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level
662   with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218-1223 (2013).
663   21.   Alvarez-Carreno, C., Becerra, A. & Lazcano, A. Molecular Evolution of the Oxygen-Binding
664   Hemerythrin Domain. *PloS one* **11**, e0157904 (2016).
665   22.   Andrews, S.C. The Ferritin-like superfamily: Evolution of the biological iron storeman from a
666   rubrerythrin-like ancestor. *Biochim. Biophys. Acta* **1800**, 691-705 (2010).
667   23.   Kartal, B. *et al.* How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol. Rev.*
668   **37**, 428-461 (2013).
669   24.   Radford, D. in Department of Molecular Genetics, Vol. Doctorate of Philosophy (University of
670   Toronto, Toronto, Canada; 2015).
671   25.   Gerlt, J.A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for
672   generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019-1037
673   (2015).
674   26.   Straub, K.L., Hanzlik, M. & Buchholz-Cleven, B.E. The use of biologically produced ferrihydrite for
675   the isolation of novel iron-reducing bacteria. *Syst. Appl. Microbiol.* **21**, 442-449 (1998).
676   27.   Li, X. *et al.* A bacterial hemerythrin-like protein MsmHr inhibits the SigF-dependent hydrogen
677   peroxide response in mycobacteria. *Frontiers in microbiology* **5**, 800 (2014).
678   28.   Hathazi, D. *et al.* Oxidative protection of hemoglobin and hemerythrin by cross-linking with a
679   nonheme iron peroxidase: potentially improved oxygen carriers for use in blood substitutes.
680   *Biomacromolecules* **15**, 1920-1927 (2014).
681   29.   Okamoto, Y. *et al.* H2O2-dependent substrate oxidation by an engineered diiron site in a
682   bacterial hemerythrin. *Chemical communications* **50**, 3421-3423 (2014).
683   30.   Richards, F.A. *Anoxic basins and fjordsin.* (London, UK; 1965).
684   31.   Ward, B.B. & Jensen, M.M. The microbial nitrogen cycle. *Frontiers in microbiology* **5**, 553 (2014).
685   32.   Fuerst, J.A. & Sagulenko, E. Nested bacterial boxes: nuclear and other intracellular
686   compartments in planctomycetes. *J. Mol. Microbiol. Biotechnol.* **23**, 95-103 (2013).
687   33.   Javidpour, P. *et al.* Investigation of Proposed Ladderane Biosynthetic Genes from Anammox
688   Bacteria by Heterologous Expression in E. coli. *PloS one* **11**, e0151087 (2016).
689   34.   Tsuda, A. *et al.* Structural and mechanistic insights into the electron flow through protein for
690   cytochrome c-tethering copper nitrite reductase. *Journal of biochemistry* **154**, 51-60 (2013).
691   35.   Mauerer, B., Crane, J., Schuler, J., Wieghardt, K. & Nuber, B. A Hemerythrin Model Complex with
692   Catalase Activity. *Angewandte Chemie* **32**, 289-291 (2003).
693   36.   Wang, Y. *et al.* Characterization of the Bacterioferritin/Bacterioferritin Associated Ferredoxin
694   Protein-Protein Interaction in Solution and Determination of Binding Energy Hot Spots.
695   *Biochemistry* **54**, 6162-6175 (2015).
696   37.   Cozzi, A. *et al.* Overexpression of wild type and mutated human ferritin H-chain in HeLa cells: in
697   vivo role of ferritin ferroxidase activity. *J. Biol. Chem.* **275**, 25122-25129 (2000).

698   38.   Zhao, G. *et al.* Iron and hydrogen peroxide detoxification properties of DNA-binding protein
699   from starved cells. A ferritin-like DNA-binding protein of Escherichia coli. *J. Biol. Chem.* **277**,
700   27689-27696 (2002).
701   39.   Klotz, M.G. *et al.* Evolution of an octahaem cytochrome c protein family that is key to aerobic
702   and anaerobic ammonia oxidation by bacteria. *Environ. Microbiol.* **10**, 3150-3163 (2008).
703   40.   Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search
704   programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).
705   41.   Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments
706   using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
707   42.   Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
708   annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-245 (2016).
709   43.   Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides
710   from transmembrane regions. *Nat. Methods* **8**, 785-786 (2011).

711

Acknowledgments

Author contributions

719   T.W.G. and P.A.S designed the research and wrote the paper. T.W.G. performed experiments.

Competing interests

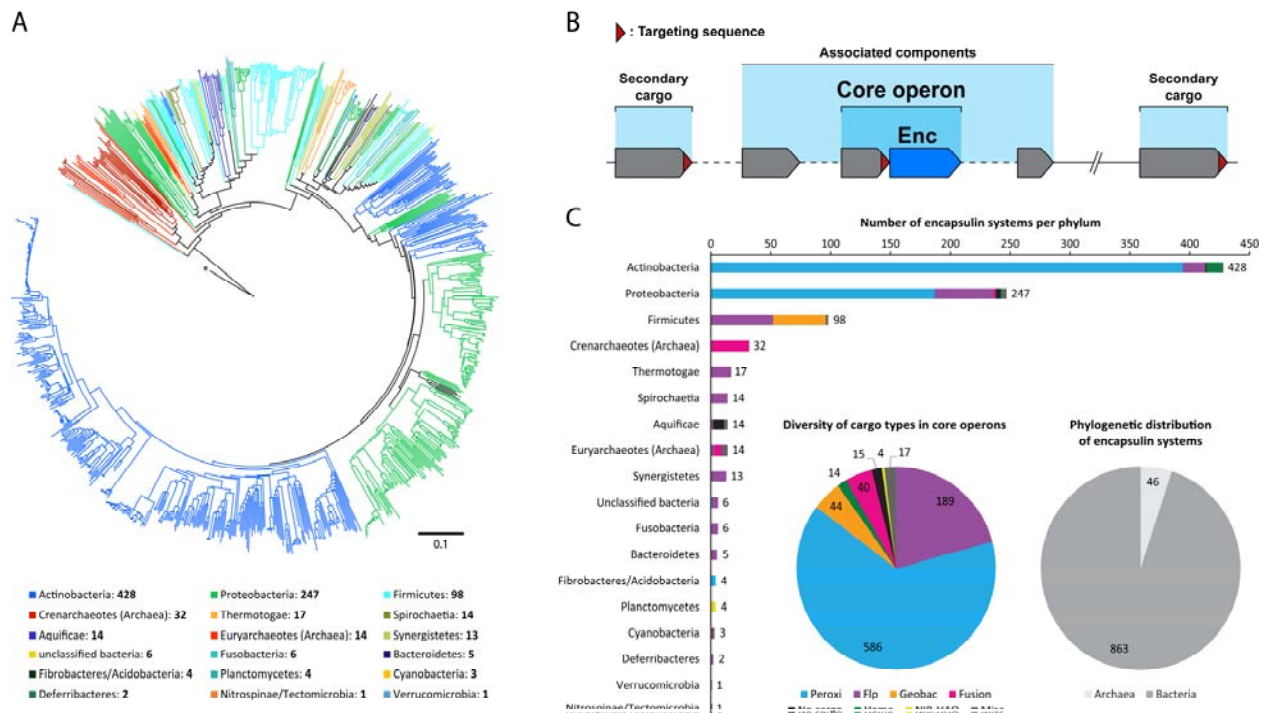721   The authors declare no competing financial interests.

722

723

724

725

726

727

728 **Figures and figure legends**

730 **Figure 1: Distribution, organization and diversity of encapsulin systems in prokaryotes.** (A) Nearest-neighbor phylogenetic tree of 909 encapsulin capsid proteins using the major capsid protein of bacteriophage HK97 as the outgroup (*) based on a ClustalOmega multiple alignment (distance scale: 0.1). The number of identified encapsulin systems per phylum is given below the phylogenetic tree. (B) Scheme depicting the general organization of encapsulin systems. The core operon consists of an encapsulin capsid gene (Enc) and a core cargo usually directly upstream. Other non-targeted but functionally associated components are found in proximity to of the core operon. Additional secondary cargo can be present throughout the genome. (C) Bar graph depicting the distribution of identified core cargo families in bacterial and archaeal phyla. The number of systems belonging to each identified core cargo type as well as the distribution of encapsulin systems in bacteria and archaea are shown as pie graphs.

741

742
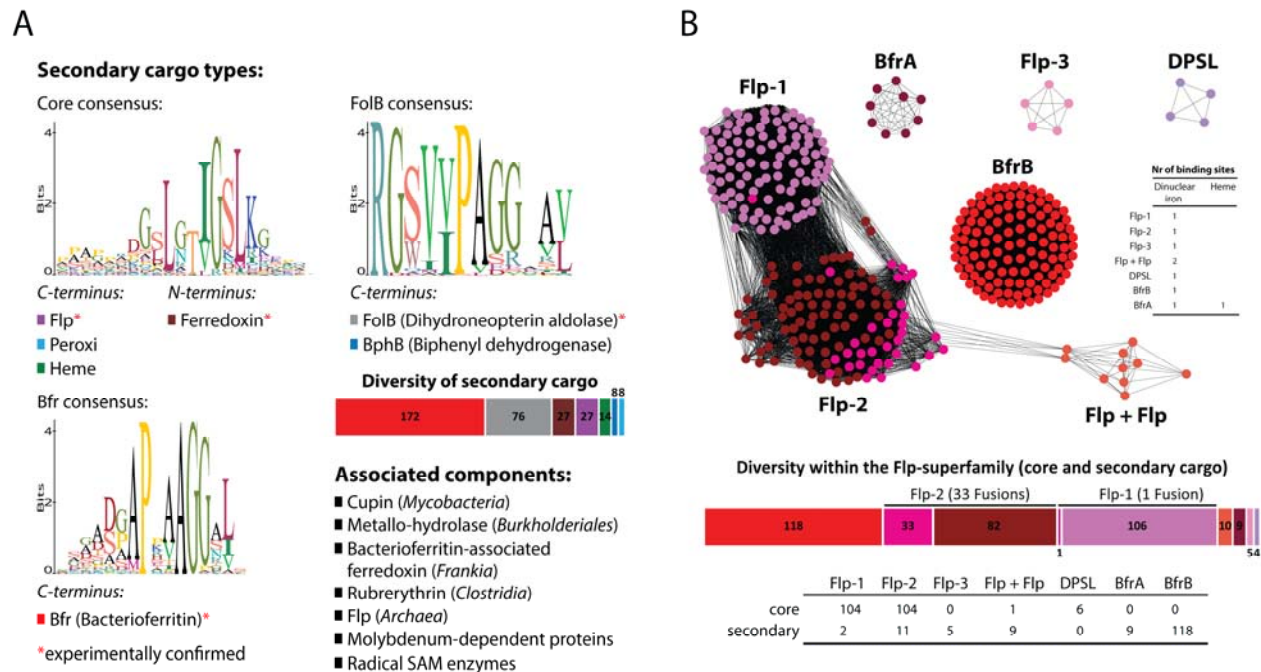
Figure 2: Secondary cargo, associated components and diversity of Flp-like cargo in encapsulin systems. (A) Shown are sequence logos of the 3 targeting peptide consensus sequences used to identify secondary cargo in all encapsulin containing genomes. The y-axis denotes positional information in bits. The secondary cargo types identified using a given consensus as a query are listed below the respective sequence logos. Experimentally confirmed secondary cargo types (either in this study or previously) are indicated with a red asterisk. The number of identified instances of different secondary cargo types are shown as a color-coded bar graph. A list of associated components identified in certain genera (shown in parentheses) is shown. (B) Sequence similarity network of all Flp-family cargo proteins identified in encapsulin systems (edge % identity: 39). The inset table indicates the presence and number of dinuclear iron centers and heme binding sites in each family. A bar graph depicts the number of proteins identified of each Flp-type with Flp-fusions depicted in pink. The bottom table shows the number of sequences of each Flp-family identified as core or secondary cargo.
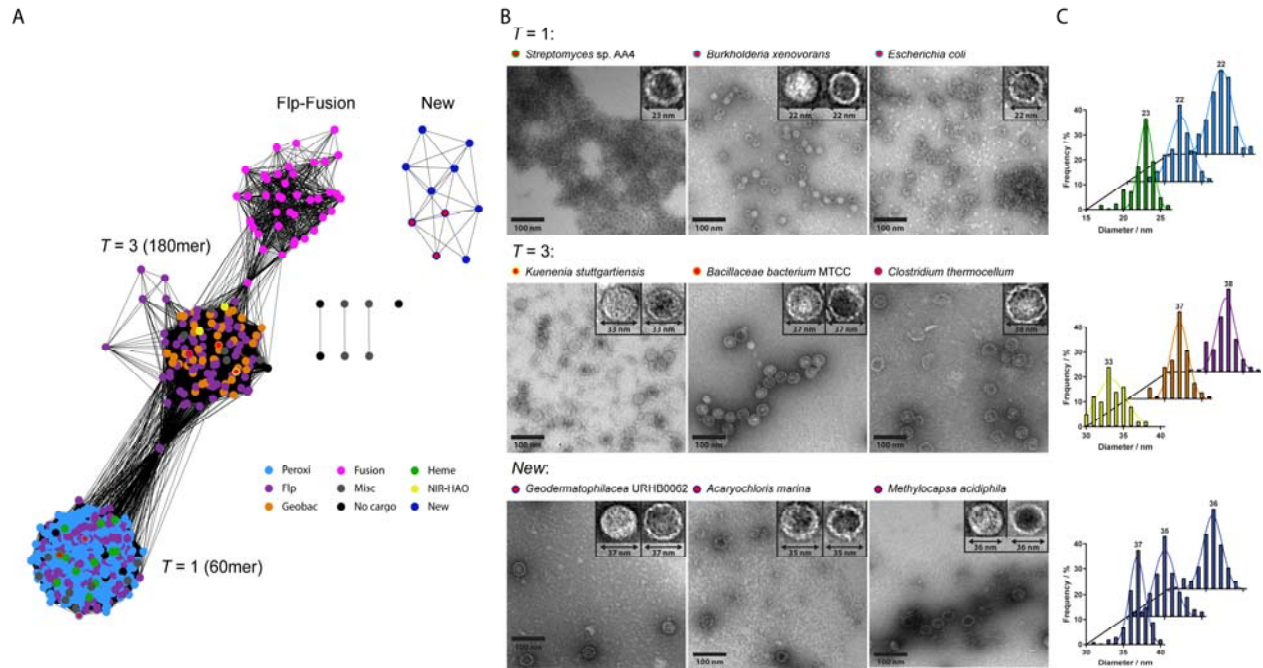
Figure 3: Classification and characterization of encapsulin capsids. (A) Sequence similarity network of all identified encapsulin capsid proteins (edge % identity: 24). Individual nodes are colored corresponding to their associated core cargo type. Red dots indicate the capsids chosen for experimental characterization. (B) Negative-stain TEM micrographs of 9 encapsulin capsids from the T=1, T=3 and New sequence clusters. Insets show magnified examples of each capsid. (C) Size-distribution of the characterized encapsulin capsids fitted using a Gaussian fit.
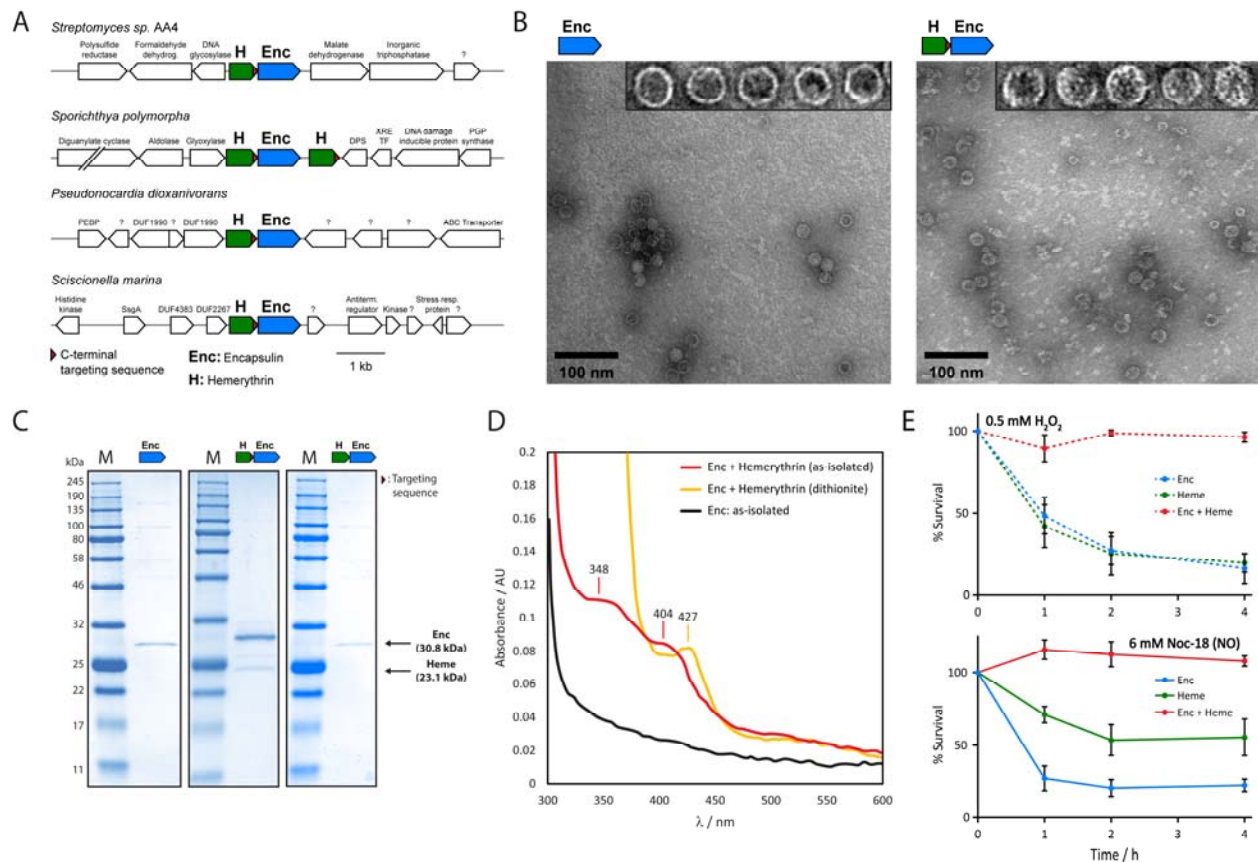
Figure 4: Characterization of the *Streptomyces* sp. AA4 Heme encapsulin system involved in oxidative and nitrosative stress resistance. (A) Selected Heme encapsulin gene clusters and their genetic surroundings. (B) Negative-stain TEM micrographs of purified encapsulin particles without (left) and with (right) Heme cargo. Insets show magnified particles. Additional protein puncta are visible when both Enc and Heme are present. (C) SDS-PAGE analysis of purified particles without (left) and with (middle) Heme. A mutant Heme missing the C-terminal TP does not co-purify with Enc particles (right). (D) UV-Vis spectroscopy of purified particles. (E) Survival of *E. coli* expressing Enc, Heme or Enc+Heme in the presence of H_2O_2 (top panel) or the NO-generator Noc-18 (lower panel). Data are represented as mean ± SD of 3 biological replicates.

782
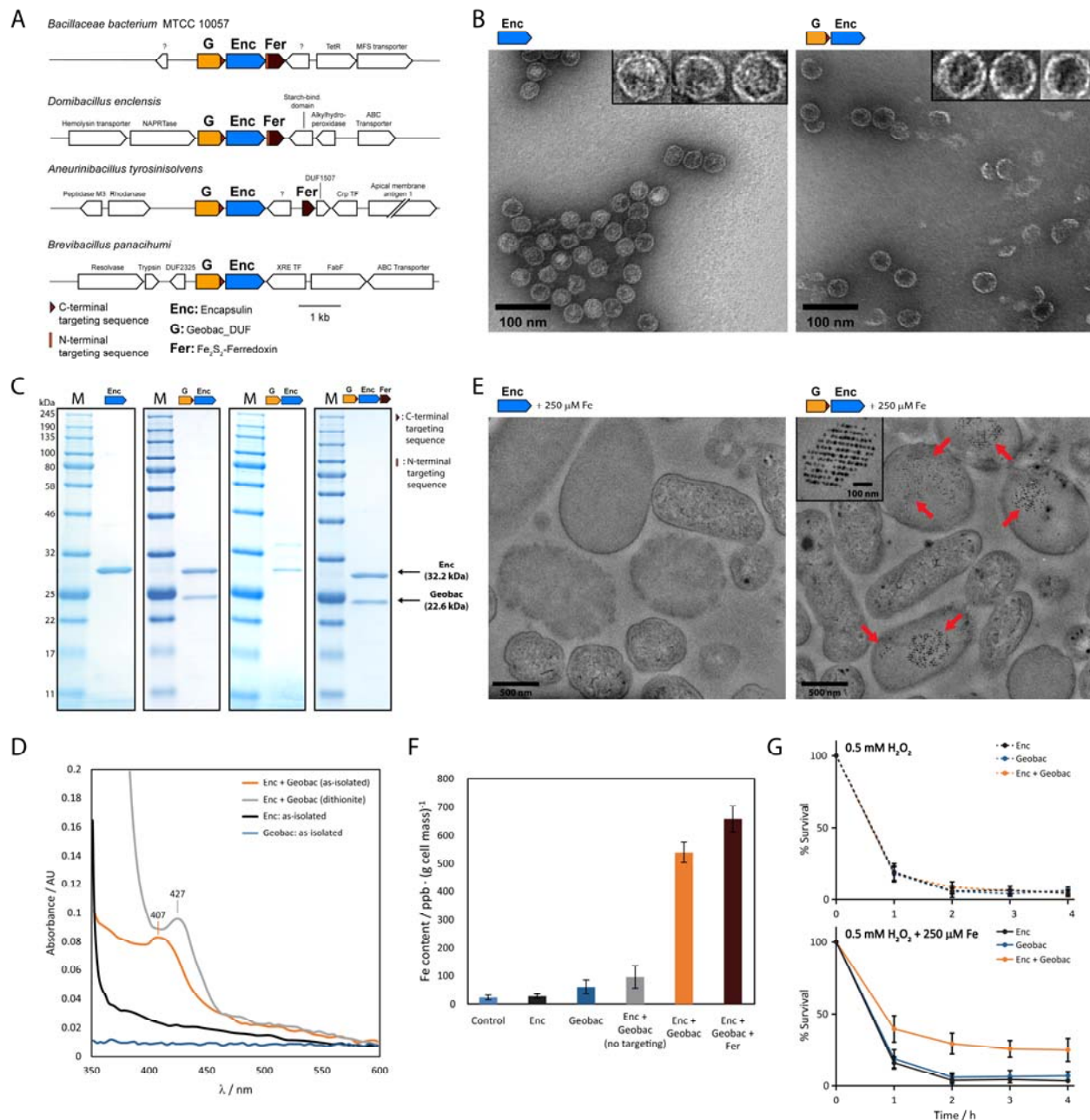
Figure 5: Characterization of the *Bacillaceae bacterium* MTCC 10057 Geobac encapsulin system involved in iron mineralization. (A) Selection of identified Geobac encapsulin gene clusters. (B) Negative-stain TEM micrographs of purified encapsulin particles resulting from *E. coli* cells expressing Enc alone or the Geobac core operon. Insets show magnified particles. (C) SDS-PAGE analysis of purified particles. The expressed gene/operon is shown above the respective gel. (D) UV-Vis analysis of purified

788 Geobac particles and controls. (E) TEM micrographs of fixed *E. coli* cells expressing Enc or Enc+Geobac in

789 the presence of iron sulfate. Intracellular electron-dense particles are indicated by red arrows. The inset

790 shows a magnification of regularly arranged electron-dense particles found in Enc+Geobac-expressing

791 cells. (F) ICP-MS analysis of washed *E. coli* cells expressing different components of the Geobac

792 encapsulin system. Data are represented as mean ± SD of 3 biological replicates. (G) Survival of *E. coli*

793 cells expressing Enc, Geobac or Enc+Geobac in the presence of H_2O_2 (top panel) or H_2O_2 and iron sulfate

794 (lower panel). Data are represented as mean ± SD of 3 biological replicates.
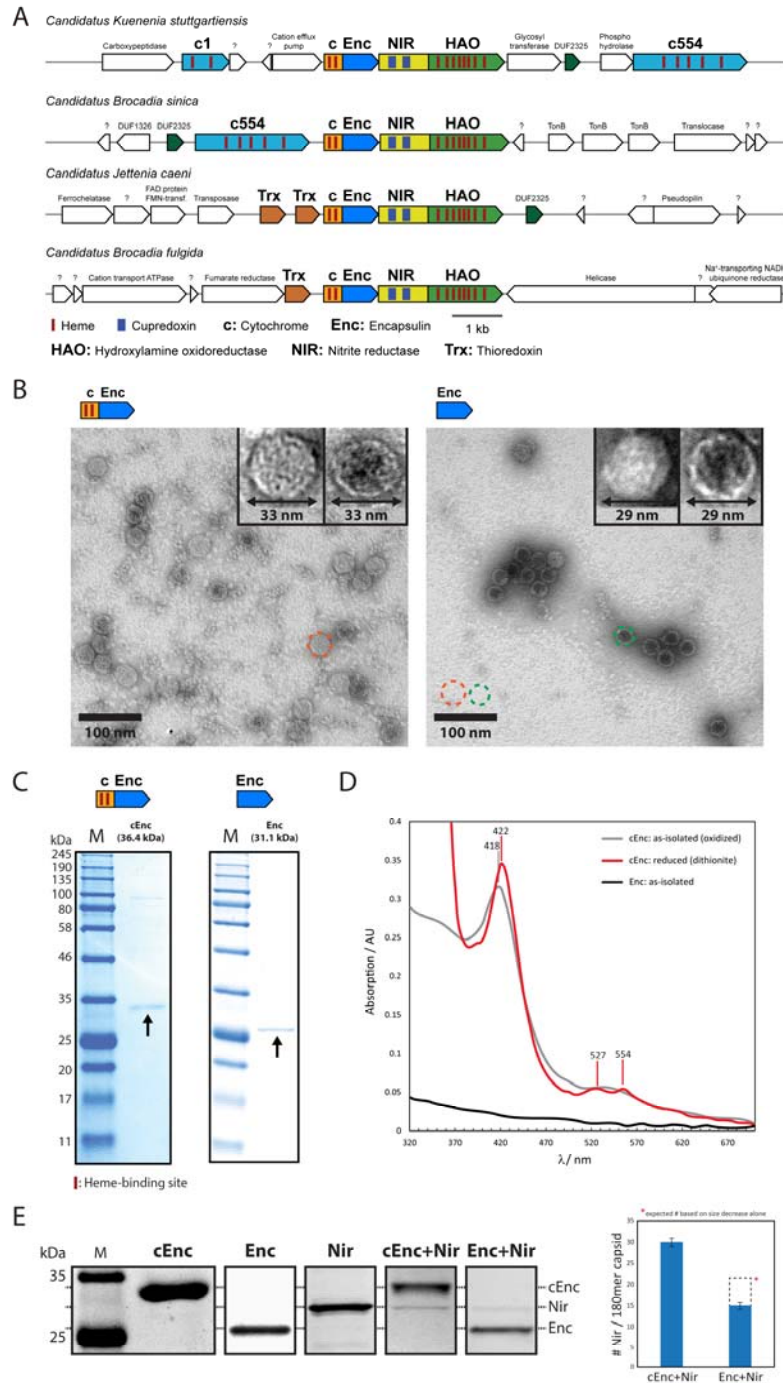
Figure 6: Characterization of the encapsulin system in the anammox bacterium *K. stuttgartiensis*. (A) Encapsulin operons found in 4 anammox bacteria. (B) TEM micrographs of purified *K. stuttgartiensis* encapsulins with (left) or without (right) N-terminal cytochrome fusion domain. Dotted circles are used to illustrate their size difference. (C) SDS-PAGE gels of cEnc and Enc. (D) UV-Vis spectroscopy of cEnc

36

800   showing the characteristic absorption pattern of c-type cytochromes. (E) SDS-PAGE gels of purified NIR

801   and different particles from *E. coli* expressing different components of the cEnc+NIR-HAO system

802   (indicated above each lane). The panel on the right shows the decrease in NIR loading when the N-

803   terminal cytochrome domain is missing. Data are represented as mean ± SD of 3 separate purifications

804   and gels.

A

Scale bar: 0.1

Legend:
- Actinobacteria: **428**
- Proteobacteria: **247**
- Firmicutes: **98**
- Crenarchaeotes (Archaea): **32**
- Thermotogae: **17**
- Spirochaetia: **14**
- Aquificae: **14**
- Euryarchaeotes (Archaea): **14**
- Synergistetes: **13**
- unclassified bacteria: **6**
- Fusobacteria: **6**
- Bacteroidetes: **5**
- Fibrobacteres/Acidobacteria: **4**
- Planctomycetes: **4**
- Cyanobacteria: **3**
- Deferribacteres: **2**
- Nitrospinae/Tectomicrobia: **1**
- Verrucomicrobia: **1**

B

◤ : Targeting sequence

Secondary cargo

Associated components

Core operon

Enc

Secondary cargo

C

Number of encapsulin systems per phylum

Phylum	Count
Actinobacteria	428
Proteobacteria	247
Firmicutes	98
Crenarchaeotes (Archaea)	32
Thermotogae	17
Spirochaetia	14
Aquificae	14
Euryarchaeotes (Archaea)	14
Synergistetes	13
Unclassified bacteria	6
Fusobacteria	6
Bacteroidetes	5
Fibrobacteres/Acidobacteria	4
Planctomycetes	4
Cyanobacteria	3
Deferribacteres	2
Verrucomicrobia	1
Nitrospinae/Tectomicrobia	1

Diversity of cargo types in core operons

Peroxi: 586
Flp: 189
Geobac: 44
Fusion: 40
Heme: 14
No cargo: 15
NIR-HAO: 4
Misc: 17

Legend: Peroxi, Flp, Geobac, Fusion, No cargo, Heme, NIR-HAO, Misc

Phylogenetic distribution of encapsulin systems

Archaea: 46
Bacteria: 863

A

Secondary cargo types:

Core consensus:

C-terminus: *N-terminus:*

- ■ Flp *
- ■ Peroxi
- ■ Heme

- ■ Ferredoxin *
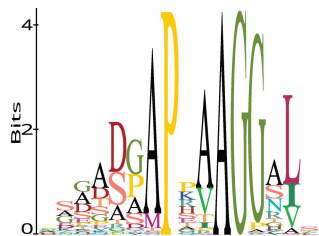
Bfr consensus:

C-terminus:

- ■ Bfr (Bacterioferritin) *

*experimentally confirmed

FolB consensus:

C-terminus:

- ■ FolB (Dihydroneopterin aldolase) *
- ■ BphB (Biphenyl dehydrogenase)
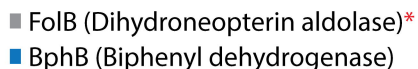
Diversity of secondary cargo

172	76	27	27	14	88

Associated components:

- ■ Cupin (*Mycobacteria*)
- ■ Metallo-hydrolase (*Burkholderiales*)
- ■ Bacterioferritin-associated ferredoxin (*Frankia*)
- ■ Rubrerythrin (*Clostridia*)
- ■ Flp (*Archaea*)
- ■ Molybdenum-dependent proteins
- ■ Radical SAM enzymes

B

Nr of binding sites

	Dinuclear iron	Heme
Flp-1	1	
Flp-2	1	
Flp-3	1	
Flp + Flp	2	
DPSL	1	
BfrB	1	
BfrA	1	1
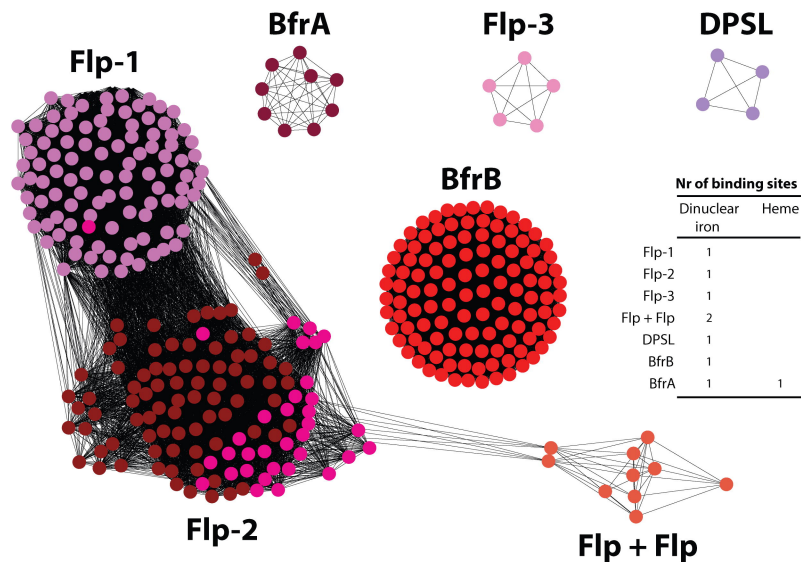
Diversity within the Flp-superfamily (core and secondary cargo)

118	Flp-2 (33 Fusions)		Flp-1 (1 Fusion)			
	33	82	106	10	9	54
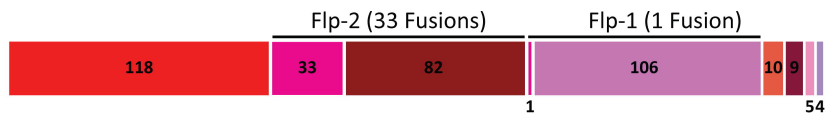
	Flp-1	Flp-2	Flp-3	Flp + Flp	DPSL	BfrA	BfrB
core	104	104	0	1	6	0	0
secondary	2	11	5	9	0	9	118

A

Flp-Fusion

New

T = 3 (180mer)

T = 1 (60mer)

Peroxi   Fusion   Heme
Flp   Misc   NIR-HAO
Geobac   No cargo New
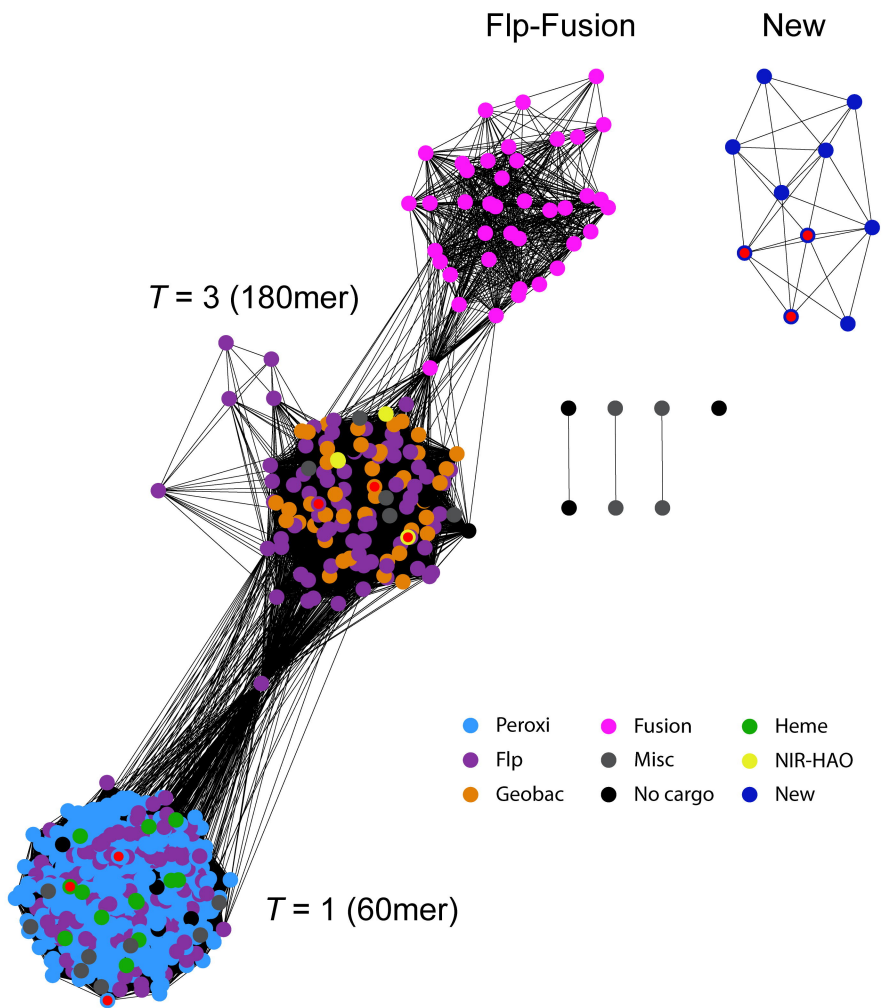
B

T = 1:

Streptomyces sp. AA4   *Burkholderia xenovorans*   *Escherichia coli*

100 nm   23 nm   100 nm   22 nm   22 nm   100 nm   22 nm

T = 3:
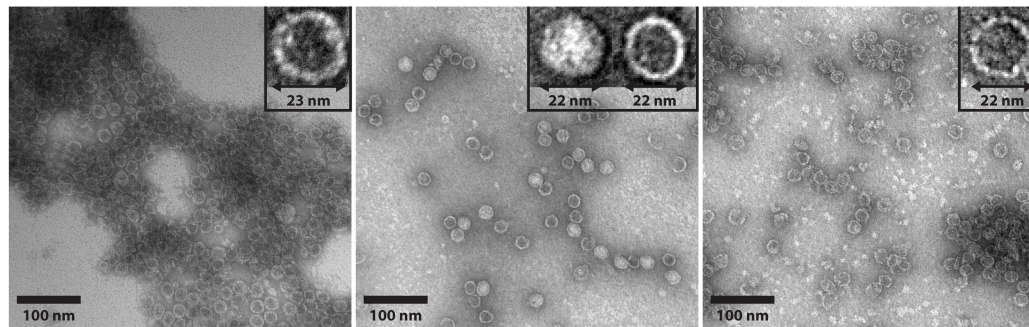
Kuenenia stuttgartiensis   *Bacillaceae bacterium* MTCC   *Clostridium thermocellum*

100 nm   33 nm   33 nm   100 nm   37 nm   37 nm   100 nm   38 nm
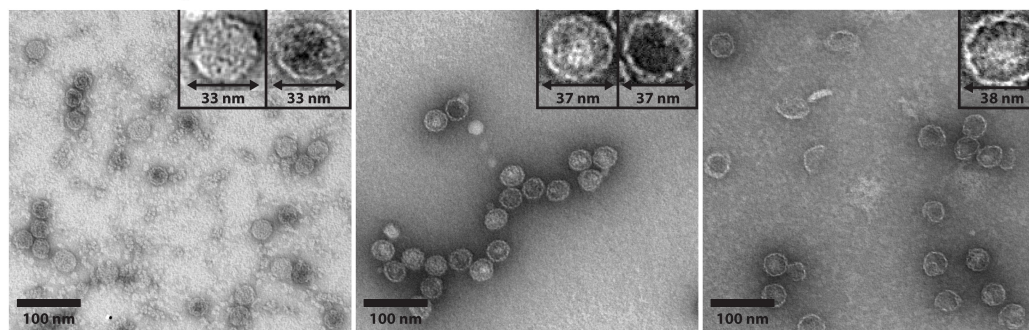
New:

Geodermatophilacea URHB0062   *Acaryochloris marina*   *Methylocapsa acidiphila*

100 nm   37 nm   37 nm   100 nm   35 nm   35 nm   100 nm   36 nm   36 nm

C

Frequency / %   23   22   22   Diameter / nm

Frequency / %   33   37   38   Diameter / nm

Frequency / %   37   35   36   Diameter / nm

A

Bacillaceae bacterium MTCC 10057

? G Enc Fer TetR MFS transporter

Domibacillus enclensis

Hemolysin transporter NAPRTase G Enc Fer Starch-bind. domain Alkylhydro-peroxidase ABC Transporter

Aneurinibacillus tyrosinisolvens

Peptidase M3 Rhodanase G Enc ? Fer DUF1507 Crp TF Apical membrane antigen 1

Brevibacillus panacihumi

Resolvase Trypsin DUF2325 G Enc XRE TF FabF ABC Transporter

▶ : C-terminal targeting sequence

▌ : N-terminal targeting sequence

Enc: Encapsulin
G: Geobac_DUF
Fer: Fe₂S₂-Ferredoxin

1 kb

B

Enc

G Enc

100 nm

100 nm

C

kDa M Enc M G Enc M G Enc M G Enc Fer

245
190
135
100
80
58
46
32
25
22
17
11

▶ : C-terminal targeting sequence

▌ : N-terminal targeting sequence

Enc (32.2 kDa)

Geobac (22.6 kDa)

D

Enc + Geobac (as-isolated)
Enc + Geobac (dithionite)
Enc: as-isolated
Geobac: as-isolated

407
427

Absorbance / AU
λ / nm

E

Enc + 250 μM Fe

G Enc + 250 μM Fe

500 nm

500 nm

100 nm

F

Fe content / ppb · (g cell mass)⁻¹

Control Enc Geobac Enc + Geobac (no targeting) Enc + Geobac Enc + Geobac + Fer

G

0.5 mM H₂O₂

Enc
Geobac
Enc + Geobac

% Survival

0.5 mM H₂O₂ + 250 μM Fe

Enc
Geobac
Enc + Geobac

% Survival

Time / h

A

Candidatus Kuenenia stuttgartiensis

Carboxypeptidase | **c1** | ? | ? | Cation efflux pump | **c Enc** | **NIR** | **HAO** | Glycosyl transferase | DUF2325 | Phospho hydrolase | **c554**

Candidatus Brocadia sinica

? | DUF1326 | DUF2325 | **c554** | **c Enc** | **NIR** | **HAO** | ? | TonB | TonB | TonB | Translocase | ? | ?

Candidatus Jettenia caeni

Ferrochelatase | ? | FAD:protein FMN-transf. | Transposase | **Trx** | **Trx** | **c Enc** | **NIR** | **HAO** | DUF2325 | ? | ? | Pseudopilin | ?

Candidatus Brocadia fulgida

? | ? | Cation transport ATPase | ? | Fumarate reductase | **Trx** | **c Enc** | **NIR** | **HAO** | Helicase | ? | Na⁺-transporting NADH: ubiquinone reductase

| | Heme | | Cupredoxin | **c:** Cytochrome | **Enc:** Encapsulin | 1 kb |

HAO: Hydroxylamine oxidoreductase   **NIR:** Nitrite reductase   **Trx:** Thioredoxin

B

c Enc   Enc

C

cEnc (36.4 kDa)   Enc (31.1 kDa)

|: Heme-binding site

D

cEnc: as-isolated (oxidized)
cEnc: reduced (dithionite)
Enc: as-isolated

418 422
527 554

Absorption / AU
λ / nm

E

kDa | M | **cEnc** | **Enc** | **Nir** | **cEnc+Nir** | **Enc+Nir**

*expected # based on size decrease alone

Nir / 180mer capsid

cEnc+Nir   Enc+Nir