Supplementary Results

The genomic and phenotypic characteristics of previously uncharacterized phages targeting obligate anaerobes

For the four crAss-like phages we isolated, *B. cellulosilyticus* phages CPB1000 and CPB1023 form clear plaque, while *B. thetaiotaomicron* phage CPB1112 and *P. distasonis* phage CPB1101 form tiny and turbid plaques (Figure S7A-D). However, they exhibit effective inhibition in liquid (Figure S7A-D). Transmission Electron Microscope (TEM) of crAss-like phage revealed a Podovirus-like morphology, with a capsid diameter of 60±2.2 nm and inconspicuous short tails (Figure S7A). These four phages have genome sizes ranging from 97,478 to 103,249 bp, with GC contents of 33-37 mol % (Table S1). They encode over 100 coding sequences (CDSs) carrying multiple tRNAs, with CPB1101 notably carrying 26 tRNAs. The genomes of these four phages are divided into two roughly equal parts with opposite transcriptional orientations, which are consistent with ΦcrAss001[1]. Additionally, we found that these four phages all carry AMGs, such as phosphoribosyl transferase and adenylyl transferase (Table S1). We further compared these four crAss-like phages to 31 previously reported isolated crAss-like phages (Figure S7E). The results showed that our four phages were clustered into three groups, each containing previously reported crAss-like phages. We noted that *B. cellulosilyticus* phages were distributed into three distinct clusters, each comprising phages that infect different bacterial species, indicating that the genomic similarity of crAss-like phages may not cluster based on the host species they infect.

For *E. lenta* phage, six isolates exhibited clear plaques without halo (Figure S8A). However, due to the low cell density in the bacterial culture, we were unable to quantify their inhibition ability in liquid (Figure S8A). TEM revealed a Podovirus-like morphology, with a polyhedral capsid of 42 ±1.9 nm in diameter and short tails of 14±2.5 nm (Figure S8A). Of the six phages isolated, only four were successfully sequenced using various methods, including high-throughput shotgun sequencing and long-read sequencing complemented by PCR amplification of gap contigs for Sanger sequencing. However, their genomic completeness was still low (33-70%). we observed 3 isolates carrying a pentapeptide repeat protein, commonly found in cyanobacteria and poorly understood (Figure S8B-D). We performed an ANI analysis between our four phages and previously reported ones and revealed high diversity among *E. lenta* phages (Figure S8E). Two reported isolates (MH626557 and MW679025) displayed partial similarity to ours (ANI 83.13-87.86%). Therefore, we performed pairwise comparisons between our phages and the two isolates. We found that the gene cluster arrangement of CPB0531 and MW679025 (completeness of 71.39%) showed a higher resemblance (Figure S8F). These observations suggest the unique genomic features among such phages, which may contribute to the incomplete sequencing results.

The 7 *R. gnavus* phages formed clear, visible plaques, while they exhibited a limited inhibitory effect on bacterial growth in liquid culture (Figure S9A). TEM revealed virions with a typical Siphovirus-like morphology, characterized by icosahedral heads (54±1.5 nm) and long flexible non-contractile tails (124±6.8 nm) possessing baseplates (Figure S9A). The genome sizes of the 7 isolates range from 34,871 to 50,654 bp and their GC contents range from 42 mol % to 47 mol %. These phages harbor 55-79 CDSs, with an annotation rate of 49-69%. Notably, these phages carry lysogenic cycle-related genes, which potentially cause a limited inhibitory effect within the liquid

culture (Figure S9B).

Four *P. vulgatus* phage isolates formed distinct plaques in agar overlays. CPB1107 and CPB1111 formed clear plaques, with CPB1107 showing a halo (Figure S10A). CPB1128 and CPB1114 formed turbid plaques initially (Figure S10A-B) but failed to form a single colony with dilution. In liquid culture, the phage exhibited a mild inhibitory effect on the host (Figure S10B). TEM revealed a Siphovirus-like morphology with a capsid diameter of 63 ± 2.9 nm and a tail length of 143 ± 2.7 nm (Figure S10B). The 4 phages have notably different genomic profiles. The genome lengths of CPB1128 and CPB1107 range from 43,568 to 43,666 bp, with a GC content of 40 mol %. They encode 54-56 CDSs, with 37% annotated. Conversely, CPB1111 and CPB1114 have larger genomes, ranging from 80,260 to 96,750 bp, with a GC content of 38 mol %. They encode 118-143 CDSs, with only 24-26% annotated (Figure S10C). Notably, we observed that CPB1107 and CPB1128 featured all CDSs oriented in the same direction, while the genomic organization of CPB1111 and CPB1114 was divided into 4 distinct regions (labeled as modules I to IV) (Figure S10C). This diverse orientation of CDSs may have biological implications, such as providing spatial utilization of genetic information and coordinating gene regulation for proper expression levels, timing, and locations.

D. longicatena phages, which have not been previously reported, form tiny, slightly turbid plaques in agar overlays and exhibit significant inhibitory effects on their bacterial hosts in liquid (Figure S11A). TEM revealed that the *D. longicatena* phages exhibit a Podovirus-like morphology, characterized by a capsid diameter of 42 ± 2.2 nm and inconspicuous short tails (Figure S11A). The genome sizes of *D. longicatena* phages range from 15,701 to 22,234 bp, with a GC content of ~ 33 mol %. We identified 17-26 CDSs with a relatively high annotation rate of 57-64% (S11B).

Supplementary Figures

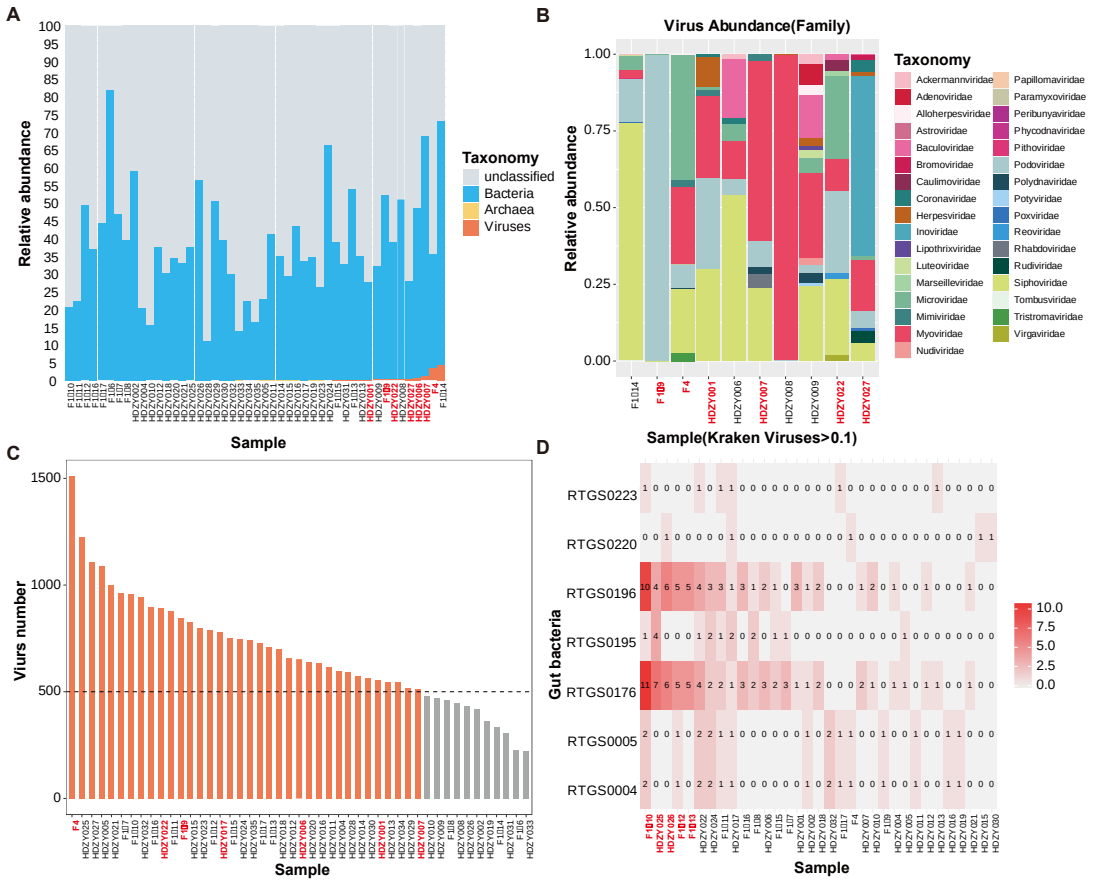


Figure S1 The metagenomic analysis of fecal samples for bacteriophage screening.

(A) The relative abundance of microbes in feces. Samples in the red font on the x-axis are the samples being chosen for the metagenomics analysis-based phage screening.

(B) The relative abundance of viral family in the fecal samples which viral relative abundance is more than 0.1% by Kraken. Samples in the red font on the x-axis are the samples being chosen for the metagenomics analysis-based phage screening.

(C) The numbers of predicted non-redundant viral sequences in feces. Samples in the red font on the x-axis are the samples being chosen for the metagenomics analysis-based phage screening.

(D) The host assignment of predicted viral sequences in feces. The y-axis is the internal ID of human gut bacteria used in this study. The red color and numbers mean the count of contigs within the fecal samples matched with bacteria CRISPR spacers. The darker the red, the greater the number of contigs matched. Only the fecal samples with at least one contig matched with bacteria CRISPR spacers are displayed. Samples in the red font on the x-axis are the samples being chosen for the metagenomics analysis-based phage screening.

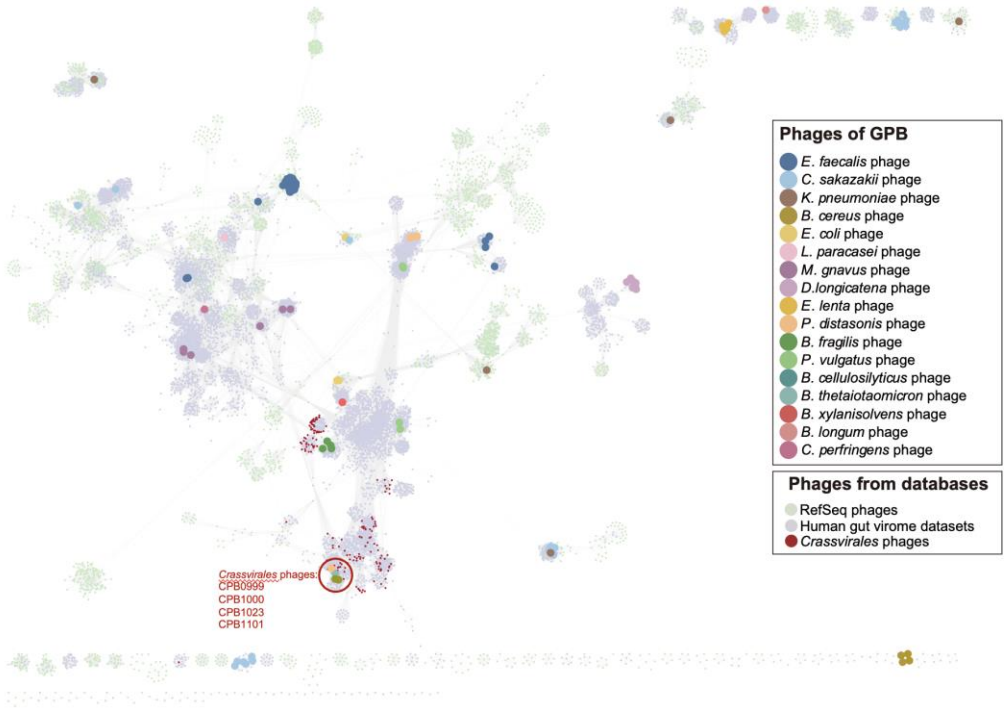
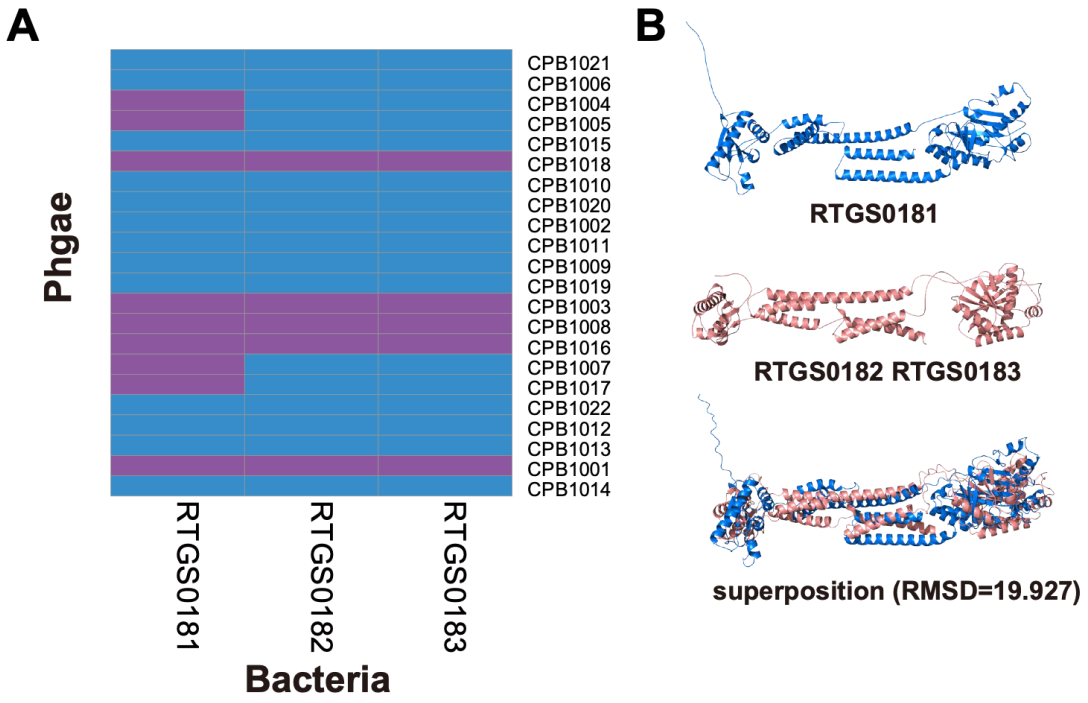


Figure S2 The gene-sharing network of GPB phages, RefSeq phages, and human gut virome datasets.

The gene-sharing network was constructed using GPB phages, RefSeq phages, the crAss-like phages dataset, and a subset from four virome databases (GPD[2], GVD[3], CHVD[4] and MGV[5]). The network includes a total of 18,112 phages, comprising 102 GPB phages, 4534 phages in the RefSeq database, and 13,476 phage contigs from human gut databases and the crAss-like phages dataset which can align with GPB phages. The resulting network comprised 17,219 nodes linked by 2,801,293 edges, with each node representing a phage. The GPB phages are labeled as big dots with different colors based on their host species. RefSeq phages are labeled as small dots with a light green color. Phages from human gut virile databases are labeled as small dots with a light purple color. crAss-like phages are labeled as small dots with red color. Four CPB phage isolates clustered with crAss-like phages are highlighted with their internal IDs.



98

99 Figure S3 The interactions between GPB phages targeting facultative anaerobes and the host
100 bacteria under both oxic and anoxic conditions.

101 (A) The host range of *C. sakazakii* phages under oxic and anoxic conditions. The purple blocks
102 mean the phage could form plaques under both oxic and anoxic conditions. The blue blocks mean
103 infection under oxic condition.

104 (B) The 3D structures of Peptidoglycan-associated lipoprotein of *C. sakazakii* RTGS0181,
105 RTGS0182, RTGS0183, and their superposition.

106

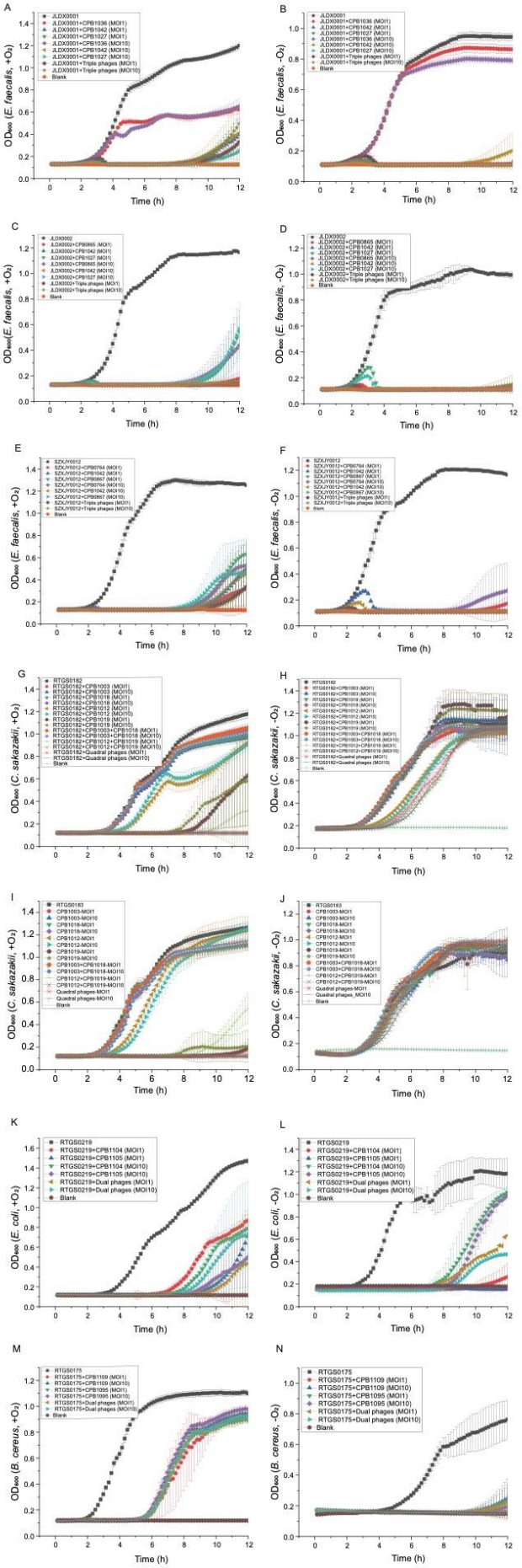

A
B
C
D
E
F
G
H
I
J
K
L
M
N
Time (h)

Figure S4 Killing curves of GPB phages targeting facultative anoxic bacteria.

(A-B) The killing curves of CPB1036, CPB1042, CPB1027 and their mix targeting *E. faecalis* JLDX0001 under oxic (A) and anoxic (B) conditions.

(C-D) The killing curves of CPB0865, CPB1042, CPB1027 and their mix targeting *E. faecalis* JLDX0002 under oxic (C) and anoxic (D) conditions.

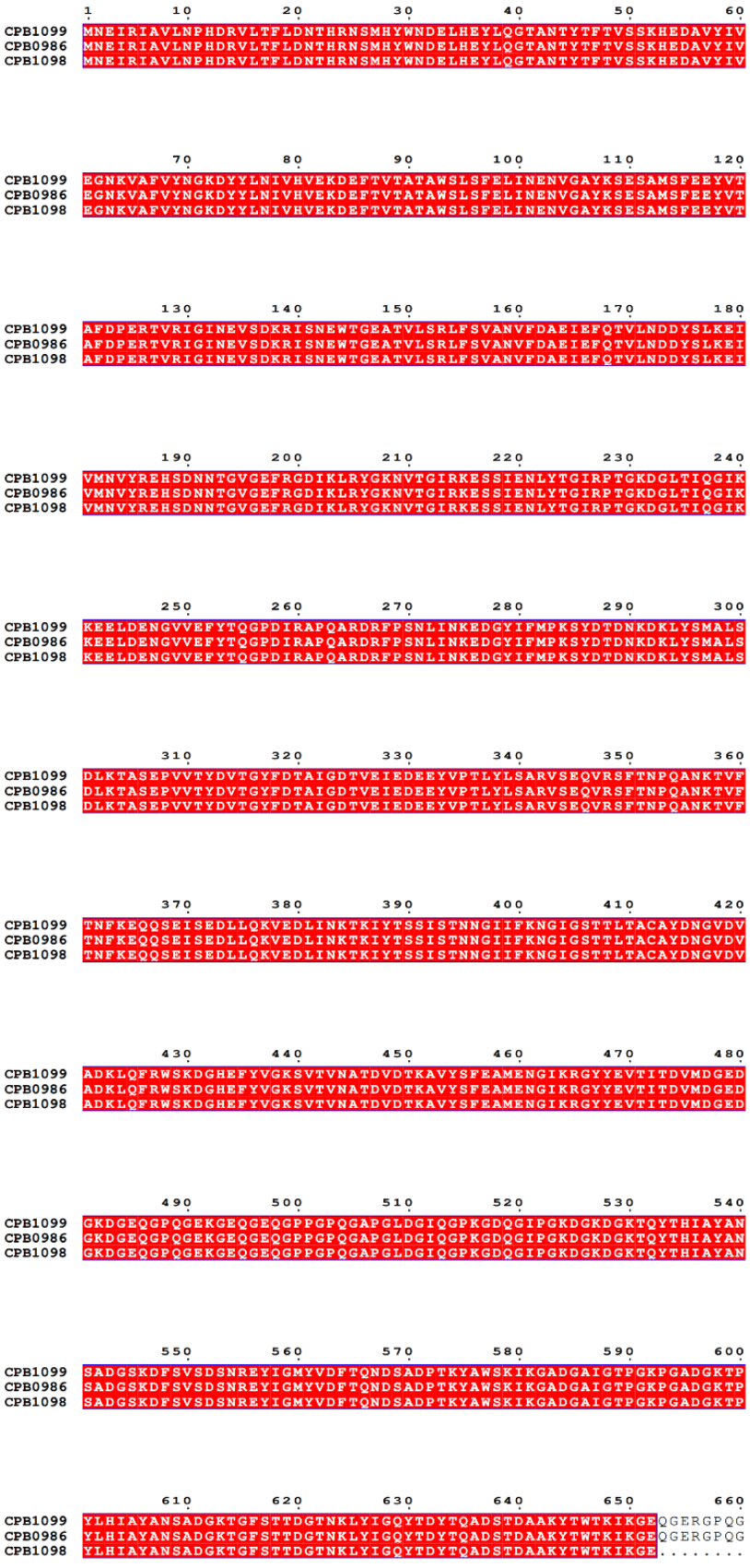
(E-F) The killing curves of CPB0764, CPB1042, CPB0867 and their mix targeting *E. faecalis* SZXJY0012 under oxic (E) and anoxic (F) conditions.

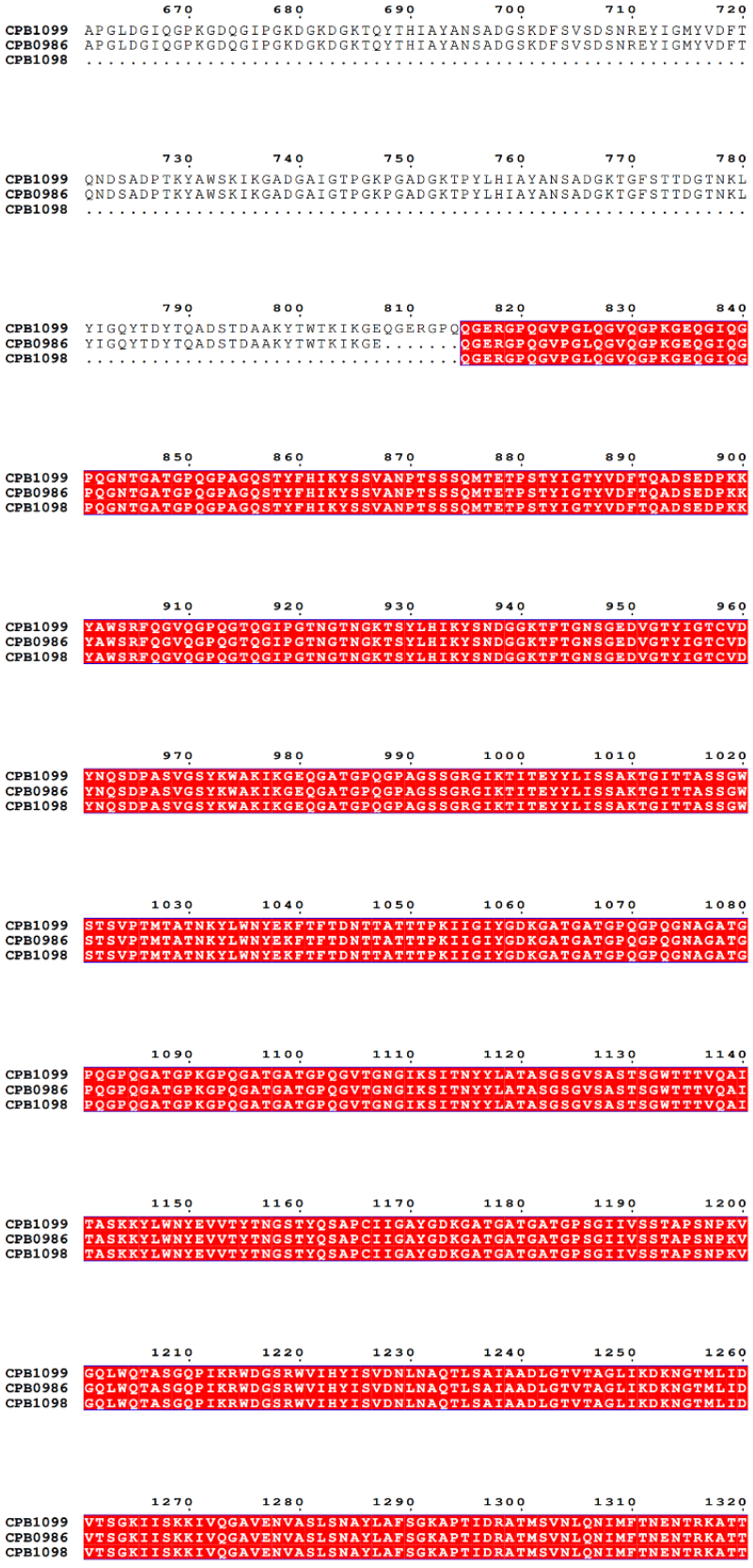
(G-H) The killing curves of CPB1003, CPB1018, CPB1012, CPB1019 and their mix targeting *C. sakazakii* RTGS0182 under oxic (G) and anoxic (H) conditions. CPB1003 and CPB1018 could form plaques on RTGS0182 under both oxic and anoxic conditions. CPB1012 and CPB1019 could only form plaques under oxic condition.

(I-J) The killing curves of CPB1003, CPB1018, CPB1012, CPB1019 and their mix targeting *C. sakazakii* RTGS0183 under oxic (I) and anoxic (J) conditions. CPB1003 and CPB1018 could form plaques on RTGS0183 under both oxic and anoxic conditions. CPB1012 and CPB1019 could only form plaques under oxic condition.

(K-L) The killing curves of CPB1104, CPB1105, and their mix targeting *E. coli* RTGS0219 under oxic (M) and anoxic (N) conditions.

(M-N) The killing curves of CPB1109, CPB1095, and their mix targeting *B. cereus* RTGS0175 under oxic (K) and anoxic (L) conditions.


1 10 20 30 40 50 60
CPB1099 MNEIRIAVLNPHDRVLTFLDNTHRNSMHYWNDELHEYLQGTANTYTFTVSSKHEDAVYIV
CPB0986 MNEIRIAVLNPHDRVLTFLDNTHRNSMHYWNDELHEYLQGTANTYTFTVSSKHEDAVYIV
CPB1098 MNEIRIAVLNPHDRVLTFLDNTHRNSMHYWNDELHEYLQGTANTYTFTVSSKHEDAVYIV
70 80 90 100 110 120
CPB1099 EGNKVAFVYNGKDYYLNIVHVEKDEFTVTATAWSLSFELINENVGAYKSESAMSFEEYVT
CPB0986 EGNKVAFVYNGKDYYLNIVHVEKDEFTVTATAWSLSFELINENVGAYKSESAMSFEEYVT
CPB1098 EGNKVAFVYNGKDYYLNIVHVEKDEFTVTATAWSLSFELINENVGAYKSESAMSFEEYVT
130 140 150 160 170 180
CPB1099 AFDPERTVRIGINEVSDKRISNEWTGEATVLSRLFSVANVFDAEIEFQTVLNDDYSLKEI
CPB0986 AFDPERTVRIGINEVSDKRISNEWTGEATVLSRLFSVANVFDAEIEFQTVLNDDYSLKEI
CPB1098 AFDPERTVRIGINEVSDKRISNEWTGEATVLSRLFSVANVFDAEIEFQTVLNDDYSLKEI
190 200 210 220 230 240
CPB1099 VMNVYREHSDNNTGVGEFRGDIKLRYGKNVTGIRKESSIENLYTGIRPTGKDGLTIQGIK
CPB0986 VMNVYREHSDNNTGVGEFRGDIKLRYGKNVTGIRKESSIENLYTGIRPTGKDGLTIQGIK
CPB1098 VMNVYREHSDNNTGVGEFRGDIKLRYGKNVTGIRKESSIENLYTGIRPTGKDGLTIQGIK
250 260 270 280 290 300
CPB1099 KEELDENGVVEFYTQGPDIRAPQARDRFPSNLINKEDGYIFMPKSYDTDNKDKLYSMALS
CPB0986 KEELDENGVVEFYTQGPDIRAPQARDRFPSNLINKEDGYIFMPKSYDTDNKDKLYSMALS
CPB1098 KEELDENGVVEFYTQGPDIRAPQARDRFPSNLINKEDGYIFMPKSYDTDNKDKLYSMALS
310 320 330 340 350 360
CPB1099 DLKTASEPVVTYDVTGYFDTAIGDTVEIEDEEYVPTLYLSARVSEQVRSFTNPQANKTVF
CPB0986 DLKTASEPVVTYDVTGYFDTAIGDTVEIEDEEYVPTLYLSARVSEQVRSFTNPQANKTVF
CPB1098 DLKTASEPVVTYDVTGYFDTAIGDTVEIEDEEYVPTLYLSARVSEQVRSFTNPQANKTVF
370 380 390 400 410 420
CPB1099 TNFKEQQSEISEDLLQKVEDLINKTKIYTSSISTNNGIIFKNGIGSTTLTACAYDNGVDV
CPB0986 TNFKEQQSEISEDLLQKVEDLINKTKIYTSSISTNNGIIFKNGIGSTTLTACAYDNGVDV
CPB1098 TNFKEQQSEISEDLLQKVEDLINKTKIYTSSISTNNGIIFKNGIGSTTLTACAYDNGVDV
430 440 450 460 470 480
CPB1099 ADKLQFRWSKDGHEFYVGKSVTVNATDVDTKAVYSFEAMENGIKRGYYEVTITDVMDGED
CPB0986 ADKLQFRWSKDGHEFYVGKSVTVNATDVDTKAVYSFEAMENGIKRGYYEVTITDVMDGED
CPB1098 ADKLQFRWSKDGHEFYVGKSVTVNATDVDTKAVYSFEAMENGIKRGYYEVTITDVMDGED
490 500 510 520 530 540
CPB1099 GKDGEQGPQGEKGEQGEQGPPGPQGAPGLDGIQGPKGDQGIPGKDGKDGKTQYTHIAYAN
CPB0986 GKDGEQGPQGEKGEQGEQGPPGPQGAPGLDGIQGPKGDQGIPGKDGKDGKTQYTHIAYAN
CPB1098 GKDGEQGPQGEKGEQGEQGPPGPQGAPGLDGIQGPKGDQGIPGKDGKDGKTQYTHIAYAN
550 560 570 580 590 600
CPB1099 SADGSKDFSVSDSNREYIGMYVDFTQNDSADPTKYAWSKIKGADGAIGTPGKPGADGKTP
CPB0986 SADGSKDFSVSDSNREYIGMYVDFTQNDSADPTKYAWSKIKGADGAIGTPGKPGADGKTP
CPB1098 SADGSKDFSVSDSNREYIGMYVDFTQNDSADPTKYAWSKIKGADGAIGTPGKPGADGKTP
610 620 630 640 650 660
CPB1099 YLHIAYANSADGKTGFSTTDGTNKLYIGQYTDYTQADSTDAAKYTWTKIKGEQGERGPQG
CPB0986 YLHIAYANSADGKTGFSTTDGTNKLYIGQYTDYTQADSTDAAKYTWTKIKGEQGERGPQG
CPB1098 YLHIAYANSADGKTGFSTTDGTNKLYIGQYTDYTQADSTDAAKYTWTKIKGE........


670 680 690 700 710 720
CPB1099 APGLDGIQGPKGDQGIPGKDGKDGKTQYTHIAYANSADGSKDFSVSDSNREYIGMYVDFT
CPB0986 APGLDGIQGPKGDQGIPGKDGKDGKTQYTHIAYANSADGSKDFSVSDSNREYIGMYVDFT
CPB1098 ....
730 740 750 760 770 780
CPB1099 QNDSADPTKYAWSKIKGADGAIGTPGKPGADGKTPYLHIAYANSADGKTGFSTTDGTNKL
CPB0986 QNDSADPTKYAWSKIKGADGAIGTPGKPGADGKTPYLHIAYANSADGKTGFSTTDGTNKL
CPB1098 ....
790 800 810 820 830 840
CPB1099 YIGQYTDYTQADSTDAAKYTWTKIKGEQGERGPQQGERGPQGVPGLQGVQGPKGEQGIQG
CPB0986 YIGQYTDYTQADSTDAAKYTWTKIKGE.....QGERGPQGVPGLQGVQGPKGEQGIQG
CPB1098 ..........QGERGPQGVPGLQGVQGPKGEQGIQG
850 860 870 880 890 900
CPB1099 PQGNTGATGPQGPAGQSTYFHIKYSSVANPTSSSQMTETPSTYIGTYVDFTQADSEDPKK
CPB0986 PQGNTGATGPQGPAGQSTYFHIKYSSVANPTSSSQMTETPSTYIGTYVDFTQADSEDPKK
CPB1098 PQGNTGATGPQGPAGQSTYFHIKYSSVANPTSSSQMTETPSTYIGTYVDFTQADSEDPKK
910 920 930 940 950 960
CPB1099 YAWSRFQGVQGPQGTQGIPGTNGTNGKTSYLHIKYSNDGGKTFTGNSGEDVGTYIGTCVD
CPB0986 YAWSRFQGVQGPQGTQGIPGTNGTNGKTSYLHIKYSNDGGKTFTGNSGEDVGTYIGTCVD
CPB1098 YAWSRFQGVQGPQGTQGIPGTNGTNGKTSYLHIKYSNDGGKTFTGNSGEDVGTYIGTCVD
970 980 990 1000 1010 1020
CPB1099 YNQSDPASVGSYKWAKIKGEQGATGPQGPAGSSGRGIKTITEYYLISSAKTGITTASSGW
CPB0986 YNQSDPASVGSYKWAKIKGEQGATGPQGPAGSSGRGIKTITEYYLISSAKTGITTASSGW
CPB1098 YNQSDPASVGSYKWAKIKGEQGATGPQGPAGSSGRGIKTITEYYLISSAKTGITTASSGW
1030 1040 1050 1060 1070 1080
CPB1099 STSVPTMTATNKYLWNYEKFTFTDNTTATTTPKIIGIYGDKGATGATGPQGPQGNAGATG
CPB0986 STSVPTMTATNKYLWNYEKFTFTDNTTATTTPKIIGIYGDKGATGATGPQGPQGNAGATG
CPB1098 STSVPTMTATNKYLWNYEKFTFTDNTTATTTPKIIGIYGDKGATGATGPQGPQGNAGATG
1090 1100 1110 1120 1130 1140
CPB1099 PQGPQGATGPKGPQGATGATGPQGVTGNGIKSITNYYLATASGSGVSASTSGWTTTVQAI
CPB0986 PQGPQGATGPKGPQGATGATGPQGVTGNGIKSITNYYLATASGSGVSASTSGWTTTVQAI
CPB1098 PQGPQGATGPKGPQGATGATGPQGVTGNGIKSITNYYLATASGSGVSASTSGWTTTVQAI
1150 1160 1170 1180 1190 1200
CPB1099 TASKKYLWNYEVVTYTNGSTYQSAPCIIGAYGDKGATGATGATGPSGIIVSSTAPSNPKV
CPB0986 TASKKYLWNYEVVTYTNGSTYQSAPCIIGAYGDKGATGATGATGPSGIIVSSTAPSNPKV
CPB1098 TASKKYLWNYEVVTYTNGSTYQSAPCIIGAYGDKGATGATGATGPSGIIVSSTAPSNPKV
1210 1220 1230 1240 1250 1260
CPB1099 GQLWQTASGQPIKRWDGSRWVIHYISVDNLNAQTLSAIAADLGTVTAGLIKDKNGTMLID
CPB0986 GQLWQTASGQPIKRWDGSRWVIHYISVDNLNAQTLSAIAADLGTVTAGLIKDKNGTMLID
CPB1098 GQLWQTASGQPIKRWDGSRWVIHYISVDNLNAQTLSAIAADLGTVTAGLIKDKNGTMLID
1270 1280 1290 1300 1310 1320
CPB1099 VTSGKIISKKIVQGAVENVASLSNAYLAFSGKAPTIDRATMSVNLQNIMFTNENTRKATT
CPB0986 VTSGKIISKKIVQGAVENVASLSNAYLAFSGKAPTIDRATMSVNLQNIMFTNENTRKATT
CPB1098 VTSGKIISKKIVQGAVENVASLSNAYLAFSGKAPTIDRATMSVNLQNIMFTNENTRKATT

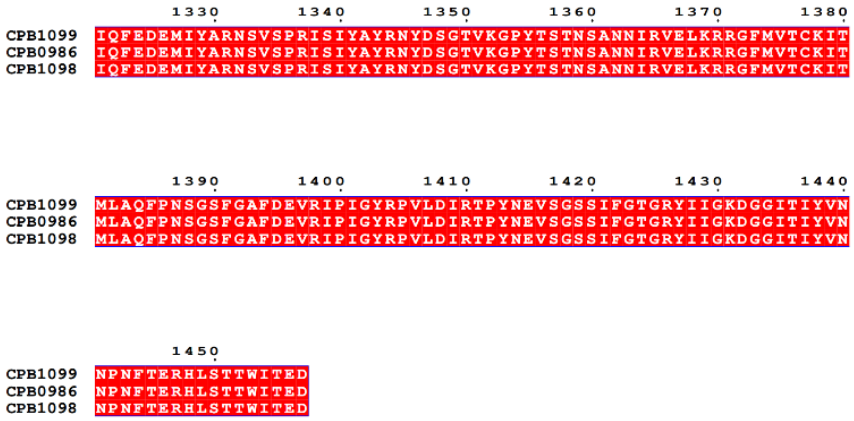
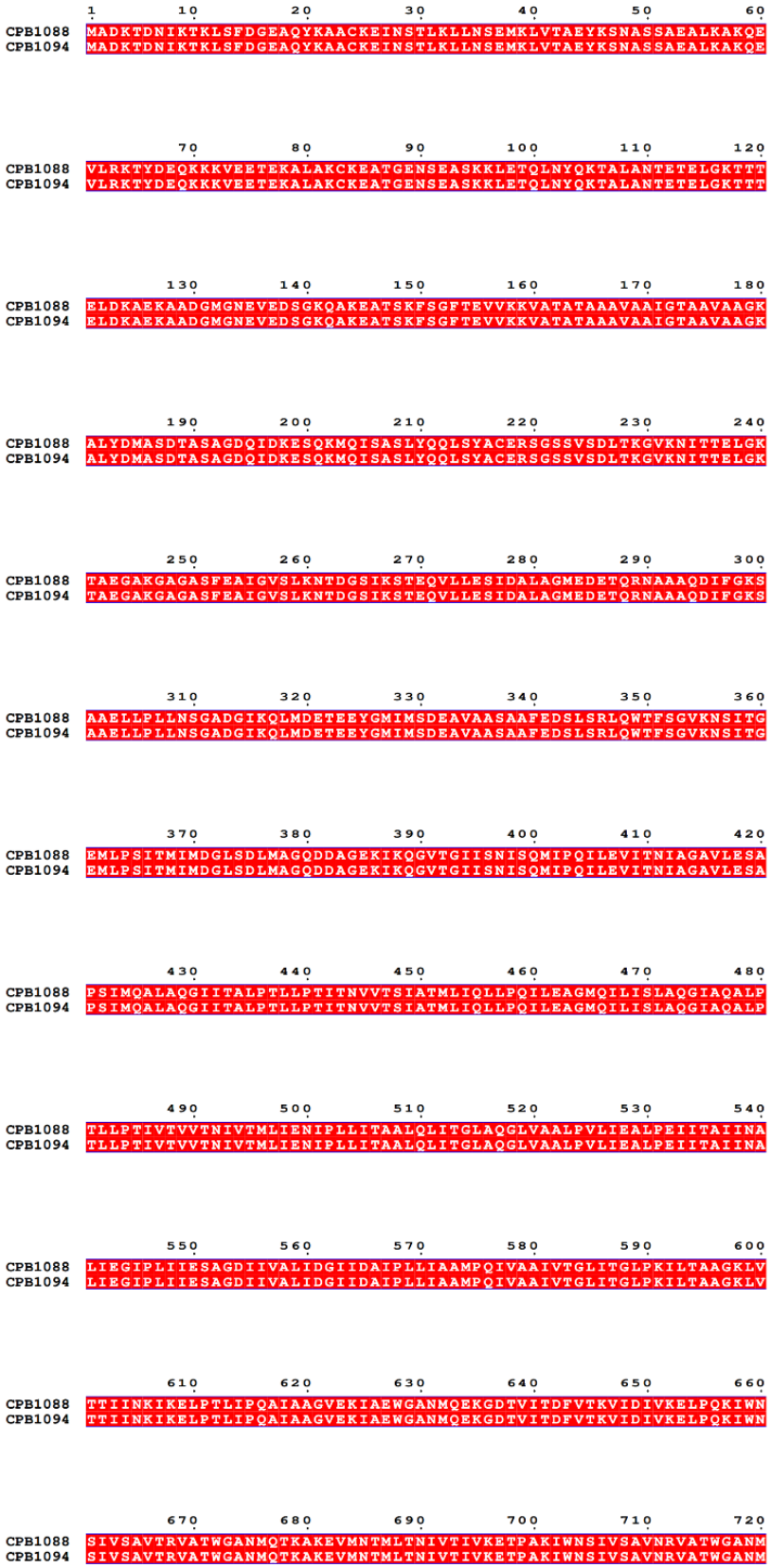


Figure S5 The multiple sequence alignment results of phage tail protein in *M. gnavus* phage Group2.

The sequences highlighted in red represent identical sequences across the inter-group. Sequences without any highlighting signify differences across the inter-group.


1 10 20 30 40 50 60
CPB1088 MADKTDNIKTKLSFDGEAQYKAACKEINSTLKLLNSEMKLVTAEYKSNASSAEALKAKQE
CPB1094 MADKTDNIKTKLSFDGEAQYKAACKEINSTLKLLNSEMKLVTAEYKSNASSAEALKAKQE
70 80 90 100 110 120
CPB1088 VLRKTYDEQKKKVEETEKALAKCKEATGENSEASKKLETQLNYQKTALANTETELGKTTT
CPB1094 VLRKTYDEQKKKVEETEKALAKCKEATGENSEASKKLETQLNYQKTALANTETELGKTTT
130 140 150 160 170 180
CPB1088 ELDKAEKAADGMGNEVEDSGKQAKEATSKFSGFTEVVKKVATATAAAVAAIGTAAVAAGK
CPB1094 ELDKAEKAADGMGNEVEDSGKQAKEATSKFSGFTEVVKKVATATAAAVAAIGTAAVAAGK
190 200 210 220 230 240
CPB1088 ALYDMASDTASAGDQIDKESQKMQISASLYQQLSYACERSGSSVSDLTKGVKNITTELGK
CPB1094 ALYDMASDTASAGDQIDKESQKMQISASLYQQLSYACERSGSSVSDLTKGVKNITTELGK
250 260 270 280 290 300
CPB1088 TAEGAKGAGASFEAIGVSLKNTDGSIKSTEQVLLESIDALAGMEDETQRNAAAQDIFGKS
CPB1094 TAEGAKGAGASFEAIGVSLKNTDGSIKSTEQVLLESIDALAGMEDETQRNAAAQDIFGKS
310 320 330 340 350 360
CPB1088 AAELLPLLNSGADGIKQLMDETEEYGMIMSDEAVAASAAFEDSLSRLQWTFSGVKNSITG
CPB1094 AAELLPLLNSGADGIKQLMDETEEYGMIMSDEAVAASAAFEDSLSRLQWTFSGVKNSITG
370 380 390 400 410 420
CPB1088 EMLPSITMIMDGLSDLMAGQDDAGEKIKQGVTGIISNISQMIPQILEVITNIAGAVLESA
CPB1094 EMLPSITMIMDGLSDLMAGQDDAGEKIKQGVTGIISNISQMIPQILEVITNIAGAVLESA
430 440 450 460 470 480
CPB1088 PSIMQALAQGIITALPTLLPTITNVVTSIATMLIQLLPQILEAGMQILISLAQGIAQALP
CPB1094 PSIMQALAQGIITALPTLLPTITNVVTSIATMLIQLLPQILEAGMQILISLAQGIAQALP
490 500 510 520 530 540
CPB1088 TLLPTIVTVVTNIVTMLIENIPLLITAALQLITGLAQGLVAALPVLIEALPEIITAIINA
CPB1094 TLLPTIVTVVTNIVTMLIENIPLLITAALQLITGLAQGLVAALPVLIEALPEIITAIINA
550 560 570 580 590 600
CPB1088 LIEGIPLIIESAGDIIVALIDGIIDAIPLLIAAMPQIVAAIVTGLITGLPKILTAAGKLV
CPB1094 LIEGIPLIIESAGDIIVALIDGIIDAIPLLIAAMPQIVAAIVTGLITGLPKILTAAGKLV
610 620 630 640 650 660
CPB1088 TTIINKIKELPTLIPQAIAAGVEKIAEWGANMQEKGDTVITDFVTKVIDIVKELPQKIWN
CPB1094 TTIINKIKELPTLIPQAIAAGVEKIAEWGANMQEKGDTVITDFVTKVIDIVKELPQKIWN
670 680 690 700 710 720
CPB1088 SIVSAVTRVATWGANMQTKAKEVMNTMLTNIVTIVKETPAKIWNSIVSAVNRVATWGANM
CPB1094 SIVSAVTRVATWGANMQTKAKEVMNTMLTNIVTIVKETPAKIWNSIVSAVNRVATWGANM

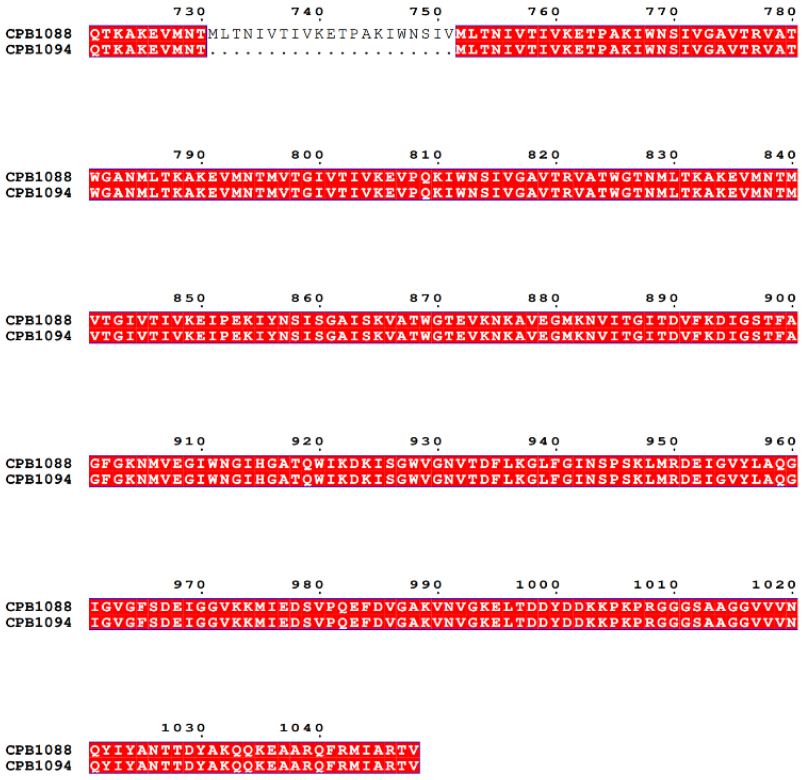
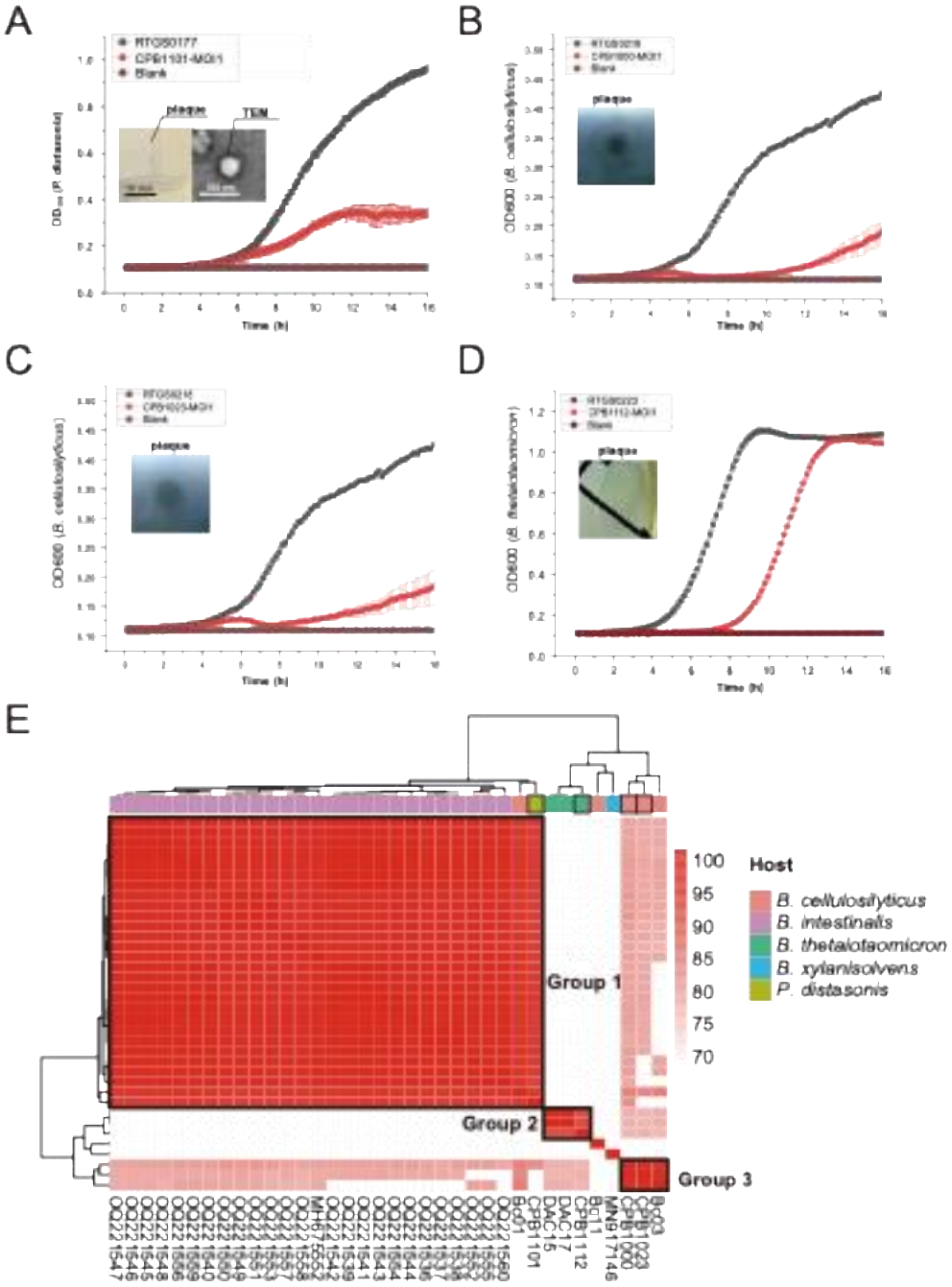


Figure S6 The multiple sequence alignment results of phage tail tape measure protein in *M. gnavus* phage Group 4.

The sequences highlighted in red represent identical sequences across the inter-group. Sequences without any highlighting signify differences across the inter-group.



146

147 Figure S7 Biological and genomic characteristics of GPB crAss-like phages.

148 (A-D) The killing curves and plaques of four GPB crAss-like phages – CPB1101 (A), CPB1000 (B),
149 CPB1023 (C) and CPB1112 (D). MOI = 1. Data are represented as mean ± SD (n=3). (A) The inset
150 left image displays the plaque of phage CPB1101 with a scale bar of 10 mm, while the inset right
151 image shows the TEM of phage CPB1101 with a scale bar of 100 nm.

152 (E) The genomic similarity of crAss-like phage isolates. The color-coded bar at the top indicates the
153 host species of phages, GPB crAss-like phages are highlighted with black boxes. The heatmap is
154 clustered based on the genomic similarity, with the intensity of the red color indicating the genomic

similarity between each two phages. Clusters containing GPB crAss-like phages are highlighted with black boxes and labeled as Group 1, Group 2, and Group 3.

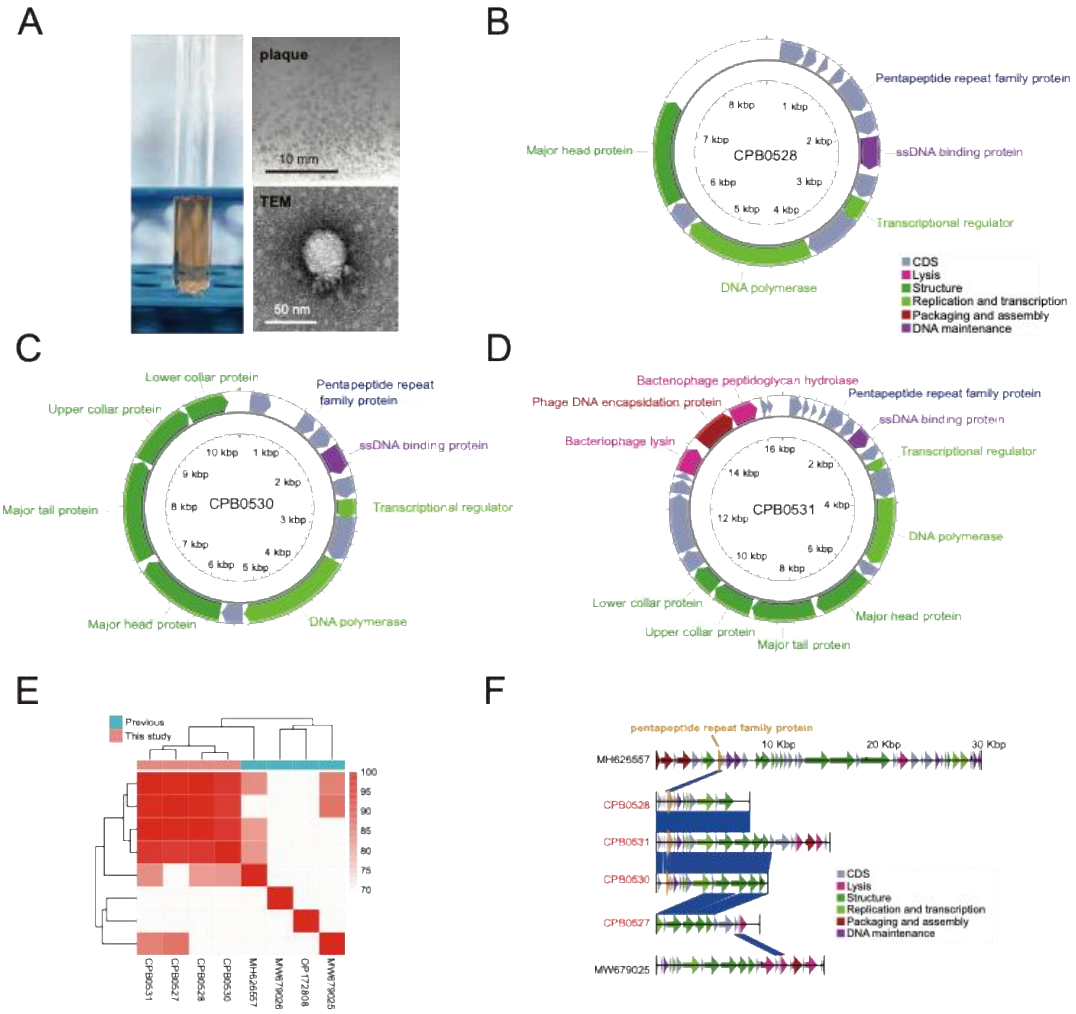


Figure S8 Biological and genomic characteristics of *Eggerthella* phages.

(A) Biological characteristics of *E. lenta* RTGS0226 and phage CPB0527. The left image shows *E. lenta* RTGS0226 in a liquid culture medium. The top right image displays the plaque of phage CPB0527 with a scale bar of 10 mm, while the bottom right image shows the TEM of phage CPB0527 with a scale bar of 50 nm.

(B-D) The genomic map of three *E. lenta* phages (B) CPB0528, (C) CPB0530 and (D) CPB0531.

(E) Genomic similarity of isolated *E. lenta* phage isolates. Phages isolated in this study are indicated by pink at the top, while previously isolated phages are indicated by blue. The heatmap is clustered based on the genomic similarity, with the intensity of the red color indicating the genomic similarity between each two phages.

(F) The multiple sequence alignment results of *E. lenta* phage isolates. Phages in the red text represent GPB phages. The arrows indicate the direction of the CDSs, with different colors representing various functional categories. Blue regions represent sequence similarity greater than 70%. Pentapeptide repeat family proteins are highlighted with yellow-bordered arrows.

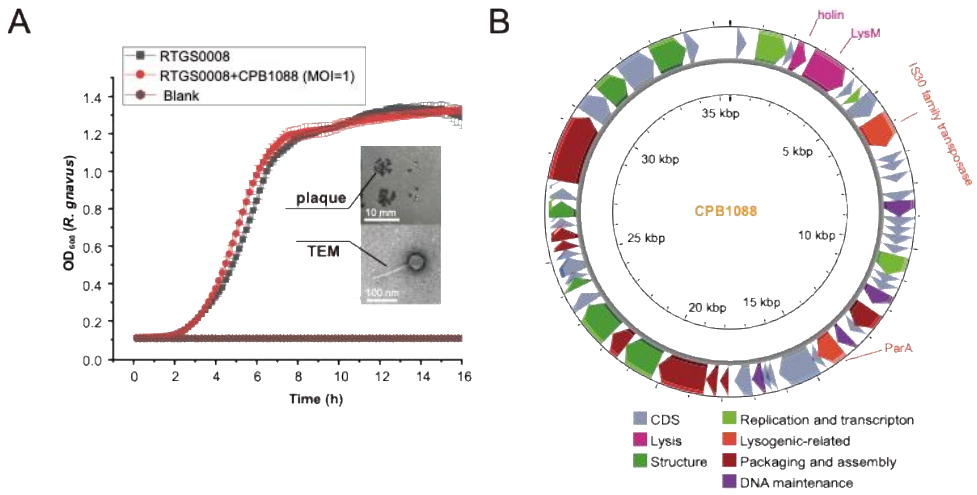
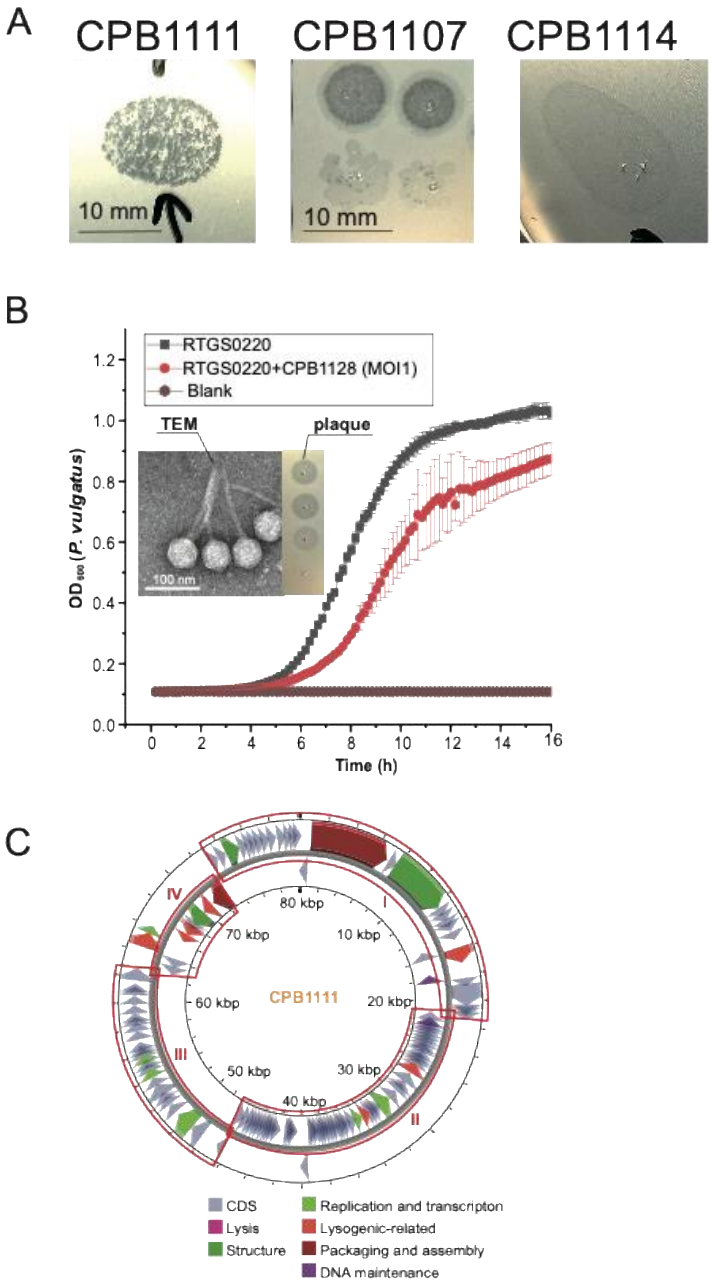


Figure S9 Biological and genomic characteristics of *Mediterraneibacter* phages.

(A) The killing curves of CP1088 targeting *M. gnavus* RTGS0008. MOI= 1. Data are represented as mean ± SD (n=3). The inset top image displays the plaque of phage CPB1088 with a scale bar of 10 mm, while the inset bottom image shows the TEM of phage CPB1088 with a scale bar of 100 nm.

(B) Genome map of *M. gnavus* phage CPB1088, with annotated genes and functional categories. Lysis proteins and lysogenic-related proteins are highlighted.



183

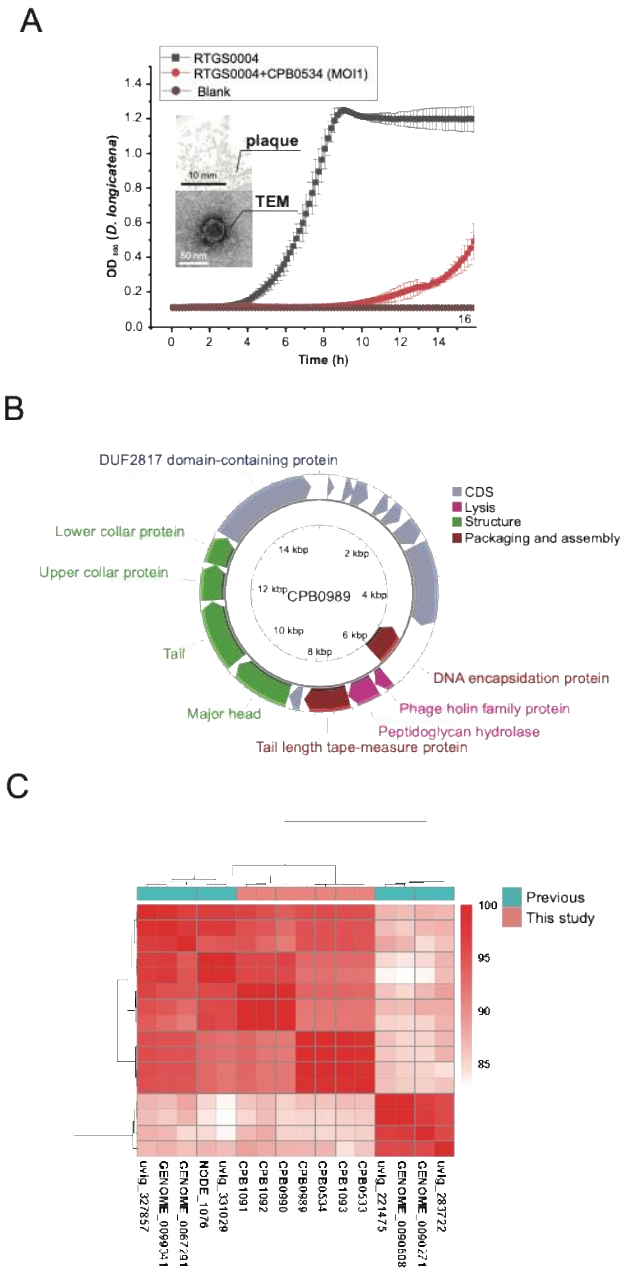
184 Figure S10 Biological and genomic characteristics of *Phocaeicola* phages.

185 (A) Plaques of *P. vulgatus* phages CPB1111, CPB1107 and CPB1114.

186 (B) The killing curves of CP1128 targeting *P. vulgatus* RTGS0220. MOI = 1. Data are represented
187 as mean ± SD (n=3). The inset left image shows the TEM of phage CPB1128 with a scale bar of
188 100 nm, while the inset right image displays the plaque of phage CPB1128 as the dilution gradient.

189 (C) Genome map of *P. vulgatus* phage CP1128, with annotated genes and functional categories. The
190 genome is divided into four parts based on different transcriptional orientations (I, II, III, IV).

191



192

193 Figure S11 Genomic map and ANI heatmap of *Dorea* phages.

194 (A) The killing curves of CP0534 targeting *D. longicatena* RTGS0004. MOI = 1. Data are
195 represented as mean ± SD (n=3). The inset top image displays the plaque of phage CPB0534 with
196 a scale bar of 10 mm, while the inset bottom image shows the TEM of phage CPB0534 with a scale
197 bar of 50 nm.

198 (B) The genomic map of *D. longicatena* phage CPB0989.

199 (C) The ANI heatmap of our *D. longicatena* phages and the previous assembled ones.

200

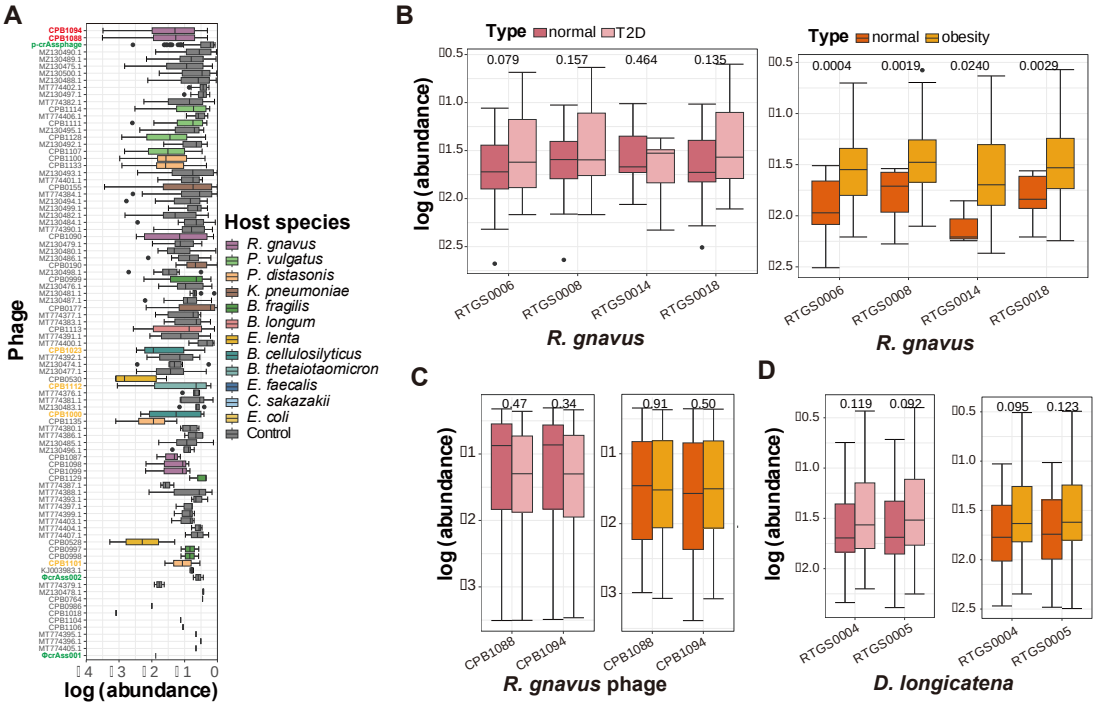


Figure S12 The relative abundances of GPB phages.

(A) The relative abundances of GPB phages (n=102) and crAss-like phages in ten cohorts from six continents. Each column represents a strain of phages, the colors of columns representing the host species of these phages (legend 'Control' represents the crAss-like phages reported). The colors of the y-axis fonts are consistent with Figure 4B.

(B) Left – The relative abundances of *M. gnavus* strains in normal healthy cohort and T2D cohort from Asia. Right - The relative abundances of *M. gnavus* strains in normal healthy cohort and obesity cohort from Asia.

(C) Left - The relative abundances of *M. gnavus* phages in normal healthy cohort and T2D cohort from Asia. Right - The relative abundances of *M. gnavus* phages in normal healthy cohort and obesity cohort from Asia.

(D) Left - The relative abundances of *D. longicatena* strains in normal healthy cohort and T2D cohort from Asia. Right - The relative abundances of *D. longicatena* strains in the normal healthy cohort and obesity cohort from Asia.

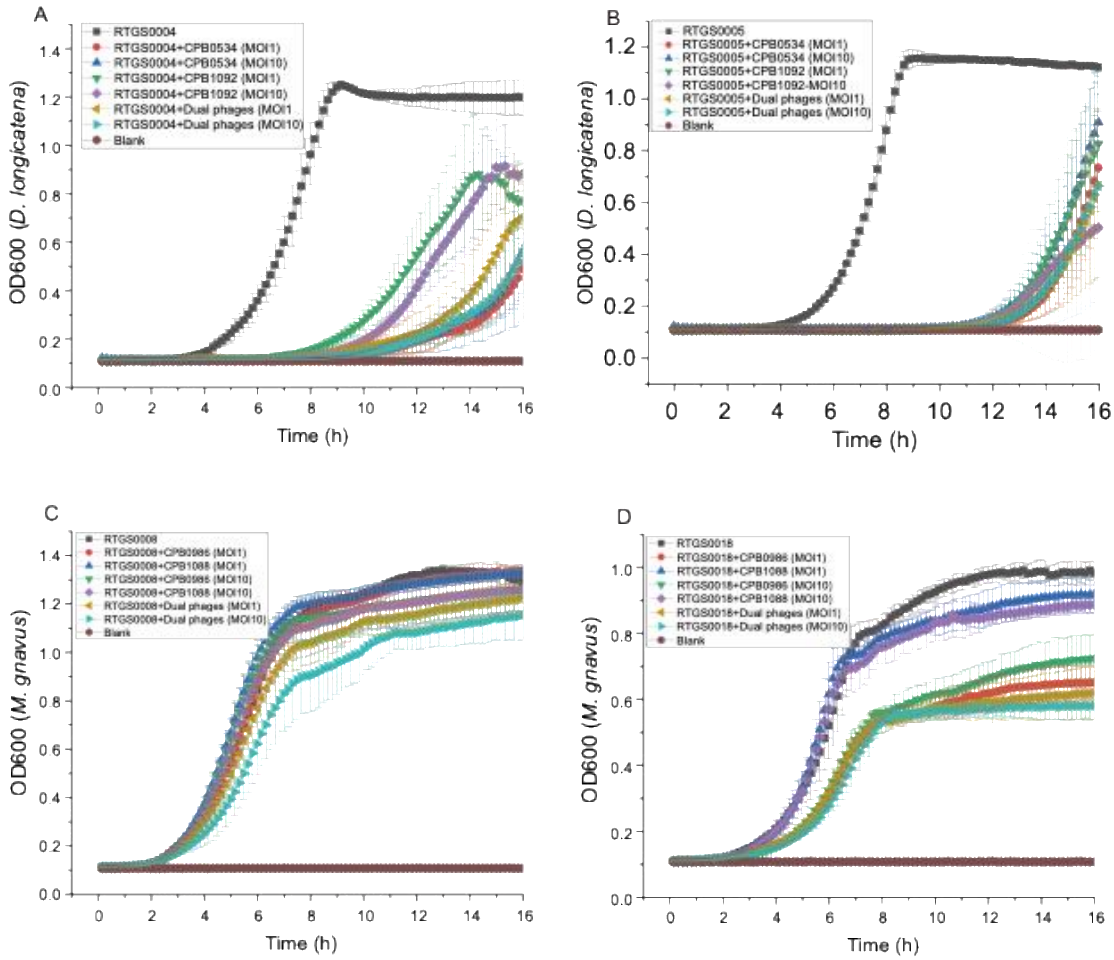


Figure S13 Killing curves of the GPB phages targeting *M. gnavus* and *D. longicatena*.

(A) The killing curves of CPB0534, CPB1092, and their mix targeting *D. longicatena* RTGS0004 under anoxic condition.

(B) The killing curves of CPB0534, CPB1092, and their mix targeting *D. longicatena* RTGS0005 under anoxic condition.

(C) The killing curves of CPB0986, CPB1088, and their mix targeting *M. gnavus* RTGS0008 under anoxic condition.

(D) The killing curves of CPB0986, CPB1088, and their mix targeting *M. gnavus* RTGS0018 under anoxic condition.

References

1 Shkoporov, A. N. *et al.* ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. *Nat Commun* 9, 4781 (2018). https://doi.org:10.1038/s41467-018-07225-7

2 Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* 184, 1098-1109 e1099 (2021). https://doi.org:10.1016/j.cell.2021.01.029

3 Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* 28, 724-740 e728 (2020). https://doi.org:10.1016/j.chom.2020.08.003

4 Tisza, M. J. & Buck, C. B. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc Natl Acad Sci U S A* 118 (2021). https://doi.org:10.1073/pnas.2023202118

5 Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol* 6, 960-970 (2021). https://doi.org:10.1038/s41564-021-00928-6