

Cyclic GMP–AMP signalling protects bacteria against viral infection

<https://doi.org/10.1038/s41586-019-1605-5>

Received: 25 June 2019

Accepted: 11 September 2019

Published online: xx xx xxxx

Daniel Cohen^{1,3}, Sarah Melamed^{1,3}, Adi Millman¹, Gabriela Shulman¹, Yaara Oppenheimer-Shaanan¹, Assaf Kacen², Shany Doron¹, Gil Amitai^{1*} & Rotem Sorek^{1*}

The cyclic GMP–AMP synthase (cGAS)–STING pathway is a central component of the cell-autonomous innate immune system in animals^{1,2}. The cGAS protein is a sensor of cytosolic viral DNA and, upon sensing DNA, it produces a cyclic GMP–AMP (cGAMP) signalling molecule that binds to the STING protein and activates the immune response^{3–5}. The production of cGAMP has also been detected in bacteria⁶, and has been shown, in *Vibrio cholerae*, to activate a phospholipase that degrades the inner bacterial membrane⁷. However, the biological role of cGAMP signalling in bacteria remains unknown. Here we show that cGAMP signalling is part of an antiphage defence system that is common in bacteria. This system is composed of a four-gene operon that encodes the bacterial cGAS and the associated phospholipase, as well as two enzymes with the eukaryotic-like domains E1, E2 and JAB. We show that this operon confers resistance against a wide variety of phages. Phage infection triggers the production of cGAMP, which—in turn—activates the phospholipase, leading to a loss of membrane integrity and to cell death before completion of phage reproduction. Diverged versions of this system appear in more than 10% of prokaryotic genomes, and we show that variants with effectors other than phospholipase also protect against phage infection. Our results suggest that the eukaryotic cGAS–STING antiviral pathway has ancient evolutionary roots that stem from microbial defences against phages.

Bacterial antiphage immune systems, such as CRISPR–Cas and restriction modification systems, tend to concentrate in ‘defence islands’ in bacterial genomes⁸; this property has facilitated the discovery of defence systems on the basis of their colocalization with known ones^{9–11}. We noticed that homologues of the gene that encodes cGAS in *V. cholerae* (*dncV*; ‘dinucleotide cyclase in *Vibrio*’) frequently tend to appear near defence genes. Out of 637 homologues of this protein that we identified through a homology search in 38,167 microbial genomes, we found that 417 (65.5%) are located in the vicinity of known defence systems (Fig. 1a). It has previously been shown that such a high propensity of colocalization with defence genes is a strong predictor that the gene under inspection has a role in phage resistance¹¹. These results therefore suggest that the gene that encodes the bacterial cGAS participates in defence against phages.

The bacterial cGAS and the effector phospholipase ‘cGAMP-activated phospholipase in *Vibrio*’ (CapV)⁷ are encoded by adjacent genes in the *V. cholerae* genome and are probably expressed in a single operon⁷. The same operon also contains two additional genes, the presence of which next to the *capV-dncV* gene pair is conserved in the majority of cases (96%, 613 out of 637 homologues), suggesting that the putative functional defence system that involves bacterial cGAS comprises these four genes (Fig. 1a). As has previously been noted^{7,12}, the two additional genes encode proteins with domains that are known to be associated with the eukaryotic ubiquitin system: the E1 and E2 domains that are typical of ubiquitin transfer and ligation enzymes, and a JAB domain,

which is similar to de-ubiquitinase enzymes that remove ubiquitin from target proteins (Fig. 1b).

To test whether the four-gene operon that contains cGAS is an antiphage defence system, we cloned the operon of *V. cholerae* serovar O1 biovar El Tor (hereafter, *V. cholerae* El Tor) or the homologous operon from *Escherichia coli* strain TW11681 into the laboratory strain *E. coli* MG1655, which naturally lacks this system. In both cases, the putative four-gene system was cloned together with its upstream and downstream intergenic regions to preserve promoters, terminators and other regulatory sequences. We then challenged *E. coli* MG1655 bacteria containing these four-gene systems with an array of phages that spans the three major families of tailed, double-stranded DNA phages (T2, T4, T6 and P1 from the *Myoviridae*; λ-vir, T5, SECPhi18 and SECPhi27 from the *Siphoviridae*; and T7 from the *Podoviridae*), as well as the single-stranded DNA phage SECPhi17, of the *Microviridae* family.

Both the *V. cholerae*-derived and the *E. coli*-derived four-gene operons conferred defence against multiple phages. The system from *E. coli* provided 10–1,800-fold protection against 6 of the 10 phages that we tested, and the system from *V. cholerae* protected against 2 of the phages (Fig. 1c, Extended Data Fig. 1). None of the systems protected against the transformation of a multi-copy plasmid (Extended Data Fig. 2). The stronger protective effect that was observed for the *E. coli*-derived system, as compared to that from *V. cholerae*, possibly stems from the higher compatibility of the former with the *E. coli* host and the coliphages

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. ²Department of Immunology, Weizmann Institute of Science, Rehovot, Israel. ³These authors contributed equally: Daniel Cohen, Sarah Melamed. *e-mail: gil.amitai@weizmann.ac.il; rotem.sorek@weizmann.ac.il

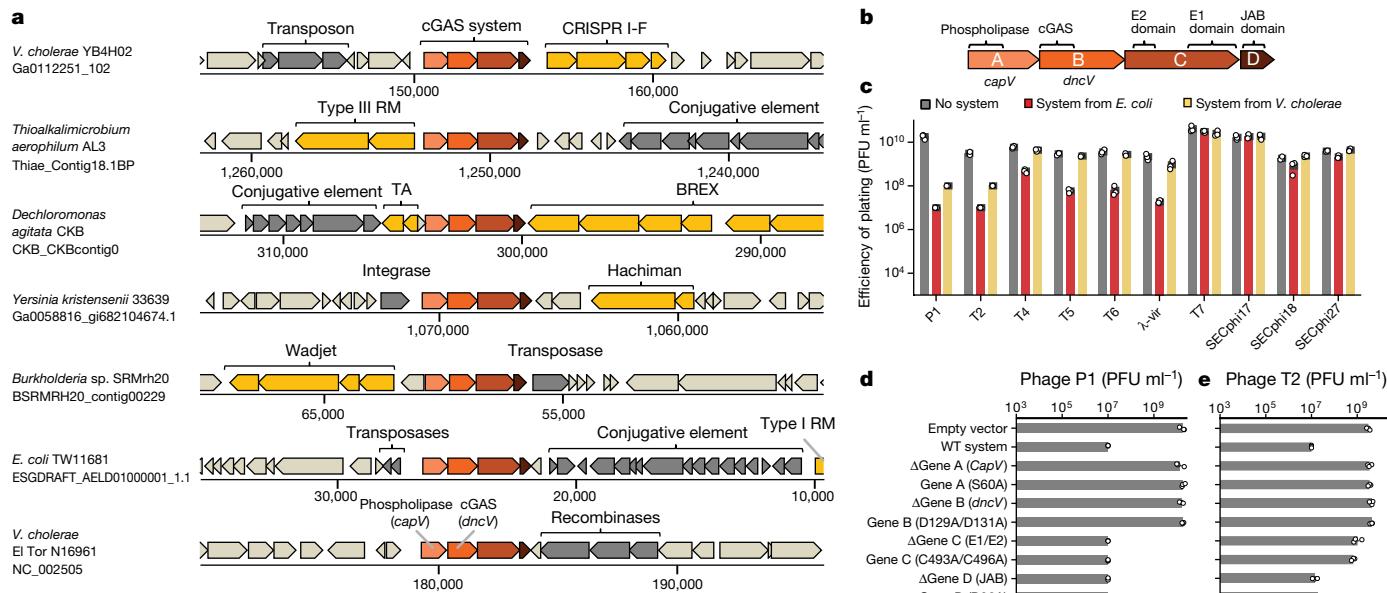


Fig. 1 | Systems containing bacterial cGAS protect against phage infection.

a, Genes that encode bacterial cGAS are part of a conserved four-gene operon that is genetically associated with antiphage defence systems. Representative instances of the four-gene putative defence systems that contain cGAS, and their genomic environments, are presented. Genes that are known to be involved in defence are shown in yellow. Genes of mobile genetic elements are in dark grey. RM, restriction modification; TA, toxin–antitoxin. BREX, Hachiman and Wadjet are recently described defence systems^{9,11}. **b**, Domain organization of the genes in the operon that contains cGAS. **c**, The four-gene operon from *V. cholerae* El Tor or *E. coli* TW11681 was cloned into *E. coli* MG1655 (Methods). The efficiency of plating is shown for ten phages infecting the control *E. coli* MG1655 strain (no system), the strain with the four-gene operon cloned from *E. coli* TW11681

(system from *E. coli*) and the strain with the four-gene operon cloned from *V. cholerae* El Tor (system from *V. cholerae*). Data represent plaque-forming units (PFU) per millilitre; bar graph represents average of three independent replicates, with individual data points overlaid. **d, e**, Efficiency of plating of phages infecting strains with the wild-type (WT) four-gene operon cloned from *E. coli* TW11681, deletion strains in that operon and strains with point mutations. Data represent plaque-forming units per millilitre; bar graph represents average of three independent replicates, with individual data points overlaid. Empty vector represents a control *E. coli* MG1655 strain that lacks the system and contains an empty vector instead. **d**, Infection with phage P1. **e**, Infection with phage T2.

that we used. We therefore proceeded with the *E. coli*-derived system for further experiments.

To determine whether the ability to produce and sense cGAMP affects defence, we experimented with mutated forms of the system. Deletion of either the cGAS-encoding or the phospholipase-encoding genes resulted in a complete loss of protection against phage infection (Fig. 1d, e, Extended Data Fig. 3a–d). Moreover, point mutations in two essential aspartate residues in the cGAMP-producing active site of cGAS (D129A/D131A)^{7,13} abolished defence, which suggests that cGAMP production is absolutely necessary for the phage-defensive properties of the system (Fig. 1d, e, Extended Data Fig. 3a–d). A S60A point mutation that inactivates the catalytic site of the CapV phospholipase also rendered the system completely inactive against all the phages that we tested, demonstrating that the cGAMP-controlled phospholipase activity of CapV is essential for defence (Fig. 1d, e, Extended Data Fig. 3a–d). These results indicate that, as in eukaryotes, the cGAS–cGAMP signalling pathway in bacteria participates in antiviral defence.

We next examined mutations in the two additional genes in the four-gene operon. Defence against the myophage P1 was not affected when the gene encoding the E1 and E2 domains or the gene encoding the JAB domain were deleted, which suggests that these two genes are not essential for protection against P1 (Fig. 1d). Accordingly, a construct that contained only the *capV-dncV* gene pair (that is, expressed only the phospholipase–cGAS pair) showed complete defence against P1 (Extended Data Fig. 4), demonstrating that the phospholipase–cGAS pair can work as a standalone defence system against some phages. However, deletion of the gene that encodes the E1 and E2 domains abolished defence against all of the other phages that were tested (T2, T4, T5, T6 and λ-vir). The presence of an intact gene that encodes the JAB-domain protein was necessary for protection against some phages

(T4, T5 and T6), but not against others (P1, T2 and λ-vir). Point mutations that are predicted to inactivate the active site of the E1 domain (C493A/C496A) or to inactivate the active site of the peptidase in the JAB domain (D38A) recapitulated the results of deletions of the entire genes, suggesting that the predicted enzymatic activities of the E1–E2 and JAB proteins are necessary for their roles in phage defence (Fig. 1d, e, Extended Data Fig. 3a–d). Our mutational analyses suggest that the four-gene operon forms a bacterial system that relies on cGAMP signalling and phospholipase activity to defend against a broad range of phages, and that the activities of the two additional genes are necessary for defence against some—but not all—phages. We denote this system the cyclic-oligonucleotide-based antiphage signalling system (CBASS).

In animals, the sensing of viral infection is the trigger that activates the production of cGAMP³. We therefore sought to examine whether phage infection triggers the production of cGAMP in the bacterial system as well. To this end, we took advantage of the fact that the CapV phospholipase is activated by cGAMP, and used this protein as a reporter for the presence of cGAMP in infected cells. We first expressed and purified the CapV phospholipase from the *E. coli* CBASS system, and measured its activity using an *in vitro* phospholipase activity assay. As expected, chemically synthesized 3'3'-cGAMP induced CapV phospholipase activity in a concentration-dependent manner (Fig. 2a). Next, we exposed the purified CapV to cell lysates that were derived from cells containing the *ΔcapV* CBASS, and which had been filtered to include small molecules only (Methods). The CapV phospholipase became markedly activated when exposed to cell lysates that were collected 40 min after infection with phage P1, which suggests that phage infection triggers cGAMP accumulation in cells that contain the CBASS (Fig. 2b). These results were corroborated by targeted mass spectrometry analysis, which detected cGAMP concentrations of 1.4–1.9 μM in lysates that were

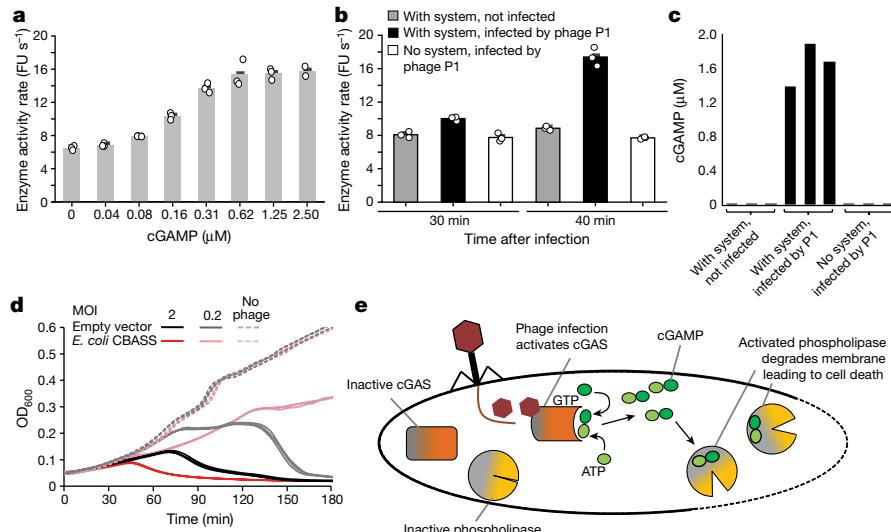


Fig. 2 | Phage infection triggers cGAMP accumulation and cell death.

a, Purified CapV protein was incubated in vitro with synthetically produced 3'3'-cGAMP in the presence of resorufin butyrate, a phospholipase substrate that emits fluorescence when hydrolysed. The x axis shows the concentration of cGAMP added (μM); the y axis shows the enzyme activity rate, measured by the accumulation rate of fluorescence units (FU) per second (Methods). Bar graph represents average of three technical replicates, with individual data points overlaid (except for the 2.5 μM concentration, for which the data represent an average of two replicates). **b**, Purified CapV protein was incubated in vitro with filtered cell lysates derived from bacterial cultures infected by phage P1 at a multiplicity of infection (MOI) of 2. Lysates were extracted from cells containing the CBASS with the *capV* gene deleted ('with system') or from control cells that lack the CBASS and contain an empty vector instead ('no system'). Lysates were collected 30 min or 40 min after initial infection. Bar graph

represents average of three technical replicates, with individual data points overlaid. **c**, Targeted mass spectrometry analysis was performed on the filtered cell lysates collected 40 min after initial infection by phage P1 at an MOI of 2. Lysates from uninfected samples were taken as control. The y axis represents the concentration of 3'3'-cGAMP in the cell lysate measured by mass spectrometry, as calculated on the basis of a calibration curve with synthetically produced 3'3'-cGAMP (Methods). Three independent replicates for each condition are shown, and each bar represents an individual replicate. **d**, Growth curves in liquid culture for bacteria that contain the CBASS and bacteria that lack the CBASS (empty vector), infected by phage P1 at 37 °C. Bacteria were infected at time 0 at an MOI of 0.2 or 2. OD₆₀₀, optical density at a wavelength of 600 nm. Three independent replicates are shown for each MOI, and each curve shows an individual replicate. **e**, Model for CBASS-based antiphage activity in bacteria.

collected 40 min after infection (Fig. 2c). CapV was not activated when exposed to lysates that were collected from uninfected cells, and only slightly activated by lysates that were collected from cells 30 min after infection. This implies that the phage component or cell state sensed by the CBASS appears in the cell relatively late in the phage-infection cycle (Fig. 2b).

It has previously been shown that, when stimulated by cGAMP, the CapV phospholipase degrades the membrane of *V. cholerae*, which leads to a loss of membrane integrity and to the arrest of cell growth or to death⁷. This, combined with our finding that the production of cGAMP is triggered by phage infection, led us to hypothesize that this defence system executes its defence via abortive infection. Abortive-infection defence systems exert their activity by causing the infected bacterial cell to 'commit suicide' before the phage replication cycle can be completed. This strategy eliminates infected cells from the bacterial population and protects the culture from a viral epidemic¹⁴. A phenotype such as this predicts that, with a high multiplicity of infection (MOI) (in which nearly all bacteria are infected in the first cycle), massive cell death will be observed in the culture, even for cells that contain the defence system.

To test this hypothesis, we infected bacteria growing in liquid cultures with phage P1 at varying MOIs and examined the culture dynamics. At an MOI of 0.2 (in which only around 20% of bacteria are initially infected by the phage) the culture of the wild-type cells collapsed owing to phage propagation, whereas the culture of the cells containing the CBASS was viable (Fig. 2d). Conversely, at an MOI of 2 (in which almost all bacteria are infected by the phage), the culture of wild-type cells and the culture of cells containing the CBASS both collapsed, which indicates cell lysis of the infected, CBASS-containing cells. Moreover, whereas at an MOI of 2 the wild-type culture started collapsing 70 min after initial infection, as has

previously been reported for phage P1¹⁵), the culture that contained the CBASS started collapsing as early as 45 min after infection—a period that is insufficient for the completion of the phage P1 replication cycle (Fig. 2d). These results were reproduced for other phages; when phages were applied at a high MOI, the cells that contained the CBASS showed lysis before the necessary time had elapsed for the completion of the phage replication cycle (Extended Data Fig. 5a). In the case of phage P1, the same phenotype was maintained for constructs that included only *capV-dncV* gene pair (that is, without the genes encoding the E1-E2 domains and JAB domain proteins), indicating that the latter two genes are not necessary for defence by abortive infection against phage P1 (Extended Data Fig. 5b).

These results were further corroborated by the staining of infected cells with propidium iodide, a fluorescent DNA-binding agent that is not membrane-permeable and cannot normally enter cells. Fluorescence-activated cell sorting (FACS) analysis showed that at 40 min after infection by phage P1 a substantial population of the cells that contain the CBASS became stained by propidium iodide, which indicates a loss of membrane integrity (Extended Data Fig. 6). This was associated with the loss of normal cell shape, as was observed via microscopy (Extended Data Fig. 7).

The above results suggest a model in which the cGAS–phospholipase defence system somehow senses a phage infection, and this sensing then triggers the production of cGAMP by the cGAS protein. The cGAMP molecule in turn activates the phospholipase, which degrades the bacterial membrane and causes cell death (Fig. 2e). This cell death before the completion of the phage replication cycle aborts the infection and prevents further propagation of the phage.

It has recently been shown that the bacterial cGAS gene (*dncV*) belongs to a large family of oligonucleotide cyclases, members of which are

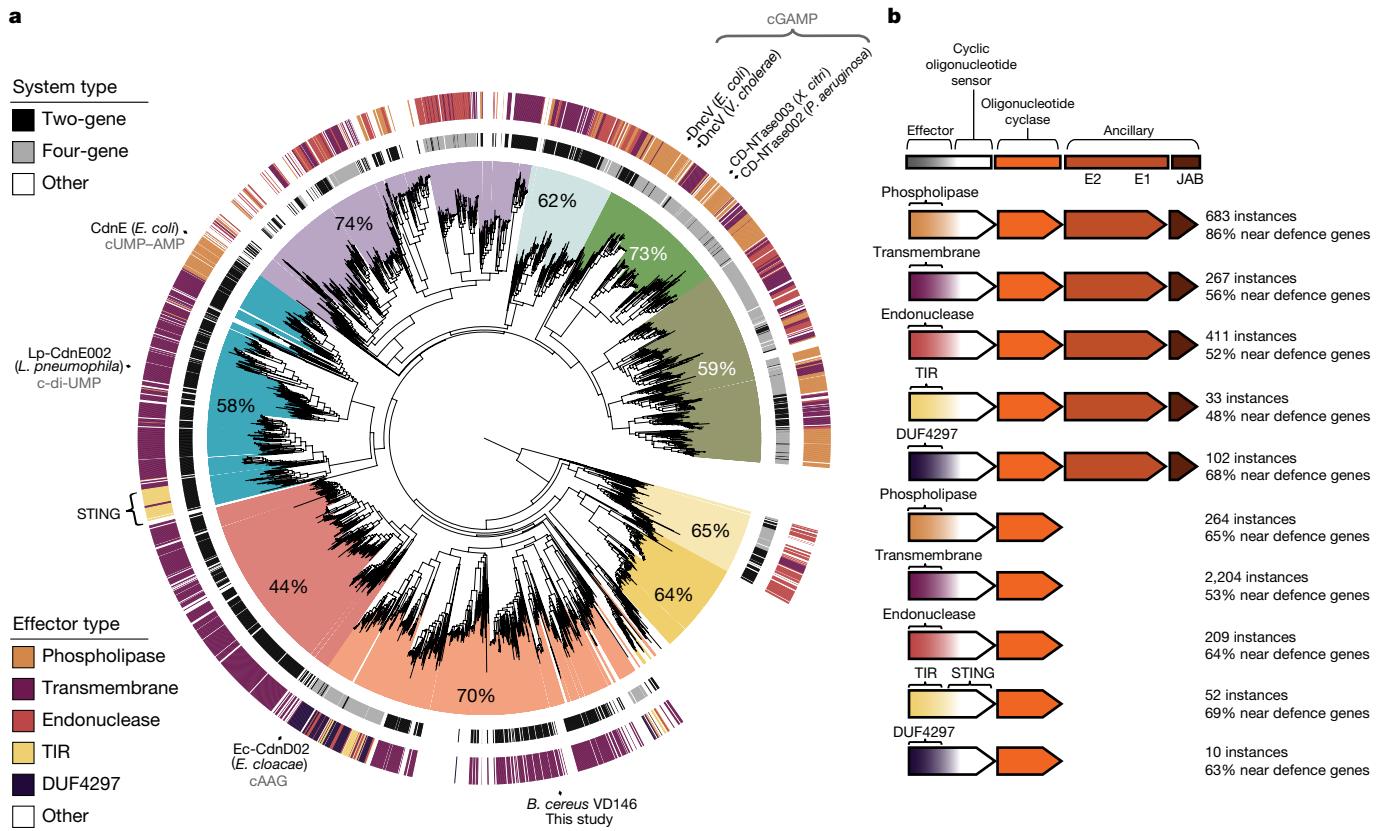


Fig. 3 | Widespread occurrence of CBASSs in prokaryotic genomes.

a, Phylogenetic tree of proteins with predicted oligonucleotide cyclase domains that were identified by searching a set of 38,167 microbial genomes (Methods). Clades are coloured following a previous colour coding¹⁶. The percentage of genes that are located near known defence systems is indicated for each clade. The outermost ring indicates the type of effector gene that is associated with the oligonucleotide cyclase gene; bottom left, colour code for the effector types. The inner ring indicates the type of system: the two-gene system (comprising an oligonucleotide cyclase and an effector) or four-gene system (that also includes

genes encoding the E1–E2 domains and JAB domain proteins). Top left, colour code for system type. Selected oligonucleotide cyclase genes that were biochemically characterized¹⁶ are indicated; the cyclic oligonucleotide that each produces appears in grey font. cAAG, cyclic trinucleotide AMP–AMP–GMP. *E. cloacae*, *Enterobacter cloacae*; *L. pneumophila*, *Legionella pneumophila*; *P. aeruginosa*, *Pseudomonas aeruginosa*; *X. citri*, *Xanthomonas citri*. **b**, Common configurations of four-gene and two-gene CBASSs. The identified number of instances of each such system, and the percentage occurrence next to known defence systems, are indicated for each configuration.

found in about 10% of all sequenced bacterial genomes¹⁶. Other genes in this family synthesize a diverse set of cyclic oligonucleotide molecules, including cyclic UMP–AMP, cyclic di-UMP and even the cyclic trinucleotide AMP–AMP–GMP¹⁶. This family of genes has been divided into several clades on the basis of sequence similarity between its members¹⁶. We found that all major clades of this gene family have a high propensity to be genetically associated with other known defence genes in defence islands (Fig. 3a). Between 44% and 74% of the genes in each clade are located near known defence systems, which suggests that this entire family of oligonucleotide cyclases participates in phage defence (Fig. 3a).

When examining the genomic environment of the 6,232 genes that we identified as belonging to this family (Supplementary Table 1, Methods), we found that in 26% of the cases (1,612 instances) they appeared as part of a four-gene operon that resembled the CBASS of *V. cholerae* (including the genes encoding the E1–E2 domains and the JAB-domain proteins). In 683 (42%) of these operons the predicted effector gene had a phospholipase domain, but in the remaining cases the phospholipase domain in the effector gene was replaced by another domain (Fig. 3b). Common alternative effector domains included endonucleases (usually of the HNH type); a domain comprising predicted transmembrane helices; a domain of unknown function (DUF4297) that has previously been shown to participate in the Lamassu phage-resistance system¹¹; and a Toll interleukin receptor (TIR) domain that was also previously shown to participate in antiphage defence¹¹ (Fig. 3b). These observations imply

that in these variants of the CBASS, the alternative domains replace the phospholipase in exerting the cell-suicide effector activity. For example, the endonuclease may degrade cellular DNA and the transmembrane-domain effector may oligomerize to form membrane pores, as in the case of the RexA–RexB abortive infection system¹⁷.

In addition, and consistent with previous reports¹⁸, we found 2,745 cases in which the genes that belong to the family of oligonucleotide cyclases appeared in the context of a two-gene operon that lacked the genes encoding the E1–E2 domains and JAB-domain proteins. The second gene in the operon also frequently included phospholipase, endonuclease or transmembrane helix domains, which suggests that these operons represent a minimal CBASS that comprises only two genes (Fig. 3b). The most common predicted effector domain in these operons was a domain that included either two or four transmembrane helices; effector domains of this type are present in over 2,000 instances of the two-gene operons that we identified.

To test whether predicted two-gene CBASSs have a role in phage resistance, we examined a two-gene system from *Bacillus cereus* VD146 that contains a predicted effector gene with four transmembrane helices (Extended Data Fig. 8a). We engineered this system into the laboratory strain *Bacillus subtilis* BEST7003 and challenged the engineered strain with an array of 11 *Bacillus* phages, as previously described¹¹ (Methods). The system conferred strong defence against one of these phages (the myophage SBSphiC), verifying its ability to defend against phages

(Extended Data Fig. 8b). Deletion of the cGAS-like gene or of the effector gene rendered the system inactive (Extended Data Fig. 8b). Finally, infection assays of bacteria that contained the two-gene system in liquid culture showed culture collapse at a high MOI, consistent with an abortive infection system (Extended Data Fig. 8c).

We observed 52 instances of a two-gene CBASS in which the effector gene contained an N-terminal TIR domain, and also a C-terminal STING domain (Fig. 3b, Extended Data Fig. 9). Such a domain arrangement of STING fused to a TIR domain is also found in primitive eukaryotes, including the oyster *Crassostrea gigas* and the annelid worm *Capitella teleta*¹. It is therefore possible that the CBASS with the TIR-STING effector gene represents the ancient evolutionary origin of the eukaryotic cGAS-STING system.

This study reveals the biological role of a large family of defence systems that is widespread in microbial genomes, but much remains unknown. The exact phage component sensed by the system is yet to be identified; it is unlikely that this component is cytoplasmic double-stranded DNA (as is the case in the animal cGAS-STING system), because bacteria do not have a nucleus and their cytoplasm therefore always contains double-stranded DNA. The role of the genes encoding the E1–E2 domains and JAB-domain proteins also remains unknown. These genes could be involved in the sensing of some phages, or in the mitigation of anti-cGAS activities that phages are likely to encode.

Accumulating evidence suggests that important components of the eukaryotic innate immune system have counterparts in bacterial immune systems. Argonaute—a central protein in the antiviral RNA interference machinery of plants, insects and animals¹⁹—has also been reported to have immunity roles in bacteria and archaea^{20,21}. TIR domains, which are essential components of the pathogen-recognizing Toll-like receptors²², are abundant in bacteria and have recently been shown to have a primary role in antiphage defence¹¹. Foreign RNA is sensed in eukaryotic cells by the oligo-adenylate synthase (OAS) protein, leading to the activation of a non-specific RNase²³; this process has recently been shown to have parallels in type III CRISPR–Cas immunity upon the sensing of phage RNA²⁴. All of these processes, in addition to our finding that cGAS signalling in prokaryotes has an antiviral role similar to that in eukaryotes, are unlikely to have been the result of parallel evolution. Instead, these observations cumulatively point to a scenario in which these defence systems first evolved in prokaryotes as means of defence against phages, and the ancient eukaryote (which was probably formed by fusion of a bacterium and an archaeon²⁵) inherited a primordial version of these systems from the prokaryotes that formed it. Under this hypothesis, these systems became the basis for the primitive immune system of the ancient eukaryote, and have evolved into the cell-autonomous immune system that we know today. If this hypothesis is correct, future studies may find homologues of additional components of the human immune system functioning as phage resistance systems in bacteria.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1605-5>.

- Kranzusch, P. J. et al. Ancient origin of cGAS-STING reveals mechanism of universal 2',3' cGAMP signaling. *Mol. Cell* **59**, 891–903 (2015).
- Margolis, S. R., Wilson, S. C. & Vance, R. E. Evolutionary origins of cGAS-STING signaling. *Trends Immunol.* **38**, 733–743 (2017).
- Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
- Ablässer, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).
- Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).
- Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**, 358–370 (2012).
- Severini, G. B. et al. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **115**, E6048–E6055 (2018).
- Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011).
- Goldfarb, T. et al. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
- Ofir, G. et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).
- Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pan-genome. *Science* **359**, eaar4120 (2018).
- Iyer, L. M., Burroughs, A. M. & Aravind, L. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. *Genome Biol.* **7**, R60 (2006).
- Kato, K., Ishii, R., Hirano, S., Ishitani, R. & Nureki, O. Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP-AMP synthase. *Structure* **23**, 843–850 (2015).
- Molineux, I. J. Host–parasite interactions: recent developments in the genetics of abortive phage infections. *New Biol.* **3**, 230–236 (1991).
- Walker, J. T. & Walker, D. H. Mutations in coliphage P1 affecting host cell lysis. *J. Virol.* **35**, 519–530 (1980).
- Whitelley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
- Snyder, L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.* **15**, 415–420 (1995).
- Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).
- Joshua-Tor, L. & Hannon, G. J. Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb. Perspect. Biol.* **3**, a003772 (2011).
- Swarts, D. C. et al. DNA-guided DNA interference by a prokaryotic argonaute. *Nature* **507**, 258–261 (2014).
- Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* **51**, 594–605 (2013).
- Akira, S. & Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **4**, 499–511 (2004).
- Zhou, A. et al. Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997).
- Kazlauskienė, M., Kostiuk, G., Venclavas, Č., Tamulaitytė, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR–Cas systems. *Science* **357**, 605–609 (2017).
- Margulis, L. Archaeal–eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl Acad. Sci. USA* **93**, 1071–1076 (1996).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Genomic identification and analysis of DncV homologues

The protein sequence of DncV (NCBI accession NP_229836) was searched against the protein sequences of all genes in 38,167 bacterial and archaeal genomes downloaded from the Integrated Microbial Genomes (IMG) database²⁶ in October 2017, using the ‘search’ option in the MMseqs2 package²⁷ (release 6-f5a1c) with default parameters. Hits with an evalue less than 1×10^{-20} were taken as homologues. The fraction of homologues found in the vicinity of known defence systems was calculated as previously described¹¹, using a positive set of defence gene families that was updated to include a recently discovered set of defence genes¹¹.

Genomic analysis of oligonucleotide cyclase genes

For the analysis in Fig. 3, the proteins from the database of 38,167 bacterial and archaeal genomes were first clustered using the ‘cluster’ option of MMseqs2²⁷ (release 2-1c7a89), with default parameters. Clusters were further aggregated into larger clusters using four additional cycles of clustering, in which—in each cycle—a representative sequence was taken from each cluster using the ‘createsubdb’ option of MMseqs2 and representative sequences were clustered using the ‘cluster’ option with the ‘–add-self-matches’ parameter. For the first additional clustering cycle, the ‘cluster’ option was run with default parameters; for the additional cycles 2–4, clustering was run with sensitivity parameter ‘–s 7.5’, and for the additional cycle 4, the ‘–cluster-mode 1’ parameter was also added.

The sequences of each cluster were aligned using Clustal Omega²⁸. Each multiple sequence alignment was scanned with HHpred²⁹ using 60% gap rule (–M 60) against the PDB_mmCIF70³⁰ and pfam31³¹ databases. Clusters with HHpred hits to one of the cGAS entries (Protein Data Bank (PDB) codes: 4LEV, 4MKP, 4O67, 5VDR, 5V8H, 4LEW, 5VDP, 4KMS, 4O68, 4O69, 5V8J, 5V8N, 5V8O, 5VDO, 5VDQ, 5VDS, 5VDT, 5VDU, 5VDV, 5VDW, 4XJ5, 4XJ1, 4XJ6, 4XJ3 and 4XJ4, and pfam PF03281) with >90% probability in the top 30 hits were taken for manual analysis. Clusters containing genes that were suspected to belong to toxin–antitoxin gene pairs were discarded. Overall, this procedure identified 30 clusters containing 6,232 predicted oligonucleotide cyclase genes from 5,150 genomes.

The genomic environments spanning 10 genes upstream and downstream of each of the 6,232 predicted oligonucleotide cyclase genes were searched to identify conserved gene cassettes and known defence genes around the oligonucleotide cyclase genes, as previously described¹¹. Predicted systems were manually reviewed and unrelated genes (for example, mobilome genes and genes of other defence systems) were omitted.

To generate the phylogenetic tree in Fig. 3a, the ‘clusthash’ option of MMseqs2 (release 6-f5a1c) was first used to remove protein redundancies (using the ‘–min-seq-id 0.9’ parameter). Sequences shorter than 200 amino acids were also removed. To further remove outlier sequences, an all-versus-all search was conducted using the ‘search’ option in MMseqs2, and proteins with less than 20 hits were manually examined. Overall, 17 outlier proteins were removed this way. The human cGAS protein (UniProt Q8N884) was added, as well as the human oligoadenylylate synthase genes (UniProt P00973, P29728 and Q9Y6K5); these were used as an outgroup. Sequences were aligned using MAFFT²⁸ with gap open penalty of ‘–op 2’. The FastTree software³² was used to generate a tree from the multiple sequence alignment using default parameters. The iTOL³³ software was used for tree visualization.

Cloning of the CBASS into *E. coli* MG1655

The following full-length constructs containing the CBASS were designed: the four-gene CBASS from *V. cholerae* El Tor N16961 (GenBank accession NC_002505) was identified as locus tags VC0178–VC0181,

and the operon was taken together with its upstream and downstream intergenic regions, spanning the nucleotide range 178,424–183,957 in GenBank accession NC_002505. The four-gene system from *E. coli* TW11681 (GenBank accession AELD00000000) was identified as locus tags ESGDRAFT_00026–ESGDRAFT_00029 and the operon, together with its upstream and downstream intergenic regions, spanned the nucleotide range 21,738–27,072 in GenBank accession AELD00000000. These two constructs were commercially synthesized and cloned by Genscript directly into the plasmid pSG1-rfp^{10,11} between the Ascl and NotI sites of the multiple cloning site.

For strains with gene deletions and point mutations, plasmids containing systems with these deletions or mutations were also commercially synthesized and cloned into the plasmid pSG1-rfp by Genscript, except for one construct in which the genes for both the E1–E2 domains and JAB domain were deleted (used for Extended Data Figs. 4, 5b). To build this construct, the pSG1-rfp backbone was amplified with primers P1-Fw + P1-Rv using KAPA HiFi HotStart ReadyMix (Kapa Biosystems KK2601) (Supplementary Table 2). Primers P2-Fw + P2-Rv were used to lift the gene pair *capV* and *dncV* with their native promoter from the plasmid that contained the full system derived from *E. coli* TW11681 (Supplementary Table 2). After gel purification of the PCR products with Zymoclean Gel DNA Recovery Kit (cat. no. D4001), the two fragments were assembled using NEBuilder HiFiDNA Assembly cloning kit (NEB E5520S) to produce the construct that lacks both gene C (encoding the protein with the E1 and E2 domains) and gene D (encoding the JAB-domain protein).

The plasmids were transformed into *E. coli* MG1655 cells by electroporation, and the resulting transformants of the wild-type and mutated systems were verified by whole genome sequencing as previously described¹⁰ to verify system integrity and a lack of mutations. A negative control was constructed as a transformant containing an empty pSG1 plasmid.

Cloning of the CBASS from *B. cereus* into *B. subtilis* BEST7003

The two-gene CBASS from *B. cereus* VD146 (GenBank accession KB976672) was identified as locus tags IK1_05630–IK1_05631, and the operon was taken together with its upstream and downstream intergenic regions, spanning the nucleotide range 60,974–63,493 in GenBank accession KB976672. The construct, as well as two additional constructs in which one of the two genes was deleted, were commercially synthesized and cloned by Genscript directly into the plasmid pSG1-rfp^{10,11} between the Ascl and NotI sites of the multiple cloning site. The plasmid was transformed into *B. subtilis* and integrated into the *amyE* locus as previously described¹¹. A negative control was constructed as a transformant containing an empty pSG1 plasmid integrated in the *amyE* locus. The resulting transformants of the wild-type and mutated systems were verified by whole genome sequencing as previously described¹⁰.

Phage cultivation

E. coli phages (P1, T4, T5, T7 and λ-vir) were provided by U. Qimron. Phages SECphi17, SECphi18 and SECphi27 were isolated in our laboratory¹¹. T2 and T6 were ordered from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (DSM16352 and DSM4622, respectively). The following *B. subtilis* phages were obtained from the *Bacillus* Genetic Stock Center (BGSC): SPO1 (BGSCID 1P4), φ3T (BGSCID 1L1), SPβ (BGSCID 1L5), SPR (BGSCID 1L56), φ105 (BGSCID 1L11), p14 (BGSCID 1L15), SPP1 (BGSCID 1P7), SP82G (BGSCID 1P5). Phage φ29 was obtained from the DSMZ (DSM 5546). Phages SBSphij and SBSphic were isolated in our laboratory¹¹. Phages were propagated on *E. coli* MG1655 or *B. subtilis* BEST7003 in liquid culture, and their titre was determined using the small drop plaque assay method, as previously described¹¹.

Plaque assays

Bacteria were mixed with MMB agar (LB + 0.1 mM MnCl₂ + 5 mM MgCl₂ + 0.5% agar), and tenfold serial dilutions of the phage lysate in MMB were dropped on top of them. After the drops were dry, plates were

incubated overnight at 37 °C for *E. coli* phage and at room temperature for *B. subtilis* phages. Plaques were counted to calculate the efficiency of plating in plaque-forming units per millilitre. For phages showing a fuzzy killing zone in which single plaques could not be counted, the lowest phage concentration in which a killing zone was observed was counted as ten plaques. Fold defence was calculated as the efficiency of plating on control bacteria divided by the efficiency of plating value obtained on bacteria containing the CBASS.

Phage-infection dynamics in liquid medium

Overnight cultures were diluted 1:100 in MMB medium and incubated at 37 °C while shaking at 250 r.p.m. until early log phase ($OD_{600} = 0.3$). One hundred and eighty microlitres of the diluted culture were transferred into wells in a 96-well plate containing 20 µl of phage lysate for a final MOI of 2, 0.2 or 0.02 as applicable. Infections were performed in triplicate and OD_{600} was followed using a TECAN Infinite 200 plate reader with measurement every 5 min.

Transformation efficiency assay

To prepare electro-competent cells, *E. coli* MG1655 cells (with or without the CBASS) were diluted 1:100 in 100 ml LB medium supplemented with ampicillin, 100 µg/ml. At $OD_{600} = 0.6$, the cells were transferred to ice for 15 min and then centrifuged for 15 min at 4,000 r.p.m. The supernatant was discarded and the pellet was resuspended in ultrapure ice-cold water. This step was repeated twice and the supernatant was removed. The resultant pellet was resuspended in 10% glycerol, centrifuged for another 10 min and the supernatant was discarded. The final pellet was mixed in 500 µl of 10% glycerol and aliquots of 50 µl were flash-frozen in liquid nitrogen and transferred immediately to -80 °C.

One hundred nanograms of pRSFDuet-1 was added to 50 µl electro-competent cells and the mixture was transferred to a Bio-Rad Gene Pulser Curvette (0.2 cm, cat. no. 165-2086). The cells were electroporated with a Bio-Rad Micropulser using the 'Ec1' setting and then immediately transferred to 1 ml LB medium to recover at 37 °C for 1 h. After incubation cells were diluted 1:100, and 100 µl were plated on LB plates containing ampicillin (100 µg/ml) and kanamycin (50 µg/ml), and incubated at 37 °C overnight. Transformation efficiency was calculated by dividing the number of transformants that grew on LB plates containing ampicillin and kanamycin by the live count grown on LB ampicillin (100 µg/ml) only.

Cloning, expression and purification of CapV from *E. coli* TW11681

The CapV protein from *E. coli* TW11681 was PCR-amplified (primers P3-Fw + P3-Rv, and P4-Fw + P4-Rv), fused with a C-terminal His tag (linker plus His tag sequence: SerGly₄His₆) and cloned into pET28b. The plasmid was transformed into BL21(DE3)pLys cells and grown at 37 °C, 250 r.p.m., until induction (1 mM IPTG) at an OD_{600} of 0.6. After induction, bacterial cell culture growth was continued for an additional 14 h at a temperature of 18 °C. Cells were then centrifuged for 10 min (3,900g, 4 °C) and pellets kept at -20 °C. Pellets were thawed on ice and resuspended with an ice-cold buffer containing 50 mM phosphate buffer (pH 7.4), 300 mM NaCl, 10% glycerol (v/v). The buffer was supplemented with 10 µg/ml lysozyme, and 1 µl benzonase-nuclease (SKU Merck). The suspension was then mixed with Lysing matrix B (MP) beads and cells were disrupted mechanically using a FastPrep-24 (MP) bead-beater device (2 cycles of 40 s, 6 m s⁻¹, at 4 °C). Cell lysate was centrifuged at 12,000g for 10 min at 4 °C and the supernatant was then mixed with Ni-NTA magnetic agarose beads (Qiagen) for 2 h (4 °C). CapV-6×His proteins bound to Ni-NTA beads were washed 3 times with a 50 mM phosphate buffer (pH 7.4), 300 mM NaCl, 10% glycerol (v/v), 20 mM imidazole and then eluted with the 50 mM phosphate buffer (pH 7.4), 300 mM NaCl, 10% glycerol (v/v), 250 mM imidazole. The eluent was loaded onto an Amicon Ultra-0.5 centrifugal filter unit 10-kDa filter (Merck) to exchange the elution buffer with the reaction buffer (50

mM phosphate buffer (pH 7.4), 300 mM NaCl, 10% glycerol (v/v)). Buffer exchange was done by centrifuging the centrifugal filter unit at 14,000g for 10 min at 4 °C 4 times with the reaction buffer. Eluted protein purification was assessed by running the sample on an SDS-PAGE gel and quantity was measured by using a Qubit Protein Assay Kit (Thermo Fisher Scientific).

Fluorogenic biochemical assay for CapV activity

The esterase activity of the 6×His-tagged CapV was probed with the fluorogenic substrate resorufin butyrate. The 6×His-tagged CapV was diluted in 50 mM sodium phosphate (pH 7.4), 300 mM NaCl, 10% (v/v) glycerol to a final concentration of 1.77 µM. To measure the linear range of CapV activation, purified 6×His-tagged CapV was incubated for 5 min with increasing concentrations of 3'3'-cGAMP (Merck), ranging from 0.04 to 2.5 µM of 3'3'-cGAMP. In parallel, 6×His-tagged CapV was incubated for 5 min with cell lysates derived from *E. coli* cells infected with phage P1 or uninfected. Subsequently, the enzyme-cGAMP or enzyme-lysate solution was added to DMSO-solubilized resorufin butyrate (stock of 20 mM mixed with 50 mM sodium phosphate (pH 7.4), 300 mM NaCl, 10% (v/v) glycerol reaching a final concentration of 64 µM (final assay DMSO concentration was 0.32%)) to a final assay volume of 50 µl, and fluorescence was measured in a 96-well plate (Corning 96-well half area black non-treated plate with a flat bottom). Plates were read once every 30 s for 10 min at 37 °C using an Infinite-200 (Tecan) with excitation and emission wavelengths of 550 and 591 nm, respectively. The enzymatic reaction velocity was measured as previously described³⁴. For this, a regression fit was calculated for the output values of each reaction over time, and the slope of this linear regression fit was used for determining the initial reaction velocity (FU s⁻¹).

Cell lysate preparation

E. coli MG1655 cells containing the *E. coli* TW11681-derived CBASS in which the *capV* gene was deleted were used for preparation of cell lysates. Cells containing an empty vector (pSG1) were used as control. Cells were grown in 100 ml MMB medium (flask size 250 ml) at 37 °C (250 r.p.m.) until reaching an OD_{600} of 0.3. Cells were then infected with 5 ml of phage P1 (titre of 10^{10} infective particles per millilitre, estimated MOI of 2). After 30 or 40 min from the initial infection, samples were collected and centrifuged at 3,900g for 5 min at 4 °C. Following centrifugation, pellets were kept on ice until resuspended in 600 µl buffer containing 50 mM sodium phosphate (pH 7.4), 300 mM NaCl and 10% (v/v) glycerol. The resuspended pellet was supplemented with 1 µl hen-lysozyme (Merck) (final hen-lysozyme concentration of 16 µg/ml). The resuspended cells were then mixed with Lysing matrix B (MP) beads and cells were disrupted mechanically using a FastPrep-24 (MP) bead-beater device (2 cycles of 40 s, 6 m s⁻¹, at 4 °C). Cell lysate was then centrifuged at 12,000g for 10 min at 4 °C and the supernatant was loaded onto a 3-kDa filter Amicon Ultra-0.5 centrifugal filter unit (Merck) and centrifuged at 14,000g for 30 min at 4 °C. The flow-through-containing substances smaller than 3 kDa – was used as the lysate sample for evaluating cGAMP production within the cells (Fig. 2).

Quantification of 3'3'-cGAMP by high-performance liquid chromatography and mass spectrometry (HPLC-MS)

Cell lysates were prepared as described in 'Cell lysate preparation'. Lysates collected at 40 min after infection were analysed by MS-Omics using ultra-performance liquid chromatography (UPLC) (Vanquish, Thermo Fisher Scientific) coupled with a high-resolution quadrupole-orbitrap mass spectrometer (Q Exactive HF Hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific). An electrospray ionization interface was used as an ionization source. Analysis was performed in positive ionization mode. A calibration series of 3'3'-cGAMP (SML1232, Sigma-Aldrich) ranging 0.001 to 50 µM was prepared and a linear regression from 0.001 to 5 µM was used for cGAMP quantification.

FACS analysis of infected cells

Overnight cultures of *E. coli* MG1655 cells containing the *E. coli* TW11681-derived CBASS, and *E. coli* MG1655 cells containing an empty vector (pSG1), were diluted 1:100 in 0.5 ml MMB and grown at 37 °C and 500 rpm to an OD₆₀₀ of 0.3. Cells were then infected with phage P1 (titre of 10¹⁰ infective particles per millilitre, estimated MOI of 2). At 40 min after infection, 40 µl of each culture were diluted into 2 ml of filtered PBS containing 1 µl of propidium iodide (Invitrogen LIVE/DEAD BacLight Bacterial Viability Kit (L7007)). The diluted bacteria were incubated in the dark in room temperature for five minutes and were then analysed by a BIORAD ZE5 Cell Analyzer. A 5-s agitation was performed and then 25,000 ungated events were recorded for both the CBASS-containing and CBASS-lacking cultures at 0.2 µls⁻¹. Forward scatter and propidium iodide fluorescence were measured using the height (*H*) parameter. FlowJo v.10 was used to analyse and visualize the data.

Microscopy of infected cells

E. coli MG1655 cells that contain the *E. coli* TW11681-derived CBASS or the same CBASS with a point mutation of the phospholipase catalytic site (S60A) were grown in MMB medium at 37 °C. When growth reached an OD₆₀₀ of 0.3, bacteria were infected with P1 phage (MOI of about 2). Five hundred microlitres of the infected samples were centrifuged at 10,000*g* for 2 min at 25 °C and resuspended in 5 µl of 1× phosphate-buffered saline (PBS), supplemented with 1 µg/ml membrane stain FM4-64 (Thermo Fisher Scientific T-13320) and 2 µg/ml DNA stain 4,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich D9542-5MG). Cells were visualized and photographed using an AxioPlan2 microscope (ZEISS) equipped with ORCA Flash 4.0 camera (HAMAMATSU). System control and image processing were carried out using Zen software version 2.0 (Zeiss).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data that support the findings of this study are available within the article and its Extended Data and Supplementary Tables. GenBank accessions, locus tags and nucleotide ranges of the CBASSs appear in the Methods. IMG gene and genome ID number, contig ID and system and effector

classification appear in Supplementary Table 1. Primer sequences for the CBASSs are available in Supplementary Table 2. Any other relevant data are available from the corresponding authors upon reasonable request.

26. Chen, I. A. et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
27. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
28. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
29. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
30. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003).
31. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
32. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
33. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
34. Lam, V. et al. Resorufin butyrate as a soluble and monomeric high-throughput substrate for a triglyceride lipase. *J. Biomol. Screen.* **17**, 245–251 (2012).
35. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).

Acknowledgements We thank A. Leavitt and S. Sharir for assistance in DNA extraction, library preparation and sequencing, A. Bernheim for assistance in data visualization, and members of the Sorek laboratory for fruitful discussions. This study was supported in part by the Israel Science Foundation (personal grant 1360/16), the European Research Council (grant ERC-CoG 681203), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, and the Knell Family Center for Microbiology. A.M. was supported by a fellowship from the Ariane de Rothschild Women Doctoral Program.

Author contributions D.C., S.M. and G.A. led the study and performed all experiments unless otherwise indicated. A.M. performed the computational analyses that appear in Figs. 1 and 3. Y.O.-S. performed the microscopy analysis that appears in Extended Data Fig. 7. G.S. assisted with the plaque assays that appear in Fig. 1 and Extended Data Figs. 1 and 3. A.K. and S.D. performed the computational analyses that led to Extended Data Fig. 8. R.S. supervised the study and wrote the paper together with the team.

Competing interests R.S. is a scientific cofounder and consultant of BiomX Ltd, Pantheon Ltd and Ecophage Ltd.

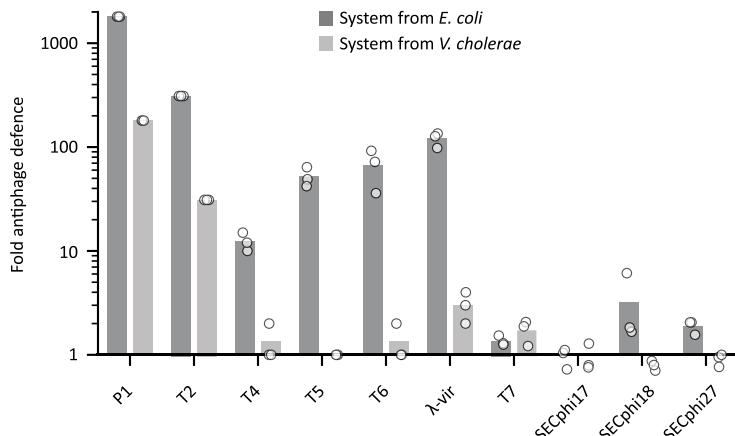
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1605-5>.

Correspondence and requests for materials should be addressed to G.A. or R.S.

Peer review information *Nature* thanks Zhijian ‘James’ Chen, Karen Maxwell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

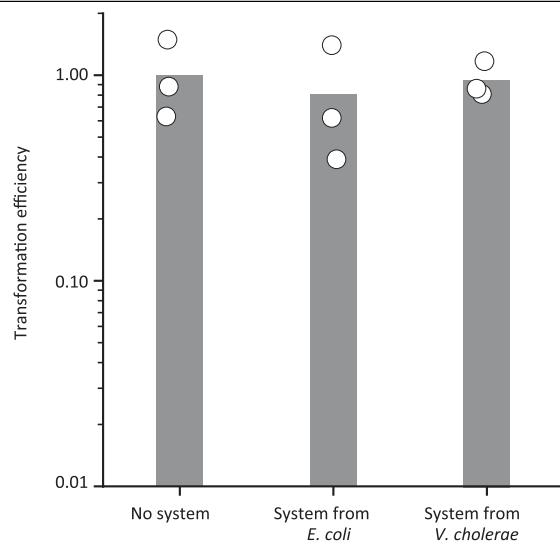
Reprints and permissions information is available at <http://www.nature.com/reprints>.



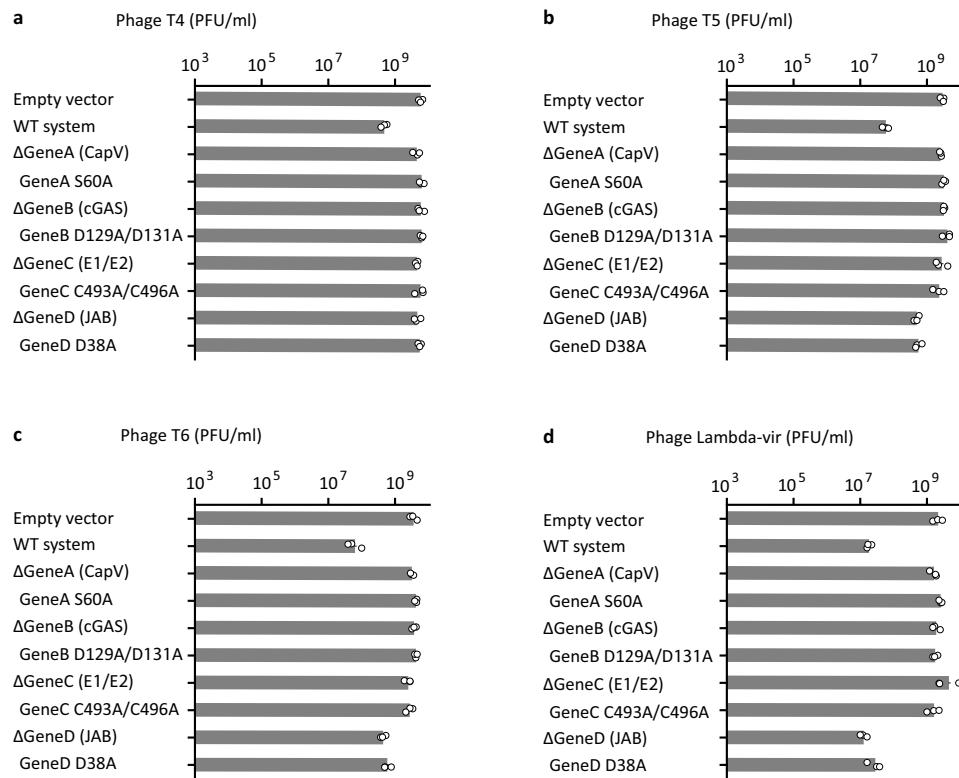
Extended Data Fig. 1 | Fold antiphage defence conferred by four-gene defence systems against various phages. The four-gene operon from either *V. cholerae* El Tor or *E. coli* TW11681 was cloned into *E. coli* MG1655 (Methods). Fold antiphage defence, as measured by plaque assays, is shown. The fold defence was calculated as the ratio between the efficiency of plating of the

phage on the operon-lacking control strain and the efficiency of plating on the operon-containing strain (Fig. 2b, Methods). Bar graph represents average of three independent replicates, with individual data points overlaid. Points that fall below the x axis (for SECphi17, SECphi18 and SECphi27) denote values lower than 1.

Article



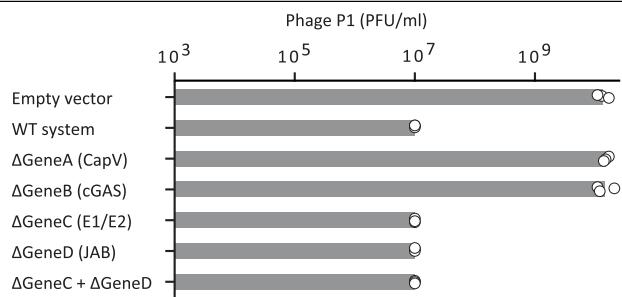
Extended Data Fig. 2 | Transformation efficiency assays. Transformation efficiency of plasmid pRSFDuet-1 into strains that contain the four-gene operon derived from *E. coli* TW11681 or from *V. cholerae* El Tor, presented as a fraction of the transformation efficiency to *E. coli* MG1655 carrying an empty vector instead of the four-gene operon. Bar graph represents average of three independent replicates, with individual data points overlaid.



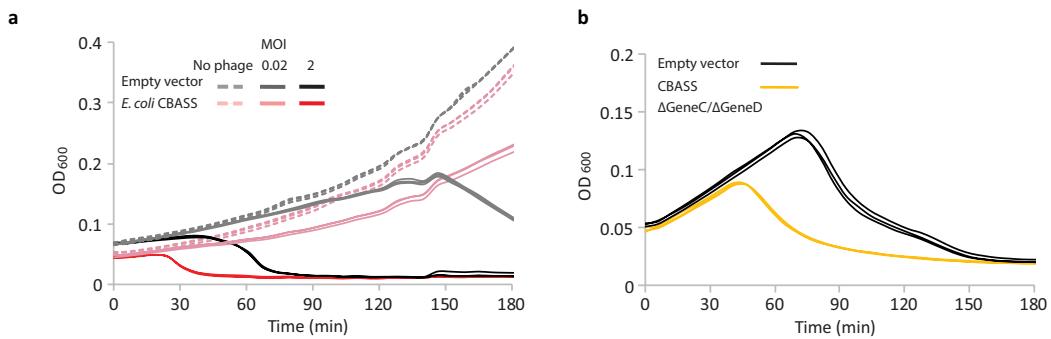
Extended Data Fig. 3 | Efficiency of plating of coliphages on defence systems with whole-gene deletions or point mutations. The efficiency of plating of phages infecting strains with the wild-type *E. coli*-derived four-gene, deletion strains and strains with point mutations. Data represent plaque-forming units per millilitre; bar graphs represent average of three independent replicates, with

individual data points overlaid. Empty vector represents a control *E. coli* MG1655 strain that lacks the system and has an empty vector instead. **a**, Infection with the phage T4. **b**, Infection with the phage T5. **c**, Infection with the phage T6. **d**, Infection with the phage λ-vir.

Article

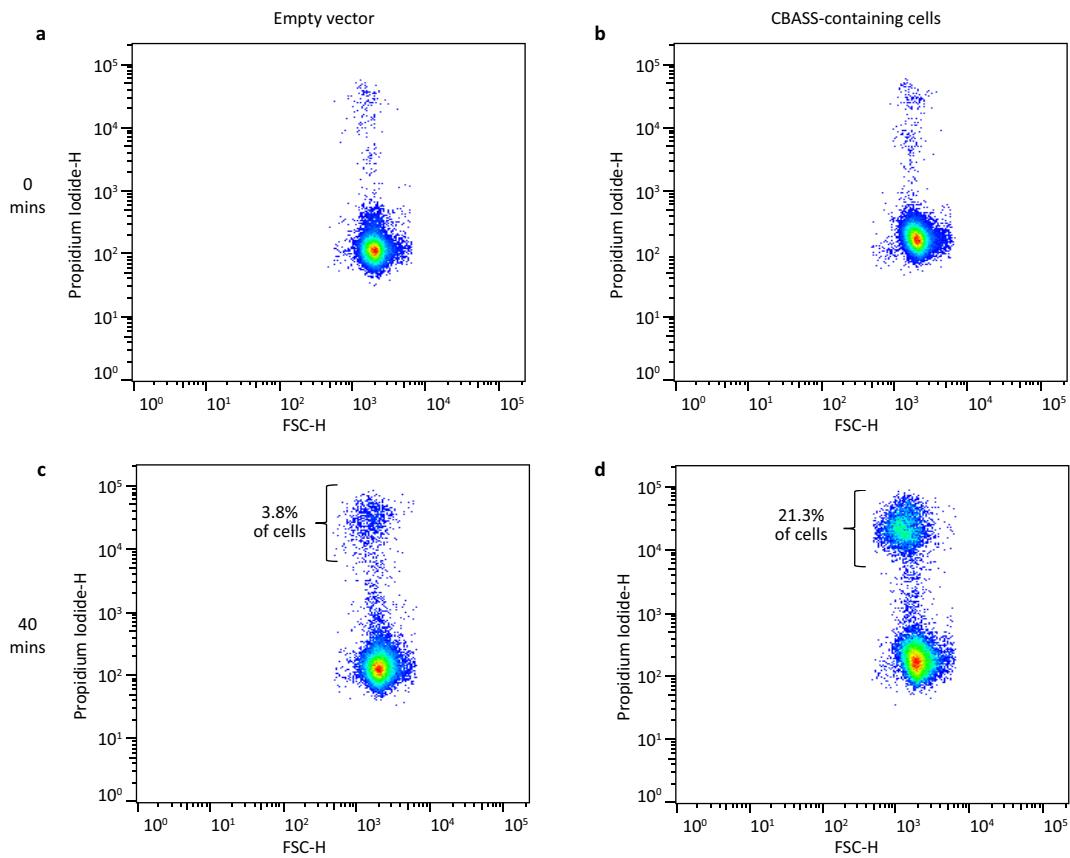


Extended Data Fig. 4 | Efficiency of plating of phage P1 on a double-deletion strain. The efficiency of plating is shown of phage P1 infecting strains with the wild-type *E. coli*-derived four-gene system, strains with individual genes deleted and a strain with two genes deleted. Data represent plaque-forming units per millilitre; bar graphs represent average of three independent replicates, with individual data points overlaid. Empty vector represents a control *E. coli* MG1655 strain that lacks the system and has an empty vector instead.



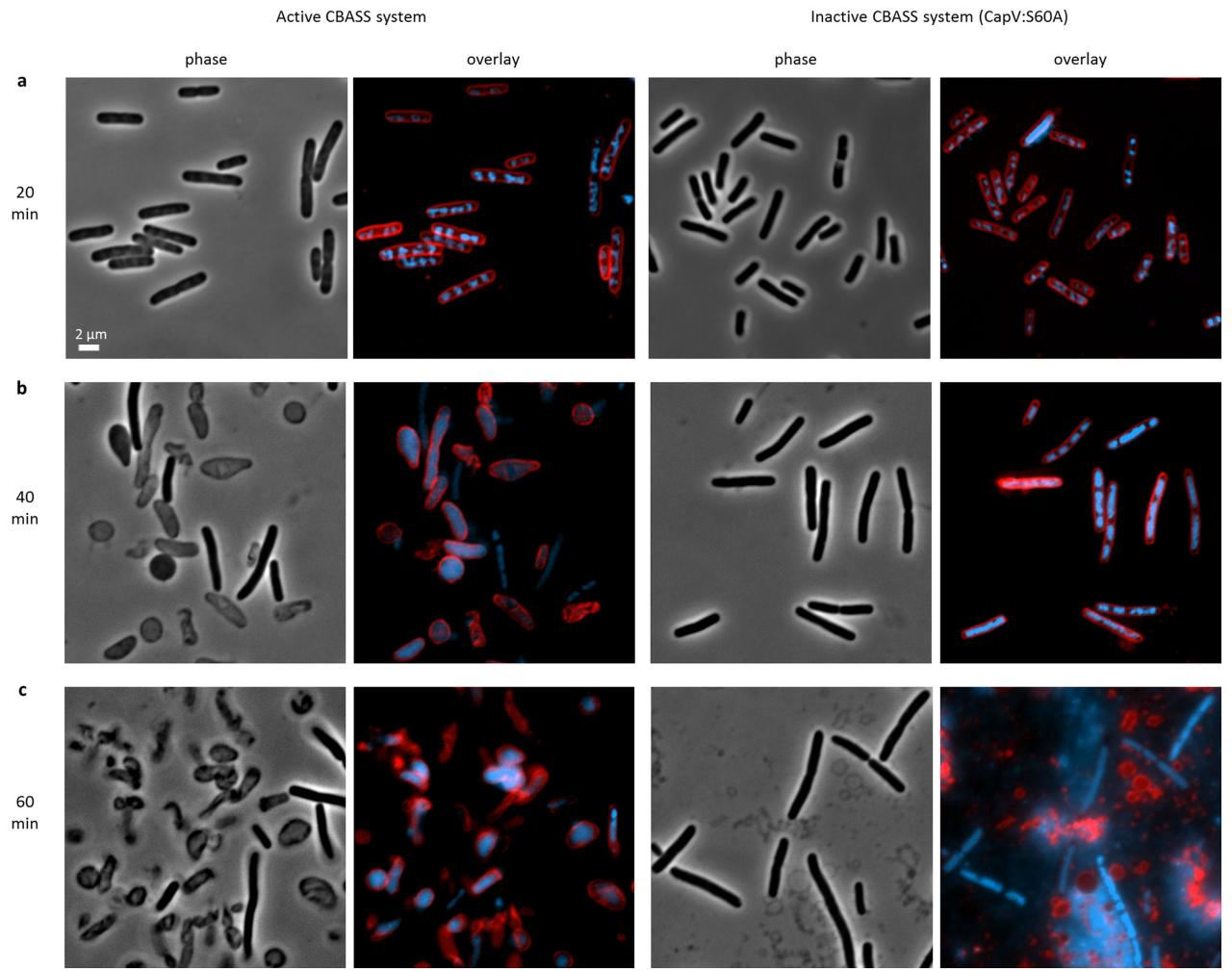
Extended Data Fig. 5 | The bacterial CBASS functions through abortive infection. **a**, Growth curves in liquid culture for CBASS-containing and CBASS-lacking (empty vector) bacteria infected by phage SECphi18 at 25 °C. Bacteria were infected at time = 0 at an MOI of 0.02 or 2. Three independent replicates for each MOI are shown, and each curve shows an individual replicate. **b**, Growth

curves in liquid culture for cells containing a minimal CBASS comprising phospholipase-cGAS (*capV-dncV*) only. Bacteria were infected at time = 0 at an MOI of 2 by phage P1. Three independent replicates for each MOI are shown, and each curve shows an individual replicate.



Extended Data Fig. 6 | Cell sorting of infected cells stained with propidium iodide. Cells containing the CBASS derived from *E. coli* TW11681, and control cells containing an empty vector, were stained with propidium iodide, a fluorescent DNA-binding agent that penetrates cells that have impaired membrane integrity. Cells were infected by phage P1 (MOI of 2) and sorted on the basis of propidium-iodide fluorescence intensity (yaxis); the xaxis represents

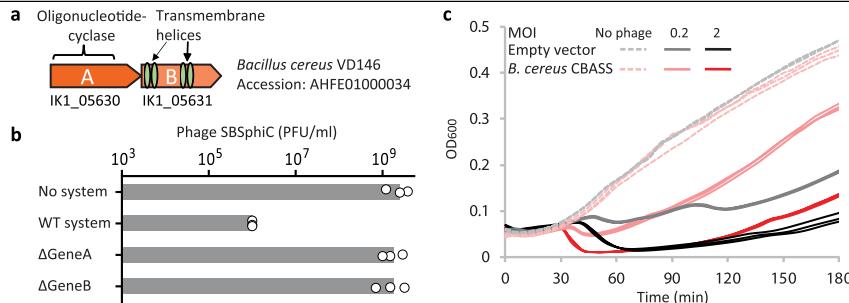
forward scatter. **a**, Uninfected cells that lack the CBASS. **b**, Uninfected cells that contain the CBASS. **c**, Cells that lack the CBASS, 40 min after infection. **d**, Cells that contain the CBASS, 40 min after infection. A large population of cells with high propidium-iodide fluorescence intensity is observed. Data from a representative replicate of two independent replicates are shown.



Extended Data Fig. 7 | Microscopy of infected cells. a–c, Phase contrast and overlay images are shown, of membrane stain (red) and DAPI (blue) images captured at 20 min (a), 40 min (b) and 60 min (c) after infection with phage P1 at an MOI of 2. The two columns on the left show *E. coli* MG1655 cells containing the CBASS derived from *E. coli* TW11681. The two columns on the right show *E. coli* MG1655 cells containing the CBASS derived from *E. coli* with a single point

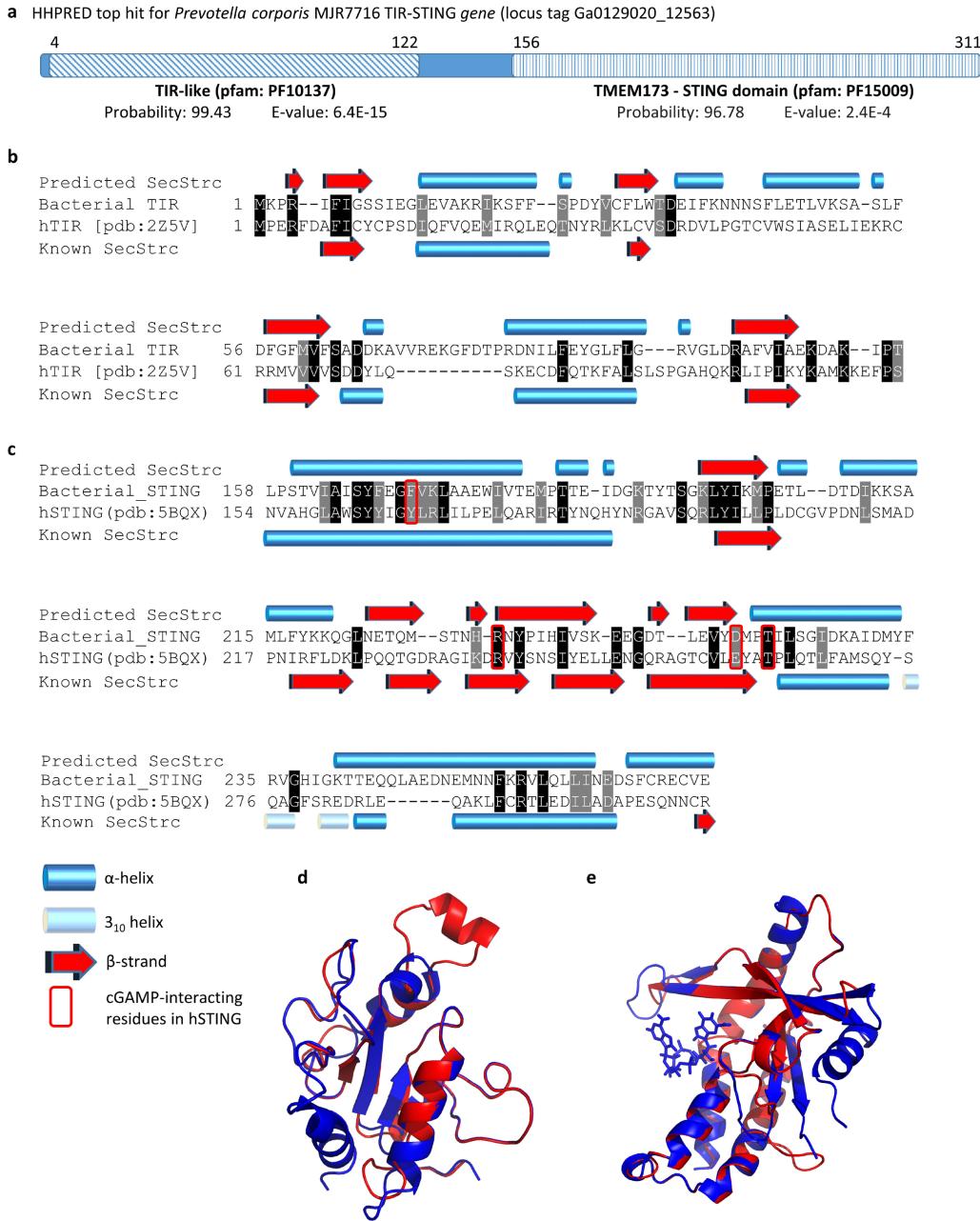
mutation that inactivates the CapV phospholipase (CapV(S60A)). Cell shape is deformed after 40 min in cells containing the CBASS, but not in cells in which the CBASS is mutated. After 60 min, phage-mediated cell lysis is observed in cells in which the CBASS is mutated. Representative images from a single replicate out of two independent replicates are shown.

Article



Extended Data Fig. 8 | A two-gene CBASS protects *Bacillus* against phage infection. **a**, Domain organization of a two-gene operon found in the *B. cereus* VD146 genome. Locus tags of the depicted genes are indicated below each gene. **b**, The two-gene operon from *B. cereus* VD146 was cloned and genetically integrated into *B. subtilis* BEST7003, which naturally lacks this system. The efficiency of plating of phage SBSphiC infecting the CBASS-lacking and CBASS-containing strains, as well as strains in which one of the two genes was deleted, is

shown. Bar graph represents average of three independent replicates, with individual data points overlaid. **c**, Growth curves in liquid culture for *B. subtilis* containing the *B. cereus* two-gene CBASS, or CBASS-lacking *B. subtilis* that contains an empty vector instead, infected by phage SBSphiC. Bacteria were infected at time = 0 at an MOI of 0.2 or 2. Three independent replicates for each MOI are shown, and each curve shows an individual replicate.



Extended Data Fig. 9 | Domain analysis and homology-based structure prediction of a bacterial TIR-STING protein. **a**, Schematics of HHpred²⁹ homology-based search results of the *Prevotella corporis* TIR-STING protein (Supplementary Table 1). **b**, Phyre2³⁵ secondary structure prediction of the TIR domain in the *P. corporis* TIR-STING protein, compared to the solved crystal structure of the human TIR domain protein MyD88 (PDB accession 2Z5V_A). **c**, Phyre2³⁵ secondary structure prediction of the STING domain in the *P. corporis* TIR-STING protein, compared to the solved crystal structure of the human

STING protein (PDB accession 5BQX_A). Black, identical residues; grey, similar residues. Secondary structure prediction for the bacterial protein appears above the alignment; secondary structure of solved human domain appears below the alignment. **d**, Structural alignment of human TIR domain protein MYD88 and the modelled bacterial TIR domain. **e**, Structural alignment of human STING domain and the modelled bacterial STING domain. In **d**, **e**, blue and red represent the structure of the human protein and the model of the bacterial domain structure, respectively.

Corresponding author(s): Rotem Sorek

Last updated by author(s): Sep 1, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

N/A

Data analysis

MMSeqs2; Clustal Omega; HHpred; MAFFT; FastTree; iTOL; Compound Discoverer 3.0; FlowJo v10; Phyre2; ZEN software version 2.0 (Zeiss)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data supporting the findings of this study are available within the manuscript, extended data and supplementary tables. GenBank accessions, locus tags and nucleotide ranges of the CBASS systems appear in the Methods section. Integrated Microbial Genomes (IMG) gene and genome ID number, contig ID, and system and effector classification appear in Supplementary Table 1. Primer sequences of the CBASS systems herein described are available in Supplementary Table 2.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	To generate the phylogenetic tree in Fig. 3a, protein redundancies were removed using the "clusthash" option of MMseqs2 (release 6-f5a1c). Sequences shorter than 200 amino acids were also removed. To further remove outlier sequences, an all-versus-all search was conducted (using the "search" option) and proteins with less than 20 hits were manually examined. Overall, 17 outlier proteins were removed this way.
Replication	Efficiency of plating (EOP) experiments and growth curves of liquid cultures were performed in triplicate across independent experiments. The fluorogenic assay for CapV activity was performed in triplicate except for the 2.5 micro-molar concentration in the calibration curve where data represent two replicates. Three technical replicates were performed to estimate variability in measurement. Mass spectrometry was performed in triplicate for cell lysates collected across independent experiments. Attempts at replication were successful.
Randomization	N/A
Blinding	N/A

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.</i>
-------------------	---

Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access and import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.
Validation	Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	<i>State the source of each cell line used.</i>
Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>
Mycoplasma contamination	<i>Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.</i>
Commonly misidentified lines (See ICLAC register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

Palaeontology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>
Dating methods	<i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i>

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	<i>For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.</i>
Wild animals	<i>Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.</i>
Field-collected samples	<i>For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.</i>
Ethics oversight	<i>Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.</i>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	<i>Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."</i>
Recruitment	<i>Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.</i>
Ethics oversight	<i>Identify the organization(s) that approved the study protocol.</i>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<i>Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.</i>
Study protocol	<i>Note where the full trial protocol can be accessed OR if not available, explain why.</i>
Data collection	<i>Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.</i>

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

ChIP-seq**Data deposition**

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

**Genome browser session
(e.g. [UCSC](#))**

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology**Replicates**

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry**Plots**

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology**Sample preparation**

See methods

Instrument

BIORAD ZE5 Cell Analyzer

Software

FlowJo v10

Cell population abundance

N/A

Gating strategy

N/A

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging**Experimental design****Design type**

Indicate task or resting state; event-related or block design.

Design specifications		<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures		<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>
Acquisition		
Imaging type(s)		<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength		<i>Specify in Tesla</i>
Sequence & imaging parameters		<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition		<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	<input checked="" type="checkbox"/> Used <input type="checkbox"/> Not used	
Preprocessing		
Preprocessing software		<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
Normalization		<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template		<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal		<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring		<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>
Statistical modeling & inference		
Model type and settings		<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested		<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:		<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See Eklund et al. 2016)		<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
Correction		<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>
Models & analysis		
n/a	Involved in the study	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Functional and/or effective connectivity
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Graph analysis
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Multivariate modeling or predictive analysis
Functional and/or effective connectivity		<i>Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).</i>
Graph analysis		<i>Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).</i>
Multivariate modeling and predictive analysis		<i>Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.</i>