SUPPLEMENTAL INFORMATION

SUPPLEMENTARY TEXT

Properties of GRM encapsulation peptides (EPs)

Glycyl radical enzymes. Secondary structure prediction of the putative N-terminal EPs of the GRM1 GREs yielded two consecutive α-helices instead of just one as is typically found in the EPs of other types of BMCs (1). Each of the helices is about 12 amino acids long. They are separated by ~4 residues predicted to be disordered, which may constitute a turn between the helices. A predicted unordered ~30 amino acid long linker connects the two helices to the N-terminal domain of the GRE. Similarly, secondary structure predictions for the insertion domains in the GRM3 and GRM5 GREs yielded two central α-helices, ~15 residues each. The helices appear to be connected to one another via a short (~5 residues) loop and are flanked by unordered, poorly conserved linkers (15 residues each on average). In the case of the GRM3 GREs only the first of the two helices shows the characteristic amphipathicity of EPs, whereas the second helix does not. Thus, the insertion in the GRM3 GREs resembles the N-terminal extension of the GRM1 GREs. In contrast, in the GRM5 GREs both helices of the insertion appear to be amphipathic. The GRM4 GREs provide yet another variation by containing only one amphipathic α-helix of 15 residues length within this insertion.

Aldehyde dehydrogenases. The GRM1 and GRM2 AldDHs exhibit C-terminal extensions of ~50-100 and ~80-100 residues, respectively. The characteristic predicted amphipathic helix of an EP is found at the ends of these extension. In contrast, the GRM3-5 AldDHs contain N-terminal extensions ranging from 30 up to 100 residues in length. These extensions exhibit the characteristics of EPs, *i.e.* long flexible linkers (usually 10-20 residues) and predicted amphipathic α-helices of 12-15 residues (1). Some GRM3 AldDHs (for example from *Rhodopseudomonas palustris* and *Rhodosprillum rubrum*) contain two amphipathic helices (12-15 residues each), which are separated from one another

by long linkers (~20 residues) in the N-terminal extension. This type of EP resembles the ones found in the GRM1, 3, and 5 GREs, which also comprise two α-helices and appears to be a unique feature of GRMs in comparison to other characterized BMCs.

Loci encoding GREs of Unknown Function (GUF)

Genomes of the GRM1 encoding organisms *D. dehalogenans, D. hafniense* DCB-2 & Y51, and *D. psychrophila* also contain an additional BMC gene cluster that we designate GUF loci. The GREs in these loci do not contain any obvious EPs, and are oriented in the opposite direction of the rest of the genes in the loci (except in *D. psychrophila*); all other known BMC loci encode their signature enzymes and other components in the same orientation (2). Phylogenetically, the GUFs appear to be more closely related to non-encapsulated GREs (Fig. 2A). Given these observations, we predict that these GUFs are not packaged into BMCs. Nevertheless, the GUF loci appear to encode all other required interior constituents of a functional BMC, with the exception of an alcohol dehydrogenase gene. Interestingly, the GRM1 locus of *D. psychrophila* only encodes a non-functional AldDH, whereas the GUF locus harbors genes for two (seemingly functional) AldDHs. They cluster together with EutE from *Salmonella enterica*, the AldDH of the canonical EUT-BMC (Fig. 2D in the main text). This is due to an apparent partial duplication within this GUF locus. All other GUF loci only encode one functional AldDH, which phylogenetically cluster closely together with AldDHs of GRM1 loci. Moreover, the GUF loci in *D. dehalogenans, D. hafniense* DCB-2 and Y51 only encode two BMC-Hs (Dataset S1) meaning these GUF loci may not form functional organelles on their own. However, it is possible that the genes in the GRM1 loci could complement the deficiencies of the GUF loci and their co-expression results in a functional metabolosome. The situation appears to be reversed in *D. psychrophila*, where the GRM1 locus lacks a BMC-P gene and only encodes a single BMC-H, whereas the GUF locus encodes six BMC-Hs as well as a BMC-P. Collectively, these observations support the

hypothesis that cooperative expression of both GRM1 and GUF loci in *D. dehalogenans, D. hafniense* DCB-2 & Y51, as well as *D. psychrophila* is required for the formation of a functional BMC.

Locus type reclassifications

As mentioned in the main text three loci (in *C. ljungdahlii*, *D. meridiei*, and *D. orientis*) that were previously annotated as belonging to the GRM3 type (2) were re-classified as GRM1 based on the phylogenetic analysis for the signature as well as the core enzymes. In detail, the GREs, the AEs, the AldDHs, and the PTACs of the corresponding loci cluster together with their counterparts of the GRM1 loci in the respective phylogenetic trees. Moreover, the corresponding loci in *C. ljungdahlii*, *D. meridiei* also encode an additional 'dud' AldDH. All three loci harbor a predicted choline/ethanolamine kinase gene, further supporting their reclassification as GRM1 loci.

Ancillary Enzymes/Proteins of GRM Loci

In addition to signature and core enzymes, BMC loci typically encode additional proteins/enzymes that are likely not directly involved in the degradation of a substrate, but instead provide supporting functions, such as co-factor recycling, and integration of the BMC function with the cytoplasm of the host and its environment (2, 3). These include genes encoding putative zinc-containing alcohol dehydrogenases in GRM5 loci, acyl kinases in GRM3 loci, putative S-adenosylmethionine synthetases in a subset of GRM3 loci, (putative) cobalamin reductases, a EutJ homolog (in most GRM1 and GRM3 loci), and various putative metabolite transporters of the cell membrane. Each GRM locus also appears to have its own (set of) transcription factor(s). The GRM1 loci encode a transcription factor of the MerR family, with a small molecule binding domain. GRM2 loci encode regulatory proteins of the TetR and LysR families. GRM3 and GRM4 loci encode

regulatory proteins of the PocR and AraC families with domains associated with two-component signaling systems. GRM5 loci have a transcription factor of the DeoR family.

Incomplete Shell Compositions of GRM Loci

The GRM1 loci of *D. psychrophila* and *Clostridium tetani* lack genes encoding BMC-P orthologs, the pentameric proteins that presumably cap the vertices of an icosahedral shell. Furthermore, those two loci encode only one or two BMC-H proteins, respectively, which is a relatively low number; the average number of genes encoding BMC-H proteins per BMC locus is 3.5 (2) and up to 5.5 for GRM1 loci. However, *D. psychrophila* also harbors a GUF locus containing two BMC-P genes, an apparent duplication. It is, therefore, possible that in *D. psychrophila* the GUF locus complements the GRM1 locus. However, no genes encoding BMC-Ps can be found anywhere in the genome of *C. tetani* which otherwise has all of the GRM1 lumen protein genes intact.

GRM fusion loci

GRM1/GRM3 fusion. The two distinct GRM loci present in the genome of *C. ljungdahlii* may instantiate an intermediate stage in the evolution of GRMs. It was surprising that the GRE and AE from the two loci, fell within different phylogenetic clades (GRM1 or GRM3), while the AldDH and PTAC from both loci were nested within the GRM1-associated clades. Upon closer inspection, it is apparent that two sections of the locus that contained a gene encoding the choline TMA-lyase type GRE (GRM1) had surprisingly high sequence identity (98%) with a large segment of the other locus containing the propanediol dehydratase type GRE (Fig. S6). This duplication apparently includes genes that are critical for GRM structure and function, such as those coding for BMC-H and BMC-P proteins, an AldDH, as well as a PTAC. Given the phylogenetic distribution of the GRE, AE, AldDH and PTAC, we inferred that a portion of a GRM1 locus was duplicated and inserted adjacent to an incomplete

GRM3 locus, thereby completing the locus as a whole structurally and enzymatically (Fig. S6). Two genes in the duplicated region appear to have been lost; interestingly, one of them is a BMC-T protein (Fig. S6). Since the incomplete GRM3 locus contained a BMC-T protein, it may be that the GRM1 BMC-T was redundant and eventually lost. Whether or not this hybrid GRM1/GRM3 locus is functional can only be determined via experimentation. We expect that the GRM1-derived AldDH and PTAC exhibit some promiscuity, accepting propionaldehyde and propionyl-CoA as substrates, respectively (see Fig. 1B for comparison). Their enzymatic efficiency will likely improve over time due to selective pressures. Based on the sequence identity, this duplication/transfer event must have happened very recently, providing us a snapshot of a potential mechanism for functional divergence of GRM loci.

PDU/GRM2 fusion. A PDU/GRM2 fusion locus in *Escherichia fergusonii* was found to contain genes encoding a GRM2 GRE (see Fig. 2A), its activating enzyme, all of the subunits of a B_{12}-dependent propanediol dehydratase, as well as two genes encoding PduL (phosphotransacylase) homologs. One of these PduL homologs clusters together with GRM2 PTACs, whereas the other one was closely related to PduL from the PDU locus of *Salmonella enterica* (Fig. 2E). These observations may indicate that this PDU/GRM2 locus has a dual-function, being used for the simultaneous degradation of propanediol and choline.

SUPPLEMENTAL REFERENCES

1. **Aussignargues C, Paasch BC, Gonzalez-Esquer R, Erbilgin O, Kerfeld CA.** 2015. Bacterial Microcompartment Assembly: The Key Role of Encapsulation Peptides. Commun Integr Biol **8:**e1039755.

2. **Axen SD, Erbilgin O, Kerfeld CA.** 2014. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. PLoS Comput Biol **10**.

3. **Kerfeld CA, Erbilgin O.** 2015. Bacterial microcompartments and the modular construction of microbial metabolism. Trends Microbiol **23:**22-34.

Table S1. BMC associated glycyl radical enzymes included in this study

Species	Locus type	Accession
Alkaliphilus metalliredigens QYMF	GRM1	YP_001321612
Alkaliphilus oremlandii OhILAs	GRM1	YP_001513926
Clostridium botulinum A str. ATCC 3502	GRM1	YP_001254614
Clostridium botulinum A3 str. Loch Maree	GRM1	YP_001787508
Clostridium botulinum B str. Eklund 17B	GRM1	YP_001885647
Clostridium botulinum B1 str. Okra	GRM1	YP_001781665
Clostridium botulinum Ba4 str. 657	GRM1	YP_002863094
Clostridium botulinum BKT015925	GRM1	YP_004396407
Clostridium botulinum E3 str. Alaska E43	GRM1	YP_001920768
Clostridium botulinum F str. Langeland	GRM1	YP_001391412
Clostridium botulinum H04402 065	GRM1	YP_005678666
Clostridium phytofermentans ISDg	GRM1	YP_001558531
Clostridium saccharolyticum WM1	GRM1	YP_003824152
Clostridium tetani E88	GRM1	NP_782070
Desulfitobacterium dehalogenans ATCC 51507	GRM1	YP_006432251
Desulfitobacterium hafniense DCB-2	GRM1	YP_002461335
Desulfitobacterium hafniense Y51	GRM1	YP_521239
Desulfosporosinus acidiphilus SJ4	GRM1	YP_006465906
Desulfotalea psychrophila LSv54	GRM1	YP_065559
Desulfotomaculum reducens MI-1	GRM1	YP_001114603
Desulfotomaculum ruminis DSM 2154	GRM1	YP_004545628
Desulfovibrio alaskensis G20	GRM1	YP_389771
Desulfovibrio desulfuricans ATCC 27774	GRM1	YP_002479937
Desulfovibrio hydrothermalis AM13	GRM1	YP_007325732
Desulfovibrio salexigens DSM 2638	GRM1	YP_002990074
Olsenella uli DSM 7084	GRM1	YP_003800310
Streptococcus dysgalactiae ATCC 12394	GRM1	YP_006013997
Streptococcus dysgalactiae RE378	GRM1	YP_006860442
Streptococcus iniae SF1	GRM1	YP_008055971
Streptococcus intermedius ATCC 27335	GRM1	EPH05251
Streptococcus intermedius JTH08	GRM1	YP_006469536
Clostridium ljungdahlii DSM 13528	GRM1*	YP_003782109
Desulfosporosinus meridiei DSM 13257	GRM1*	YP_006624072
Desulfosporosinus orientis DSM 765	GRM1*	YP_004973312
Aeromonas hydrophila ML09-119	GRM2	YP_008042577
Aeromonas hydrophila ATCC 7966	GRM2	YP_855870
Enterobacter aerogenes KCTC 2190	GRM2	YP_004592209
Escherichia coli 536	GRM2	YP_672437
Escherichia coli IAI39	GRM2	YP_002410085
Escherichia coli O7:K1 str. CE10	GRM2	YP_006146235
Escherichia coli UTI89	GRM2	YP_543898
Klebsiella oxytoca E718	GRM2	YP_006496423
Klebsiella oxytoca KCTC 1686	GRM2	YP_005017919
Klebsiella pneumoniae 342	GRM2	YP_002240668
Klebsiella variicola At-22	GRM2	YP_003441426
Proteus mirabilis HI4320	GRM2	YP_002152417
Raoultella ornithinolytica B6	GRM2	YP_007875383
Serratia marcescens FGI94	GRM2	YP_007343022
Shimwellia blattae DSM 4481	GRM2	YP_006318380
Vibrio furnissii NCTC 11218	GRM2	YP_004991566
Clostridium beijerinckii NCIMB 8052	GRM3	YP_001311127
Clostridium cf. saccharolyticum K10	GRM3	YP_007849774
Clostridium ljungdahlii DSM 13528	GRM3	YP_003779353
Clostridium novyi NT	GRM3	YP_878969
Escherichia coli APEC O1	GRM3	YP_859768
Escherichia coli CFT073	GRM3	NP_756397
Escherichia coli ED1a	GRM3	YP_002399331
Escherichia coli S88	GRM3	YP_002394164
Oscillibacter valericigenes Sjm18-20	GRM3	YP_004881456
Pectobacterium wasabiae WPP163	GRM3	YP_003261427
Raoultella ornithinolytica B6	GRM3	YP_007872971
Rhodobacter capsulatus SB 1003	GRM3	YP_003578348
Rhodopseudomonas palustris BisB18	GRM3	YP_531045
Rhodospirillum rubrum ATCC 11170	GRM3	YP_425991
Rhodospirillum rubrum F11	GRM3	YP_006047199
Vibrio sp. EJY3	GRM3	YP_005023152
Shewanella putrefaciens CN-32	GRM4	YP_001181743
Shewanella sp. W3-18-1	GRM4	YP_961832
Clostridium phytofermentans ISDg	GRM5	YP_001558291
Roseburia inulinivorans DSM 16841	GRM5	EEG94456
Ruminococcus obeum A2-162	GRM5	YP_007805196
Ruminococcus sp. SR1/5	GRM5	YP_007783755
Ruminococcus torques L2-14	GRM5	YP_007786118
Desulfitobacterium dehalogenans ATCC 51507	GUF[#]	YP_006428626
Desulfitobacterium hafniense DCB-2	GUF[#]	YP_002456869
Desulfitobacterium hafniense Y51	GUF[#]	YP_516649
Desulfotalea psychrophila LSv54	GUF[#]	YP_066763
Escherichia fergusonii ATCC 35469	PDU/GRM	YP_002383121

*Reclassified, previously GRM3. [#]Reclassified, previously GRM1.

Table S2. GRM loci lacking genes encoding the iron-containing type of alcohol dehydrogenase (pfam00465).

GRM loci types	Species
GRM1	*Desulfosporosinus acidiphilus* SJ4
GRM1 (GRM3 outlier)	*Clostridium ljungdahlii DSM 13528*
GRM1 (GRM3 outlier)	*Desulfosporosinus meridiei* DSM 13257
GRM1 (GRM3 outlier)	*Desulfosporosinus orientis* DSM 765
GRM1 & GUF	*Desulfitobacterium dehalogenans* ATCC 51507
GRM1 & GUF	*Desulfitobacterium hafniense* DCB-2
GRM1 & GUF	*Desulfitobacterium hafniense* Y51
GRM1 & GUF	*Desulfotalea psychrophila* LSv54
GRM2	*Escherichia coli* IAI39
GRM5	*Clostridium phytofermentans* ISDg
GRM5	*Roseburia inulinivorans* DSM 16841

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. GRE active site comparisons. A.) Stereo view superposition of a GRM3 GRE (*R. palustris* BisB18) homology model with the crystal structure of the glycerol-bound form of the B_{12}-independent glycerol dehydratase (GDH) (PDB 1R9D). GRM3 GRE residues that are divergent from the ones in the GDH are colored in orange. The other residues are conserved in both active sites. Residues are numbered according to the GDH structure. Cysteine 433 is the residue that harbors the radical during catalysis. Serine 642 is replaced by a valine residue in all GRM3, GRM4, and GRM5 GREs. This valine would form an unfavorable short contact to the hydroxyl group of glycerol. This hydroxyl group is absent in propanediol. Tyrosine 339 of the GDH is replaced by a phenylalanine in all the GRE3-5, providing a more hydrophobic for the terminal methyl group of propanediol together with the conserved valine residue. **B.)** Stereo view superposition of homology models for the GRM1 GRE choline trimethylamine-lyase (CHL) from *D. alaskensis* G20 (blue) and the GUF of *D. hafniense* Y51 (orange). The first letters and the numbering correspond to the residues of the CHL. The second letters represents the residues that differ in the GUF. All residues shown are in close proximity to the catalytic radical cysteine (C489 in the CHL), and are conserved in the CHLs or GUFs, respectively. The only exception is threonine 431, which can be replaced by serine in some of the GRM1 GREs.

Figure S2. Homology model of the GRM2 GRE and schematics of possible protein interactions. A.) Dimer of the B_{12}-independent glycerol dehydratase (GDH) (PDB 1R9D). **B.)** Homology model (created with RaptorX) of the GRM2 GRE that includes the N-terminal ~350 residues long extension (outlined in red). The modeled extension resembles the first part of the catalytic domain of the GRE (outlined in yellow). The latter is shown in the same orientation as the dark blue subunit of the GDH in panel A. **C.)** Schematic for a possible coalescence of the GRM2 GRE within the microcompartment.

The coalescence may be facilitated by the N-terminal partial catalytic domain, by replacing another catalytic domain in the dimer interface. **D.)** Schematic for a possible role of the N-terminal extension domain in the interaction with the BMC shell, similar to encapsulation peptides. Note that in contrast to the other BMC associated GREs, the GRM2 GREs do not contain any apparent encapsulation peptides.

Figure S3. Alignment of the insertion domains within the activating enzymes containing additional Fe-S cluster motifs. Sequence alignment used to construct the phylogeny of the activating enzymes (Figure 2B) depicting the insertion domains that comprise the ferredoxin-like [4Fe-4S] cluster motifs. Cysteine residues are highlighted in orange. Species names and GRM classifications are provided on the left, whereas the total number of [4Fe-4S] clusters present in the respective activating enzyme is noted on the right.

Figure S4. Phylogenetic tree for the GRM prevalent iron containing type of alcohol dehydrogenases (ADHs). The Maximum Likelihood tree of the ADHs was inferred from 67 amino acid sequences. The counterparts from the canonical PDU and EUT BMCs, PduQ and EutG, respectively, have been included for comparison. Bootstrap values for important nodes are represented as filled circles (above 50%) and empty circles (below 50%).

Figure S5. Phylogenetic trees of BMC-H, BMC-T and BMC-P shell proteins

Maximum Likelihood trees of the BMC-H, BMC-T, and BMC-P proteins were inferred from 354, 67 and 82 amino acid sequences, respectively. Bootstrap values for important nodes are represented as filled circles (above 50%) and empty circles (below 50%).

Figure S6. Proposed evolution of the *C. ljungdahlii* GRM locus. The borders of duplication and deletion were determined based on dot plot analysis (Figure S1). The order of operations was inferred based on the observation that the PduL homolog and AldDH in both loci were nested within the GRM1 clades in their respective phylogenetic trees. CHL – choline trimethylamine-lyase, PDH - propanediol dehydratase.

SUPPLEMENTAL FIGURES

Figure S1

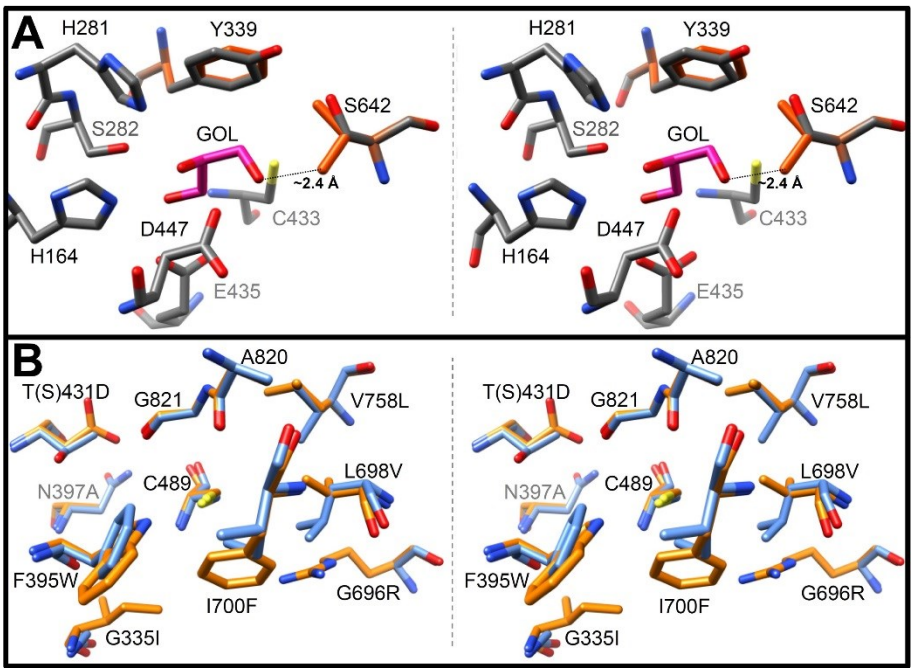


Figure S2.

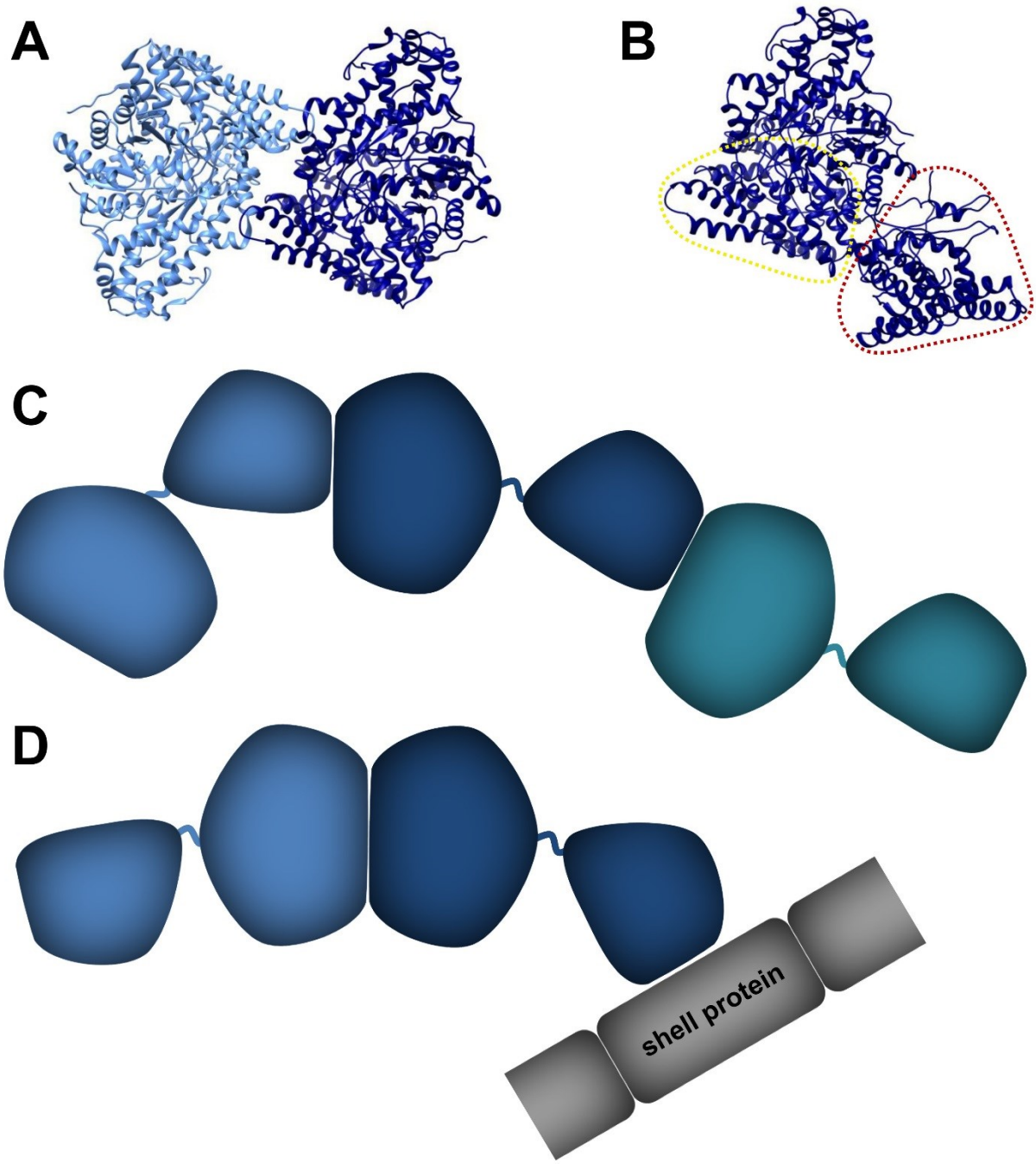


Figure S3.

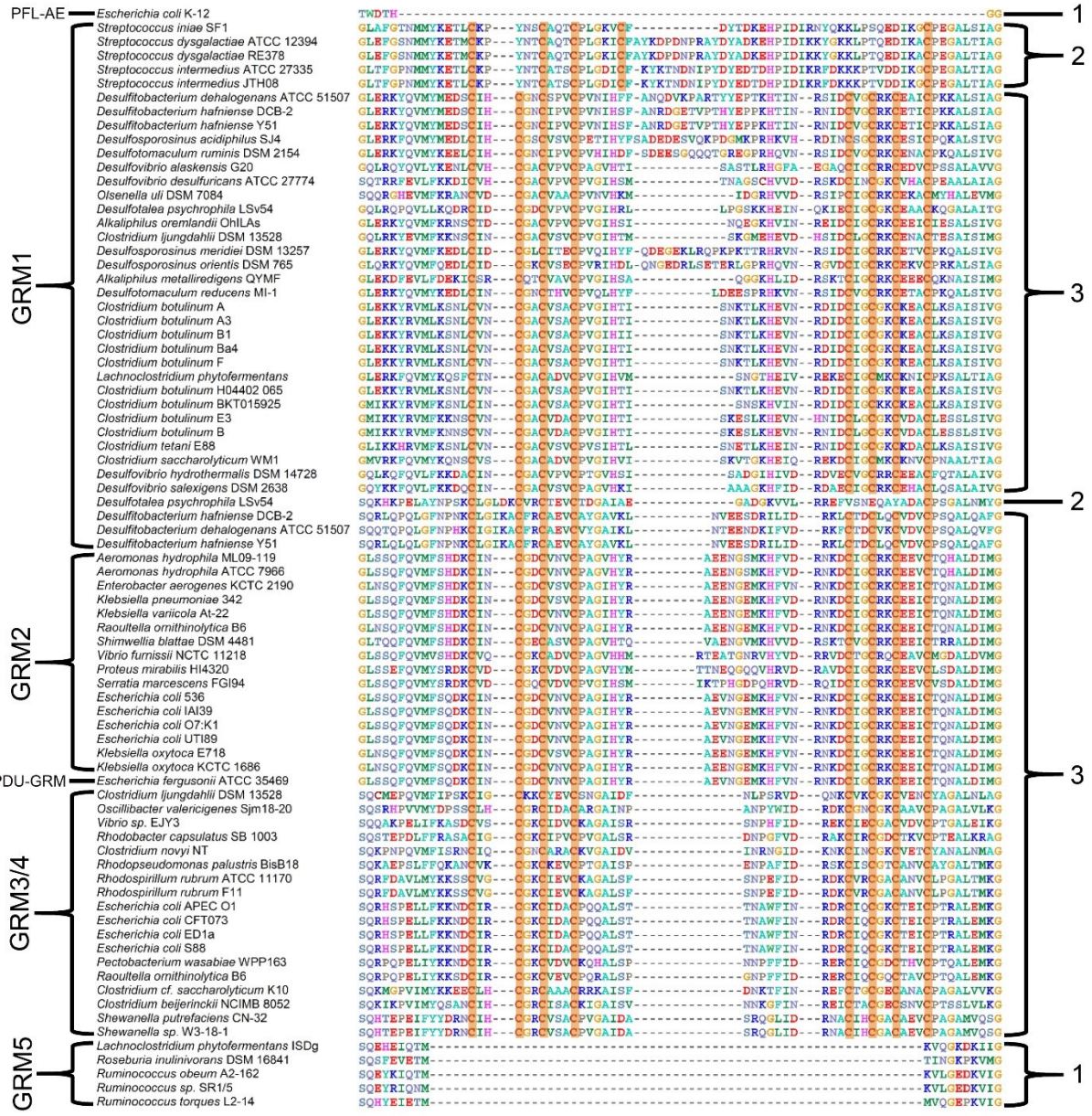

PFL-AE
GRM1
GRM2
PDU-GRM
GRM3/4
GRM5
1
2
3
2
3
1
Escherichia coli K-12
Streptococcus iniae SF1
Streptococcus dysgalactiae ATCC 12394
Streptococcus dysgalactiae RE378
Streptococcus intermedius ATCC 27335
Streptococcus intermedius JTH08
Desulfitobacterium dehalogenans ATCC 51507
Desulfitobacterium hafniense DCB-2
Desulfitobacterium hafniense Y51
Desulfosporosinus acidiphilus SJ4
Desulfotomaculum ruminis DSM 2154
Desulfovibrio alaskensis G20
Desulfovibrio desulfuricans ATCC 27774
Olsenella uli DSM 7084
Desulfotalea psychrophila LSv54
Alkaliphilus oremlandii OhILAs
Clostridium ljungdahlii DSM 13528
Desulfosporosinus meridiei DSM 13257
Desulfosporosinus orientis DSM 765
Alkaliphilus metalliredigens QYMF
Desulfotomaculum reducens MI-1
Clostridium botulinum A
Clostridium botulinum A3
Clostridium botulinum B1
Clostridium botulinum Ba4
Clostridium botulinum F
Lachnoclostridium phytofermentans
Clostridium botulinum H04402 065
Clostridium botulinum BKT015925
Clostridium botulinum E3
Clostridium botulinum B
Clostridium tetani E88
Clostridium saccharolyticum WM1
Desulfovibrio hydrothermalis DSM 14728
Desulfovibrio salexigens DSM 2638
Desulfotalea psychrophila LSv54
Desulfitobacterium hafniense DCB-2
Desulfitobacterium dehalogenans ATCC 51507
Desulfitobacterium hafniense Y51
Aeromonas hydrophila ML09-119
Aeromonas hydrophila ATCC 7966
Enterobacter aerogenes KCTC 2190
Klebsiella pneumoniae 342
Klebsiella variicola At-22
Raoultella ornithinolytica B6
Shimwellia blattae DSM 4481
Vibrio furnissii NCTC 11218
Proteus mirabilis HI4320
Serratia marcescens FGI94
Escherichia coli 536
Escherichia coli IAI39
Escherichia coli O7:K1
Escherichia coli UTI89
Klebsiella oxytoca E718
Klebsiella oxytoca KCTC 1686
Escherichia fergusonii ATCC 35469
Clostridium ljungdahlii DSM 13528
Oscillibacter valericigenes Sjm18-20
Vibrio sp. EJY3
Rhodobacter capsulatus SB 1003
Clostridium novyi NT
Rhodopseudomonas palustris BisB18
Rhodospirillum rubrum ATCC 11170
Rhodospirillum rubrum F11
Escherichia coli APEC O1
Escherichia coli CFT073
Escherichia coli ED1a
Escherichia coli S88
Pectobacterium wasabiae WPP163
Raoultella ornithinolytica B6
Clostridium cf. saccharolyticum K10
Clostridium beijerinckii NCIMB 8052
Shewanella putrefaciens CN-32
Shewanella sp. W3-18-1
Lachnoclostridium phytofermentans ISDg
Roseburia inulinivorans DSM 16841
Ruminococcus obeum A2-162
Ruminococcus sp. SR1/5
Ruminococcus torques L2-14

Figure S4.

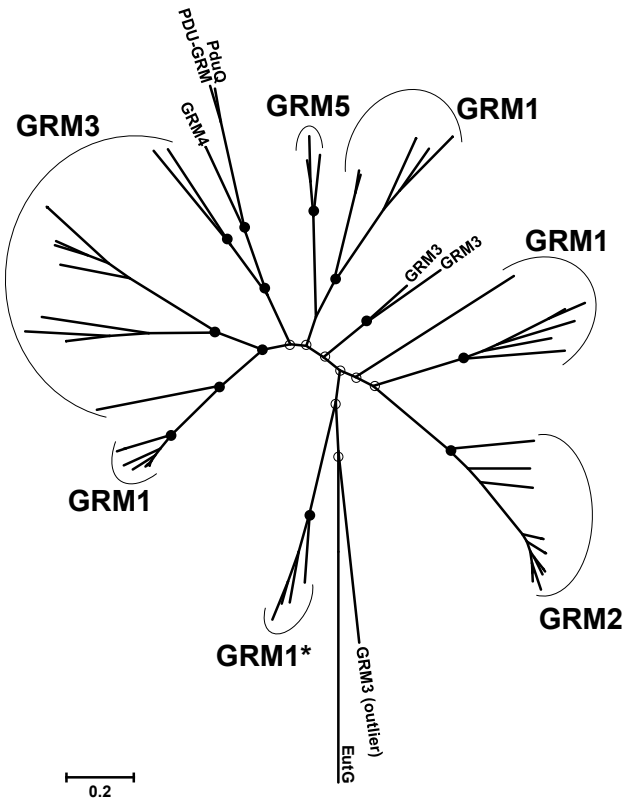

PduQ
PDU-GRM
GRM4
GRM3
GRM5
GRM1
GRM3
GRM3
GRM1
GRM1
GRM1*
GRM3 (outlier)
EutG
GRM2
0.2

Figure S5.

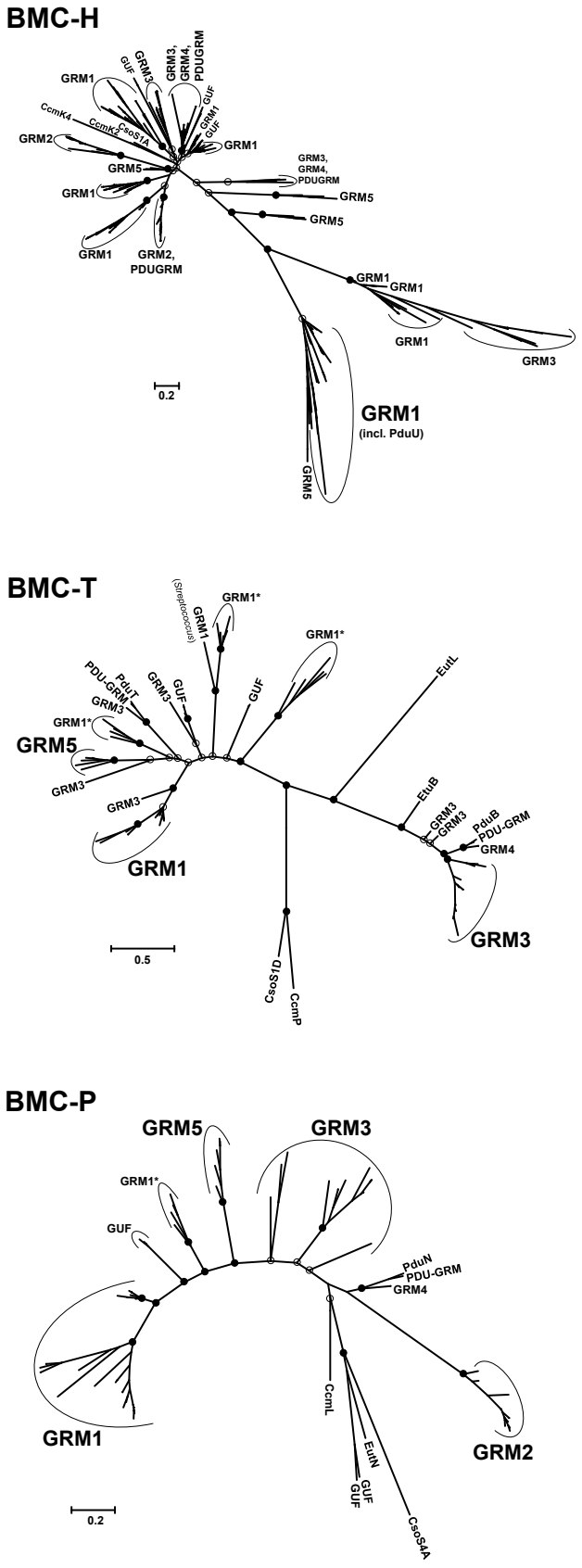
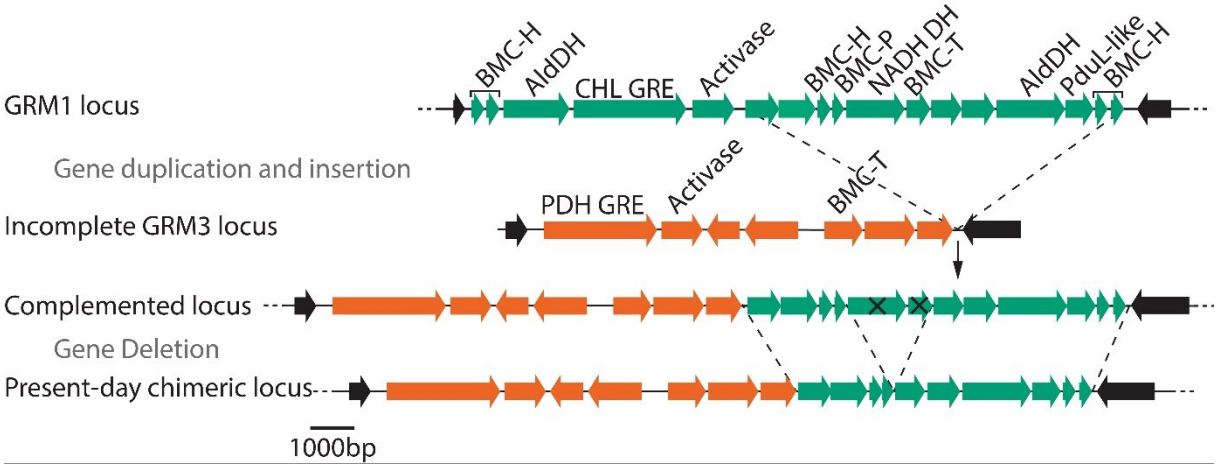

BMC-H
GRM1
GUF
GRM3
GRM3, GRM4, PDUGRM
GUF
CcmK4
CcmK2
CsoS1A
GRM1
GUF
GRM2
GRM1
GRM5
GRM3, GRM4, PDUGRM
GRM1
GRM5
GRM5
GRM1
GRM2, PDUGRM
GRM1
GRM1
GRM1
GRM3
0.2
GRM1
(incl. PduU)
GRM5
BMC-T
GRM1
(Streptococcus)
GRM1*
GRM1*
EutL
PDU-GRM
PduT
GRM3
GRM3
GUF
GUF
GRM1*
GRM5
GRM3
GRM3
EtuB
GRM3
GRM3
PduB
PDU-GRM
GRM4
GRM1
GRM3
0.5
CsoS1D
CcmP
BMC-P
GRM5
GRM3
GRM1*
GUF
PduN
PDU-GRM
GRM4
CcmL
EutN
GUF
GUF
CsoS4A
GRM1
GRM2
0.2

Figure S6.


GRM1 locus
BMC-H
AldDH
CHL GRE
Activase
BMC-H
BMC-P
NADH DH
BMC-T
AldDH
PduL-like
BMC-H
Gene duplication and insertion
Incomplete GRM3 locus
PDH GRE
Activase
BMC-T
Complemented locus
Gene Deletion
Present-day chimeric locus
1000bp