Title: Discovery of small alarmone synthetases and their inhibitors as toxin-antitoxin loci

Authors: Steffi Jimmy[1,2], Chayan Kumar Saha[1], Constantine Stavropoulos[1], Abel Garcia-Pino[3], Vasili Hauryliuk[1,2,4,*], Gemma C. Atkinson[1,*]

[1]Department of Molecular Biology, Umeå University, Building 6K, 6L University Hospital Area, 901 87 Umeå, Sweden
[2]Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Building 6K and 6L, University Hospital Area, SE-901 87 Umeå, Sweden
[3]Cellular and Molecular Microbiology, Faculté des Sciences, Université Libre de Bruxelles, 12 rue des Professeurs Jeener et Brachet, 6041 Gosselies, Belgium
[4]University of Tartu, Institute of Technology, 50411 Tartu, Estonia

[*] denotes the corresponding authors

Contact details of corresponding authors:
Vasili Hauryliuk: vasili.hauryliuk@umu.se, +46 907 850 807
Gemma C. Atkinson: gemma.atkinson@umu.se, +46 706070315

Summary
Under stressful conditions, bacterial RelA-SpoT Homologue (RSH) enzymes synthesise the alarmone (p)ppGpp, a nucleotide messenger. (p)ppGpp rewires bacterial transcription and metabolism to cope with stress, and at high concentrations inhibits the process of protein synthesis and bacterial growth to save and redirect resources until conditions improve. Single domain Small Alarmone Synthetases (SAS) are RSH family members that contain the (p)ppGpp synthesis (SYNTH) domain, but lack the hydrolysis (HD) domain, and regulatory C-terminal domains of the long RSHs such as Rel, RelA and SpoT. We have discovered that multiple SAS subfamilies can be encoded in broadly distributed bicistronic operon architectures in bacteria and bacteriophages that are reminiscent of those typically seen in toxin-antitoxin (TA) operons. We have validated five of these SASs as being toxic (toxSASs), and shown that the toxicity can be neutralised by five distinct neighbouring antitoxin genes that act at the protein level.

Introduction

Bacteria encounter a variety of different environment conditions during their life cycle, to which they need to respond and adapt in order to survive. This can include slowing down their growth and redirecting their metabolic resources during nutritional stress, until conditions improve and the growth rate can increase. One of the main signals that bacteria use for signalling stress are the alarmone nucleotides ppGpp and pppGpp, collectively referred to as (p)ppGpp(1). At high concentrations (p)ppGpp is a potent inhibitor of bacterial growth(2), targeting transcription(3), translation(4) and ribosome assembly(5). (p)ppGpp is produced and degraded by proteins of the RelA/SpoT homologue (RSH) superfamily, named after the two *Escherichia coli* representatives – multi-domain 'long' RSH factors RelA and SpoT(6). In addition to 'long' RSHs, bacteria can encode single-domain RSHs: Small Alarmone Synthetases (SAS)(7-10) and Small Alarmone Hydrolases (SAH)(6,11).

It is currently unknown why some bacteria carry multiple SASs and SAHs, which can belong to many different subfamilies. Conservation of gene order through evolution can reveal potentially interacting proteins and shed light of the cellular role of proteins(12). Therefore, we developed a computational tool – FlaGs, standing for Flanking Genes – for analysing the conservation of genomic neighbourhoods, and applied it to our updated database of RSH sequences classified into subfamilies. Surprisingly, we find that some subfamilies of SAS can be encoded in apparently bicistronic, often overlapping, conserved gene architectures that are reminiscent of toxin-antitoxin (TA) loci(13-16). The potential for SAS toxicity is supported by the observation that when (p)ppGpp is over-produced – for example, if synthesis by RelA is not balanced by hydrolysis by SpoT – the alarmone becomes toxic and inhibits growth(17).

The first direct evidence that RSH toxicity *per se* might be a *bona fide* function of some SASs was provided by Dedrick and colleagues(18). They showed that gp29, a SAS encoded by the mycobacterial Cluster N bacteriophage Phrann is exceedingly toxic to *M. smegmatis.* This toxicity is countered by co-expression of its neighbouring gene (gp30) – a proposed inhibitor of the SAS. Neither the molecular mechanism of gp29-mediated toxicity nor its neutralisation by gp30 are known. The gp29-mediated abrogation of growth is proposed to be a defence mechanism against co-infection by other bacteriophages, such as Tweety and Gaia(18).

The regulatory interplay between gp29 and gp30 is typical of that seen in toxin-antitoxin (TA) systems. When expressed, the toxin protein abolishes bacterial growth – and its toxicity can be efficiently countered by the protein or RNA antitoxin. Known toxins can act in a number of ways, most commonly targeting translation by cutting or modifying the ribosome(19), translational factors(20), tRNAs(21) or mRNAs(22). Similarly, antitoxins counteract the toxins through different mechanisms: through base-pairing of the antitoxin RNA with the toxin mRNA (type I TA systems(23)), direct protein-protein inhibition (type II(24)), inhibition of the toxin by the antitoxin RNA (type III(25)), or by indirect nullification of the toxicity (type IV(26)).

In this study we have uncovered the evolutionary diversity of SAS-based toxin (toxSAS) TA systems using sensitive *in silico* sequence searching and gene neighbourhood analysis. We have experimentally validated five SAS subfamilies as *bona fide* TA pairs and demonstrated through mutagenesis that the toxicity of SASs is strictly dependent on a functional (p)ppGpp synthetase active site. All of the discovered antitoxin classes are proteins, and while four classes are strictly specific in counteracting only the cognate toxSAS, one class can universally neutralize all of the toxSASs. This antitoxin encodes a (p)ppGpp degrading enzyme – SAH – that is, it acts as a type IV antitoxin degrading the molecular product of toxSAS synthetic activity.

Results

Updated RSH phylogeny across the tree of life

Our previous evolutionary analysis of the RSH protein family applied high-throughput sensitive sequence searching of 1072 genomes from across the tree of life(6). Since the number of available genomes has grown dramatically in the last decade(27), we revisited the evolution of RSHs, taking advantage of our new computational tool, FlaGs to analyse the conservation of gene neighbourhoods that might be indicative of functional associations(28). FlaGs clusters neighbourhood-encoded proteins into homologous groups and outputs a graphical visualisation of the gene neighbourhood and its conservation along with a phylogenetic tree annotated with flanking gene conservation. We identified and classified all the RSHs in 24072 genomes from across the tree of life using our previous Hidden Markov Model (HMM)-based method. We then carried out phylogenetic analysis to identify new subfamilies, generated new HMMs and updated the classification in our database (Supplementary **Tables S1** and **S2**). We have identified 30 subfamilies of SASs, 11 subfamilies of SAHs, 13 subfamilies of Long RSHs (**Figure 1**).

Putative toxSAS TA modules are widespread in Actinobacteria, Firmicutes and Proteobacteria

We ran FlaGs on each of all the subfamilies and discovered that Small Alarmone Synthetase (SAS) genes can be frequently found in conserved bicistronic (sometimes overlapping) loci that are characteristic of toxin-antitoxin (TA) loci. Five SAS subfamilies displaying particularly well conserved TA-like bicistronic arrangements: FaRel, FaRel2, PhRel, PhRel2 and CapRel (**Figure 2**, **Supplementary File S1, Supplementary Table S3**) were selected for further investigation. Among bacteria, PhRel (standing for Phage Rel, the group to which Gp29(18) belongs) and FaRel are found in multiple species of Firmicutes and Actinobacteria (hence the "Fa" prefix), along with representatives of various Proteobacteria; FaRel2 is found in multiple Actinobacteria, and Firmicutes, while PhRel2 is found in firmicutes in addition to bacillus phages. CapRel as a subfamily can be found in a wide diversity of bacteria (including Cyanobacteria, Actinobacteria and Proteobacteria, hence the "Cap" prefix. Each TA-like subfamily has a different putative antitoxin, with the exception of PhRel and CapRel, which have a homologous putative antitoxin (**Figure 2**). PhRel and CapRel are sister groups in the RSH phylogeny with medium support (81% MLB RAxML, 96% UFB IQ-TREE **Figure 1, Supplementary Text S1**), suggesting the TA arrangement has been conserved during the diversification of these groups from a common ancestor.

The potential antitoxins are named with an 'At' prefix to the SAS name. AtFaRel is a predicted SAH, in the PbcSpo family (**Figure 1**), and AtPhRel2 is a GepA (Genetic Element protein A) family homologue. GepA proteins, which carry the DUF4065 domain have previously been associated with TA loci(29-31), and are related to the proteolysis-promoting SocA antitoxin of the SocB toxin(32). The other potential antitoxins (AtCapRel, AtFaRel2 and AtPhRel) are small proteins with no detectable homology to proteins of known function.

Bicistronic toxSAS operons encode bona fide *type II and type IV TA pairs*

We tested whether SASs encoded in conserved bicistronic regions act as *bona fide* TA systems using a toxicity neutralisation assay in *Escherichia coli* BW25113 strain(33). Putative toxSAS and antitoxin genes were expressed under the control of arabinose- and IPTG-inducible promoters, respectively(33). We have verified five toxSASs as toxic components of *bona fide* TA pairs: *Bacillus subtilis* la1a PhRel2 (**Figure 3A**), *Coprobacillus* sp. D7 FaRel2 (**Figure 3B**), *Mycobacterium* phage Phrann PhRel (gp29) (**Figure 3C**) and *Cellulomonas marina* FaRel (**Figure 3D**). Importantly, co-expression of putative antitoxins restored bacterial viability in all of the cases. Despite the well-conserved bicistronic organisation, *Mycobacterium tuberculosis* AB308 CapRel (**Figure 3E**) initially displayed no detectable toxicity. Thus, we added a strong Shine-Dalgarno motif (5'-AGGAGG-3') to increase the translation initiation efficiency in order to drive up the expression levels. In the case of *Mycobacterium sp.* AB308 CapRel, the protein became toxic. Importantly, this toxicity is readily

counteracted by the antitoxin AtcapRel (**Figure 3E**). Mycobacterium phage Squirty PhRel(18) did not display significant toxicity even when the expression was driven with a strong Shine-Dalgarno sequence (**Supplementary Figure S1A**). Additionally, we tested well-studied bacterial SASs that are not encoded in bicistronic arrangements (*Staphylococcus aureus* RelP(34,35) and *Enterococcus faecalis* RelQ(36,37)). We detected no toxicity, even with strong Shine-Dalgarno motifs (**Supplementary Figure S1B**).

The validated toxSAS toxins differ in the strength of the toxic effect in our system (**Figure 3A-E**): i) FaRel2 and PhRel2 are exceedingly potent and no bacterial growth is detected upon expression of these toxins from the original pBAD33 vector, ii) FaRel and PhRel are significantly weaker and small colonies are readily visible and iii) CapRel is weaker still, with toxicity requiring the introduction of a strong Shine-Dalgarno sequence in the pBAD33 vector. We have validated this toxicity gradation by following the bacterial growth in liquid culture. In good agreement with the plating experiments, induction of FaRel2 and PhRel2 expression completely abrogated bacterial growth in minimal MOPS medium, while FaRel, PhRel and CapRel decrease the growth rate 4.2, 7.5 and 18.4 times, respectively (**Figure 3**).

Next we tested whether enzymatic activity is responsible for the toxicity of toxSASs. To do so, we substituted a conserved tyrosine in the so-called G-loop for alanine. This residue is critical for binding the nucleotide substrate and is highly conserved in (p)ppGpp synthetases(38). All of the tested mutants – PhRel2 Y173A (**Figure 4A)**, FaRel2 Y128A, PhRel Y143A and FaRel Y175A – were non-toxic (**Supplementary Figure S2**). Therefore, we concluded that production of a toxic alarmone is, indeed, the universal causative agent of growth inhibition by toxSASs.

We then investigated whether toxSAS antitoxins inhibit toxSASs on the level of RNA (as in type I and III TA systems) or protein (as in type II and IV TA systems). The former scenario is theoretically possible, since, as we have shown earlier, *E. faecalis* SAS RelQ binds single-strand RNA and is inhibited in a sequence-specific manner(36). To discriminate between the two alternatives, we mutated the start codon of the *aTphRel2, aTfaRel2* and *aTphRel* antitoxin ORFs to a stop codon, TAA. Since all of the mutants fail to protect from the cognate toxSAS (**Figure 4B** and **Supplementary Figure S3**), we concluded that they all act as proteins, that is, are type II or IV antitoxins.

C. marina ATfaRel efficiently degrades (p)ppGpp and cross-inhibits all identified toxSAS SASs

The antitoxin ATfaRel is a member of the PbcSpo subfamily of SAH hydrolases (**Figure 1A**). This suggests it acts via degradation of the alarmone nucleotide produced by the toxSAS (and therefore as a type IV TA system that does not require direct physical interaction of the TA pair). Therefore, we hypothesised that ATfaRel is able to mitigate the toxicity of all of the identified toxSAS classes through alarmone degradation. This is, indeed, the case (**Figure 5A**). To test if the hydrolysis activity is strictly necessary for antitoxin function, we generated a point mutant of ATfaRel (D54A). The corresponding mutation of the active site abolishes (p)ppGpp hydrolysis by Rel from *Streptococcus dysgalactiae* subsp. *equisimilis*, Rel_{Seq}(39). As expected, the D54A mutant is unable to counteract the toxicity from FaRel (**Figure 5B**). We used metabolic labelling with ^{32}P-orthophosphoric acid combined with TLC separation and autoradiography to assess the accumulation and degradation of nucleotide alarmones upon expression of the *C. marina* FaRel toxSAS and ATfaRel SAH. The expression of FaRel results in an accumulation of ^{32}P-ppGpp and this accumulation is counteracted by wild type – but not D54A substituted – ATfaRel (**Figure 5C**). The location of the SAS immediately downstream of the SAH raises the question of whether this gene pair has evolved from fission of a long RSH. However, if this was the case, we would see faRel and ATfaRel in the long RSH part of the phylogeny, which we do not see (**Figure 1**).

Class II antitoxins protect only from cognate toxSAS toxins

The gp29-mediated abrogation of growth is employed by the Phrann phage as a defence mechanism against super-infection by other phages, such as Tweety and Gaia(18). This raises the question of cross-inhibition between toxSAS TA systems: do all of the identified antitoxins inhibit all of the toxSASs (similarly to how the type IV antitoxin SAH ATfaRel protects from all of the tested toxSASs, see **Figure 5A** and **Table 1**) or is the inhibition specific to toxSAS subfamilies TAs? Therefore, we exhaustively tested pairwise combinations of all of the toxSASs with all of the antitoxins (**Table 1** and **Supplementary Figure S4**). ATphRel2, ATfaRel2, ATphRel, and ATcapRel antitoxins could not counteract their non-cognate toxSAS (**Table 1**, **Supplementary Figure S4**), demonstrating that different classes provide specific discrimination of self from non-self.

Table 1 | Cross-talk amongst identified toxSAS and their antitoxins as well as standalone phage-encoded SAHs. Toxicity neutralisation assays are presented in (Figure 5A and **Supplementary Figures S4 and S5**). Plus (+) and minus (-) symbols indicate ability and inability of the antitoxin to neutralise toxicity, respectively.

	Mycobacterium sp. AB308 ATcapRel	*B. subtilis* la1a ATphRel2	*Coprobacillus* sp. D7 ATfaRel2	*Mycobacterium* phage Phrann gp29 ATphRel	*C. marina* ATfaRel SAH	*Salmonella* phage PVP-SE1 SAH (PbcSpo)	*Salmonella* phage SSU5 SAH (PaSpo)
CapRel	+	-	-	-	+	+	+
PhRel2	-	+	-	-	+	+	+
FaRel2	-	-	+	-	+	+	+
PhRel	-	-	-	+	+	+	+
FaRel	-	-	-	-	+	+	+

Numerous SASs and SAHs are encoded in prophage-derived regions of bacterial genomes

Our initial search has identified 13 SASs in bacteriophage genomes, five of which we have confirmed as toxSASs (**Figures 2 and 3**). However, this is likely to be an underestimate for two reasons. Firstly, the currently sequenced phage genomes are a small sample of their entire diversity(40), and secondly, as prophages reside in bacterial genomes, their genes may not be identified as phage in origin. To detect SAS genes that may be phage in origin but reside in bacterial genomes, we used the tool PHASTER (PHAge Search Tool Enhanced Release(41)), taking a region of DNA equivalent to four upstream and four downstream genes around each SAS and SAH gene (one representative strain per bacterial species). In addition to the already identified phage-encoded CapRel, PhRel and PhRel2, we find 18 prophage regions around representatives in groups belonging to CapRel, FpRel, FpRel2, FunRel and RelP (**Supplementary Table S4**). It is notable that of RelP and RelQ (the two most broadly distributed SASs), RelP but not RelQ can be phage-associated. An evolutionary history that includes transduction may be part of the reason why the operon structure of RelP is less well conserved across genera compared with RelQ (**Supplementary File S1)**. SAHs are found in many more prophages and prophage-like regions than SASs (90 versus 63 instances, **Supplementary Table S4**). We have tested SAHs encoded by Salmonella phages PVP-SE1(42) (PbcSpo subfamily) and SSU5(43) (PaSpo subfamily) in toxicity neutralisation assays against validated toxSASs. Like the *C. marina* SAH ATfaRel, both of these stand-alone phage-encoded SAHs efficiently mitigate the toxicity of all the tested toxSASs (**Table 1** and **Supplementary Figure S5**).

Discussion

Using our tool FlaGs we have made the striking and surprising discovery that multiple SAS subfamilies can be encoded in TA-like bicistronic architectures. Through subsequent validation, we have found that the organisation of SAS genes into bicistronic architectures along with a downstream – or in one case upstream – gene is a strong indicator of toxicity. Identification of bicistronic architectures has been used as a starting point for prediction of TAs previously (44,45). However, these studies have focussed on species that do not encode toxSASs, and therefore these TA systems were not detected. By being associated with novel antitoxins, toxSASs have also avoided identification in "guilt by

association" analysis of thousands of genomes (46). This long-term obscurity is despite toxSAS-containing subfamilies being broadly distributed, present in 239 genera of 15 phyla of bacteria. Thus it is likely that there are other previously unknown TA systems to be found that are identifiable through searching for conservation of gene neighbourhoods across disparate lineages, as we have done with FlaGs.

The RSH protein family is widespread; most likely being present in the last common ancestor of bacteria. Thus for billions of years, these proteins have been used by bacteria to regulate their growth rate in response to their environment by synthesising and hydrolysing nucleotide alarmones. Paradoxically, the very ability of an alarmone to downregulate growth for continued survival is also what gives it toxic potential. We have identified 30 subfamilies of SAS, five of which we have validated as containing toxins, and two of which we have validated as non-toxic (RelP and RelQ). It is likely that SASs exist on a continuum in terms of toxicity, with an antitoxin only being required at a certain level of toxicity. This is supported by the observation that not all toxSASs have the same level of toxicity, with one (*Mycobacterium tuberculosis* AB308 CapRel) requiring a strong Shine-Dalgarno in order to observe any toxicity in our system. Of our five validated toxSASs, there are four different homologous groups of antitoxins. This – and the lack of a multi-subfamily toxSAS-specific clade in phylogenetic analysis – suggests toxic SASs could have evolved independently multiple times from non-toxic SASs. In the evolution of ToxSAS-antiToxSAS modules from non-toxic SAS, it is unlikely that the toxic component evolved before the regulatory antitoxin, as this would be detrimental to fitness. Rather, it is more likely that a SAS became regulated by a neighbouring gene, which relaxed enzymatic constraints on the SAS, allowing it to evolve increased alarmone synthesis rates.

Thus, we suggest that it may be useful to consider Type II TAs in the broadest sense as two-gene regulatory systems of enzyme activity rather than the toxicity always being the main function in itself; especially as toxicity is usually defined using overexpression in a heterologous system(47). Nevertheless, some toxSASs do seem to have a specific role that depends on the toxicity: inhibition of superinfection in the case of the phage PhRel2-ATPhRel2 (Gp29-Gp30) toxSAS TA pair(18). In this system, PhRel encoded in a prophage protects Mycobacteria from an infection by a second phage. Phage infection has previously been linked to alarmone accumulation and stringent response in bacteria (48-50). Presumably this is an example of a so called abortive infection mechanism(51), where infected hosts are metabolically restricted, but the larger population is protected. A corollary of alarmone-mediated phage inhibition is that incoming phages could bypass this system by encoding alarmone hydrolases. Indeed, we have found a variety of different SAHs in different phage genomes, and prophage-like regions of bacterial genomes suggesting there could be cross-talk between ToxSASs and SAHs during infection and superinfection.

Materials and Methods

Computational analysis of RSH sequences

Predicted proteomes from 24072 genomes were downloaded from NCBI, selecting one representative genome per species for Archaea, Bacteria, Eukaryotes and all Viruses. The sequences from our previous RSH database (6) were extracted on a subfamily basis, aligned with MAFFT v7.313(52) and hidden Markov models (HMMs) were made with HMMer v3.1b2(53). All genomes were scanned with the HMMs to identify RSH family members and classify them into subfamilies with E value cut-off thresholds that were previously determined(6): E^{-4} and E^{-5} for SYNTH and HD domains, respectively. HMMs of the HD and SYNTH domains were used to determine the (p)ppGpp synthesising and hydrolysing domains present in the identified sequences. Sequences, taxonomy of the source organism, domain composition and subfamily memberships were stored in a MySQL database. To update the classifications, phylogenetic analysis was carried out of representatives one per genus based on taxonomy. Three data sets of sequences, Long RSHs, SASs and SAHs were extracted and aligned as above. Phylogenetic analysis was carried out with FastTree v2.1.10(54) after

removing positions containing more than 50% gaps, and any extremely divergent proteins that could not be confidently aligned. The three resulting trees and alignments were examined by eye with FigTree v1.4.2 and Aliview v1.2.0(55) to identify groups that are appear to be distinct, that is are comprised of mostly orthologues, have a distinct domain structure, and, ideally, have strong support for monophyly. Eight of our subfamilies are paraphyletic in that they contain one or monophyletic groups nested within their diversity: MixSpo, AaRel, CapRel, FpRel, FpRel, MixRel, PRel3, and Rel. To further resolve relationships among subfamilies, trees were then made with RaxML v8.2.10(56) and IQTree v1.6.6(57) on the Cipres Science Gateway v 3.3 portal (58), excluding sequences that could not be assigned to subfamilies in the FastTree tree. For Maximum Likelihood phylogenetic analyses, we used the LG model of substitution, which is the best-fit model for our dataset, as predicted by IQTree. RAxML was run with 100 bootstrap replicates to give a value (maximum likelihood boostrap, MLB percentage) for how much of the input alignment supports a particular branch in the tree topology. In the case of IQ-TREE, the ultrafast bootstrapping (UFB) approximation method was used to compute support values for branches out of 1000 replicates. Trees from RAxML and IQTree were visualized with FigTree and subfamilies were inspected for whether they contain mostly orthologues with at least moderate (>60%) bootstrap support. Overall, we could classify sequences into 13 subfamilies of LongRSHs, 30 subfamilies of SASs and 11 subfamilies of SAHs. The sequences of each subfamily were aligned and used to make HMMs, as above. The sequences in the MySQL database were re-scanned with the updated subfamily HMMs and the database was updated, as reproduced here as an Excel file (**Supplementary Table S1**).

To make the representative tree of **Figure 1**, we selected taxa from the RSH database to sample broadly across the tree of life, and cover all subfamilies of RSHs. Since we have many protein sequences in our classified subfamily, we made a python script to select a set of 15 divergent representatives (exceptions are 145 representatives for Rel, 80 representatives for RelA and SpoT) based on taxonomy for each subfamily. To find the desired number of representatives, the script randomly makes selection of unique representative from each taxonomic level (eg, phylum, class, order genus, species) and calculates the total number of unique representatives on each taxonomic level. So then if the total number of unique representatives are sufficient enough to make the set of representatives, the script selects desired number of unique representatives from that taxonomic level. For example, to make a list of 15 divergent representatives for FaRel subfamily, if we found total 20 representatives from phylum then we randomly selected 15 out of 20. And if found less than 15 then we just take as much as we get from phylum and rest we took from the next level (eg, class or order) using the same criteria so that selection are divergent across the tree of life. The ppGpp hydrolase (HD) domain-containing dataset and the ppGpp synthetase (SYNTH) domain-containing dataset were separately aligned with MAFFT with the L-ins-i strategy with >50% gaps were removed. After alignment curation, our HD domain-containing alignment contained 698 amino acid positions from 519 sequences, and the SYNTH domain-containing alignment contained 699 amino acid positions from 722 sequences (**Supplementary Text S1**). The two domain alignments were used for Maximum Likelihood phylogenetic analysis using RaxML and IQTree were made as above.

To detect if bacterial SAS or SAH genes are located in bacteriophage-like sequence regions, we used the PHASTER URLAPI(41). To create the input nucleotide data sets, we made a pipeline that takes the nucleotide sequence containing the four up and downstream genes around each SAS or SAH genes that are present in our RSH Database.

Construction of plasmids

All bacterial strains and plasmids used in the study are listed in **Supplementary Table S5**. Oligonucleotides and synthetic genes were synthesized by Eurofins. Toxin ORFs were amplified using primers containing SacI and HindIII restriction sites and cloned in pBAD33 vector. To make the constructs with a strong Shine-Dalgarno motif, the 5'-AGGAGG-3' sequence was incorporated into the pBAD33 vector. The full start codon context including the Shine-Dalgarno motif and

intervening sequence was therefore 5'-AGGAGGAATTAAATG-3'. Antitoxin ORFs were amplified using primers containing EcoRI and HindIII restriction sites and cloned in a pKK223-3 vector. Ligation mixes were transformed by heat-shock in *E. coli* DH5α. PCR amplifications were carried out using Phusion polymerase, purchased from ThermoScientific along with restriction enzymes and T4 ligase. Point mutations were introduced using QuikChangeKit (Agilent). All final constructs were re-sequenced by Eurofins.

Toxicity neutralisation assays

Toxicity-neutralisation assays were performed on LB medium (Lennox) plates (VWR). *E. coli* BW25113 strains transformed with pBAD33 (encoding toxins) and pKK223-3 (encoding antitoxins) were grown in liquid LB medium (BD) supplemented with 100 µg/ml carbenicillin (AppliChem) and 20 µg/ml chloramphenicol (AppliChem) as well as 1% glucose (repression conditions). Serial ten-fold dilutions were spotted (5 µl per spot) on solid LB plates containing carbenicillin and chloramphenicol in addition to either 1% glucose (repressive conditions), or 0.2% arabinose combined with 1mM IPTG (induction conditions). Plates were scored after an overnight incubation at 37°C.

Growth assays

Growth assays were performed in liquid MOPS minimal medium (1x MOPS mixture (AppliChem), 0.132 M K_2HPO_4 (VWR Lifesciences), 1mg/ml thiamine (Sigma), 0.1% casamino acids (VWR Lifesciences) and the carbon source – either 0.5% glycerol (VWR Chemicals) or 1% glucose). The media was supplemented with carbenicillin and chloramphenicol. Overnight cultures were grown in MOPS medium supplemented with 1% glucose at 37°C. The cultures were diluted to a final OD_{600} of 0.01 in MOPS medium supplemented with 0.5% glycerol, 0.2% arabinose and 1mM IPTG. Growth was then monitored using a Bioscreen C Analyzer (Oy Growth Curves Ab Ltd) at 37°C for 10 hours.

Measurement of cellular (p)ppGpp levels by thin layer chromatography (TLC)

Overnight cultures were pre-grown at 37°C in liquid MOPS medium supplemented with carbenicillin, chloramphenicol and 1% glucose, diluted to a final OD_{600} of 0.05 in MOPS medium supplemented with 0.5% glycerol and antibiotics. The cultures were grown at 37°C until an OD_{600} of 0.5, at which point they were re-diluted to an OD_{600} of 0.05, and 500 µl were transferred to 2 ml Eppendorf tubes and spiked with 2.5 µCi of ^{32}P-orthophosphoric acid (Perkin Elmer). The cultures were grown for two generations (OD_{600} 0.2) and expression of the toxins and antitoxins was induced by addition of 0.2% arabinose and 1 mM IPTG (final concentrations), respectively. At 0, 5, 15 and 30 minutes post-induction, 50 µl samples were transferred to 1.5 mL Eppendorf tubes containing 10 µl of 2 M formic acid and pelleted for 2 minutes at 14,000 rpm 4°C. 10 µl of the resultant supernatant was spotted on PEI Cellulose TLC plates (Merck). The nucleotides were resolved in 1.5 M KH_2PO_4, pH 3.4 (VWR Chemicals). Plates were dried and imaged on Phosphorimager Typhoon FLA 9500 (GE Healthcare).

Acknowledgments

We are grateful to the Protein Expertise Platform (PEP) Umeå University and Mikael Lindberg for constructing the plasmids used in this work and to Anaïs Poirier for help with experiments. This work was supported by the funds from European Regional Development Fund through the Centre of Excellence for Molecular Cell Technology (TT and VH); Estonian Science Foundation grants (PUT37 to VH, IUT2-22 to TT); the Molecular Infection Medicine Sweden (MIMS) (VH); Swedish Research council (grant 2017-03783 to VH, 2015-04746 to GCA); Ragnar Söderberg foundation (VH)

Author contributions

GCA and VH conceived the study, coordinated the study, and drafted the manuscript with input from all other authors. GCA and VH designed bioinformatics analyses and experiments and analysed the data, with input from AGP. SJ and CS performed experiments, and CKS performed bioinformatics analysis. All authors have read and approved the manuscript as submitted.

Figures and legends

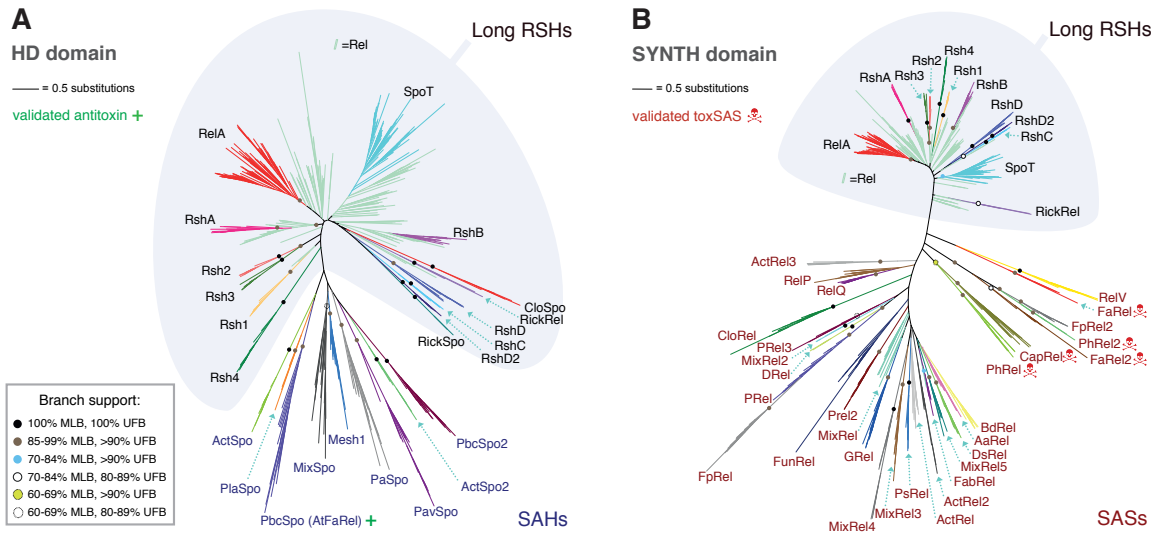

Figure 1 | Maximum likelihood phylogenies of the (p)ppGpp hydrolase (A) and synthetase (B) domains. Trees were generated from RaxML and IQ-TREE analyses of alignments of representatives across the RSHs family with (**A**) the (p)ppGpp hydrolase (HD) domain-containing dataset (698 amino acid positions, 519 sequences), and (**B**) the ppGpp synthetase (SYNTH) domain-containing dataset (699 amino acid positions, 722 sequences). Shading behind the branches indicates the boundary between multi-domain type (long) RSHs and single domain (small) RSHs. The long RSH groups also contain members that seem to have secondarily lost domains to become single domain (members of the RickSpo and RickRel groups). The inset box shows the legend for subfamily and intersubfamily support; support values within subfamilies and that are less that 60% MLB are not shown. Branch length is proportional to the number of substitutions per site (see scale bar). The red skull and crossbones icon indicates those subfamilies of SASs that we have confirmed with toxicity neutralisation assays to contain toxSASs. The SAH group PbcSpo that we have found contains an antitoxin is indicated with a green plus sign. Alignments used for phylogenetic analysis, and trees will all branch support values are available in **Supplementary Text 1.**

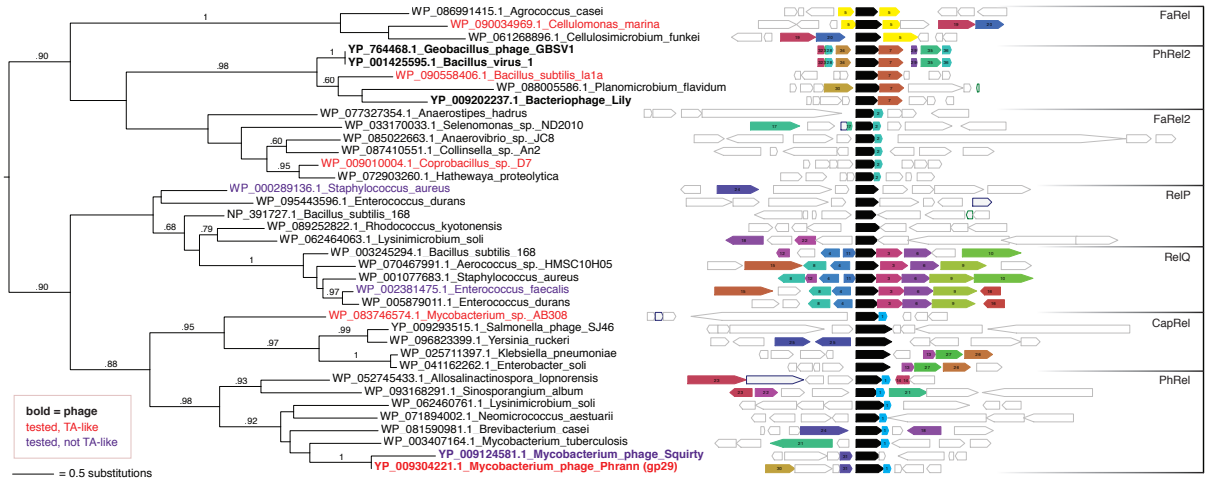

Figure 2 | Conservation of neighbourhood around SAS genes. Genes that encode proteins belonging to a homologous cluster in more than one genomic neighbourhood are coloured and numbered (see **Supplementary Table S3** for the identity of clusters with flanking gene accession numbers). The SAS gene is shown in black, and non-conserved genes are uncoloured. Validated TAs are boxed in red, and have red taxon names. SASs that we have tested and are non-toxic have purple taxon names. Bacteriophage names are in bold.

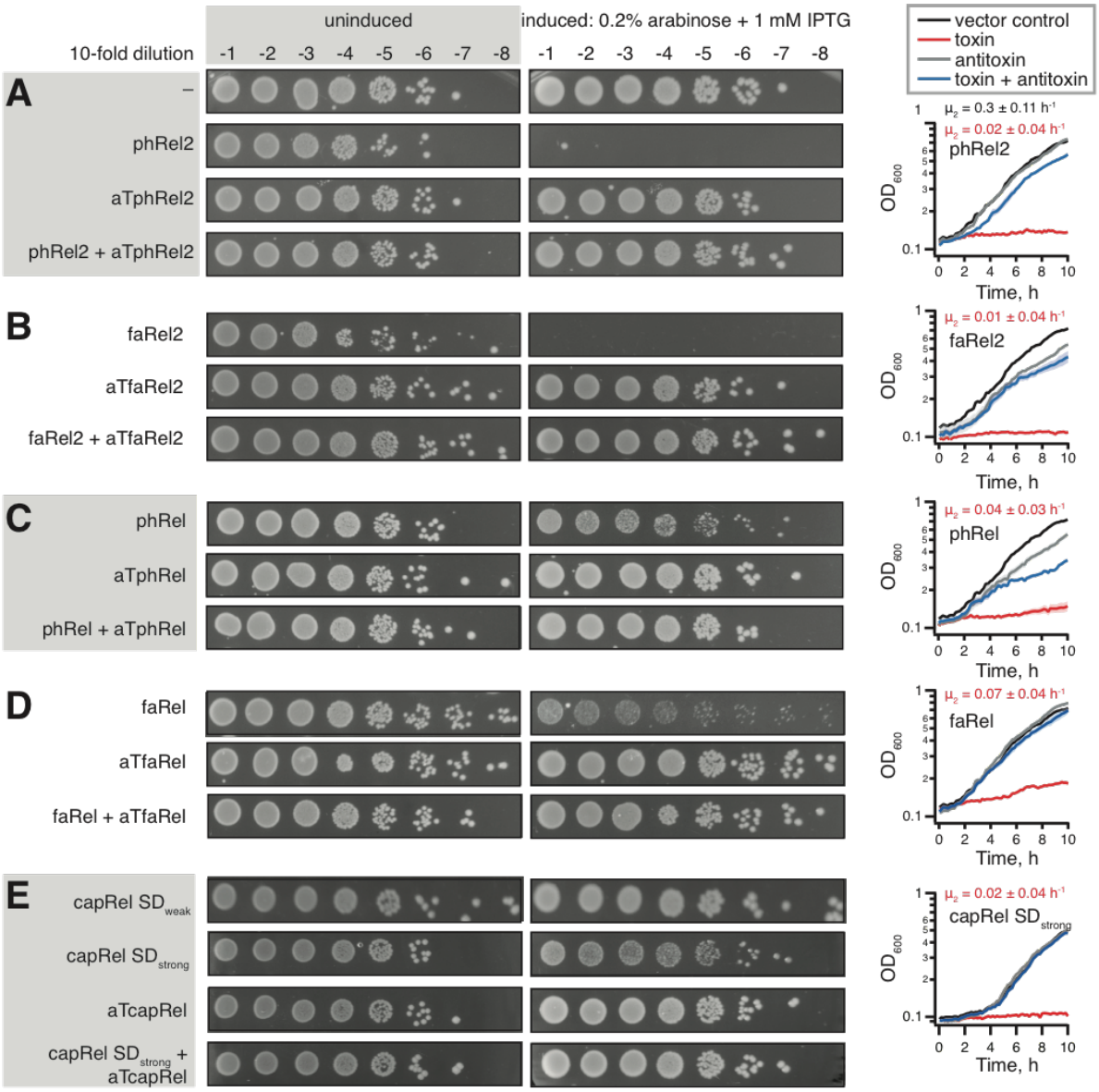

Figure 3 | Bicistronic toxSAS operons encode *bona fide* TA systems. Representatives of groups of bicistronically encoded toxSAS validated as TA-pairs: (**A**) *Bacillus subtilis* la1a PhRel2:AtPhRel2, (**B**) *Coprobacillus* sp. D7 FaRel2:AtFaRel2, (**C**) *Mycobacterium* phage Phrann PhRel:AtPhRel (gp29:gp30) (**D**) *Cellulomonas marina* FaRel:AtFaRel, and (**E**) *Mycobacterium* sp. AB308 CapRel:AtCapRel. To perform the toxicity neutralisation assays on LB plates, overnight cultures of *E. coli* strains transformed with pBAD33 and pKK223-3 vectors or derivatives expressing putative toxSAS toxins and antitoxins, correspondingly, were serially diluted from 10^1 to 10^8-fold and spotted on LB medium supplemented with appropriate antibiotics as well as either 1% glucose (repression conditions, left) or 0.2% arabinose and 1 mM IPTG (induction conditions, right). To assay the toxicity in liquid media, bacteria were grown at 37°C in MOPS minimal media supplemented with 0.5% glycerol, 0.2% arabinose and 1mM IPTG (induction conditions). The growth curves represent the geometric mean of three biological replicates and shading represents the standard error. μ2 signifies the growth rate in the presence of toxin and antitoxin (see inset box).

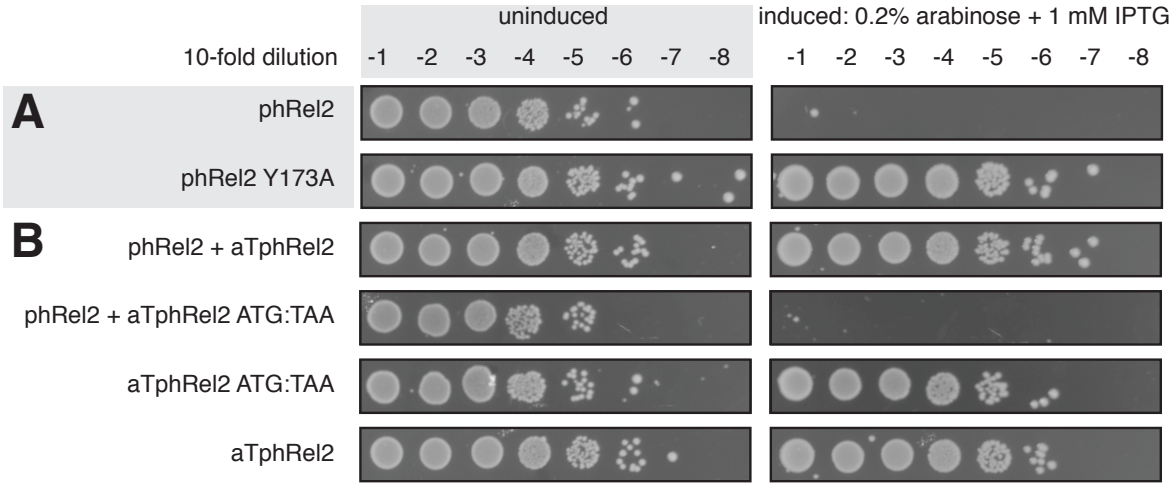

Figure 4 | Active site mutations abrogate toxicity of toxSASs, and toxSAS antitoxins work as proteins, not RNA. (**A**) Active site mutation Y173A renders PhRel2 toxSAS non-toxic. Analogous experiments with all other identified toxSAS support the essentiality of the enzyme function for toxicity (**Supplementary Figure S2**). (**B**) Mutation of the start codon to stop renders the *aTphRel2* antitoxin ORF unable to protect from the PhRel2 toxin. Equivalent experiments of other toxSASs are presented in **Supplementary Figure S3**.

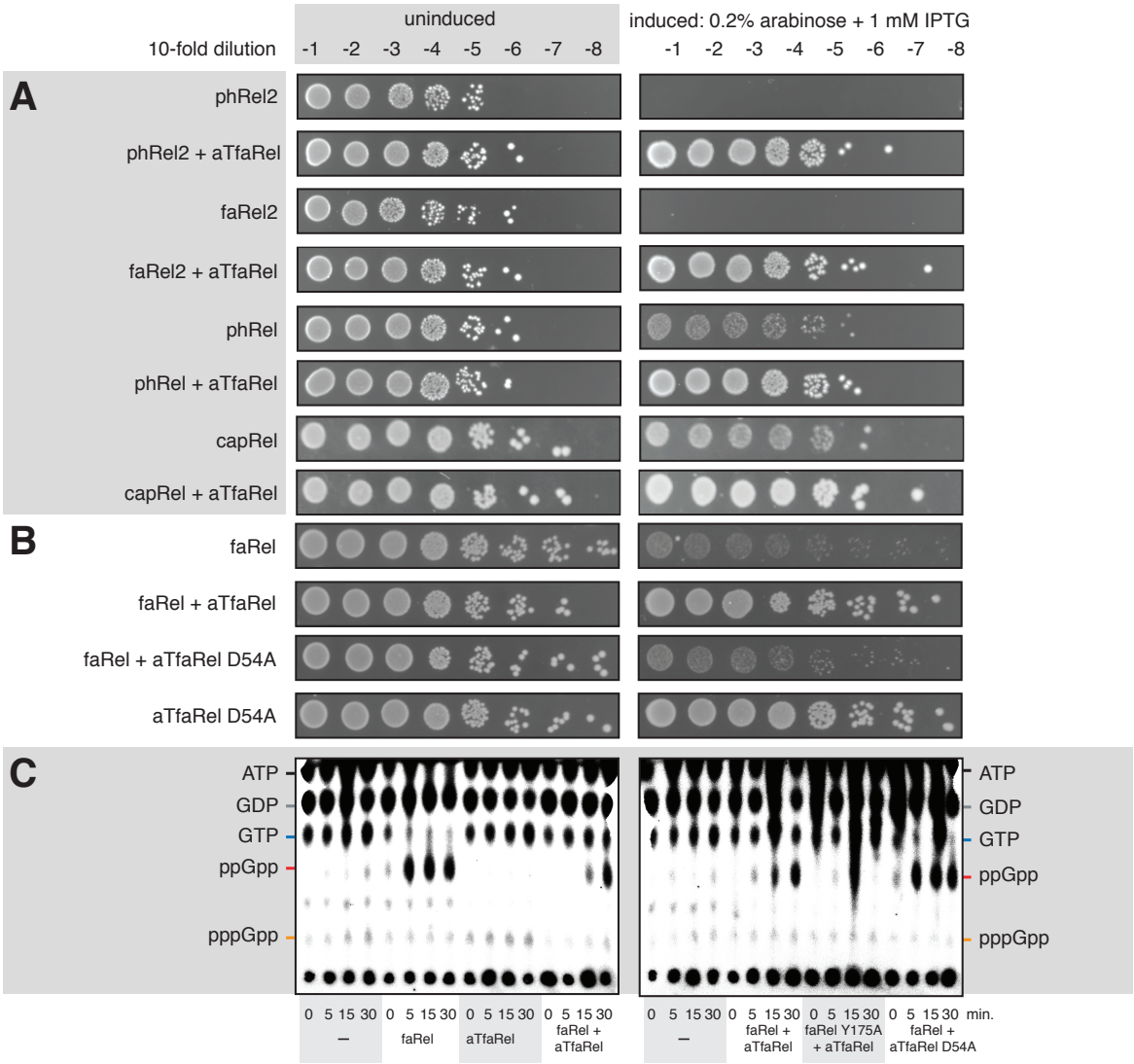

Figure 5 | The *C. marina* ATfaRel SAH universally counteracts all identified toxSASs and efficiently degrades (p)ppGpp produced by the FaRel toxSAS in live *E. coli* cells. (**A**) *C. marina* ATfaRel neutralises the toxicity of all identified toxSAS toxins. (**B**) Toxicity neutralisation by *C. marina* ATfaRel is abolished by the D54A mutation that inactivates the hydrolytic activity of ATfaRel. (**C**) The expression of *C. marina* FaRel leads to the accumulation of the alarmone ppGpp. Alarmone accumulation is efficiently counteracted by wild type ATfaRel but not its enzymatically-inactive D54A mutant. Autoradiograms of a representative TLC plate and a biological replicate are presented (Figure 5C and Supplementary Figure S6).

References

1. Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T. and Gerdes, K. (2015) Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol*, **13**, 298-309.
2. Ryals, J., Little, R. and Bremer, H. (1982) Control of rRNA and tRNA syntheses in Escherichia coli by guanosine tetraphosphate. *Journal of bacteriology*, **151**, 1261-1268.
3. Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S.E. and Murakami, K.S. (2018) Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol Cell*, **69**, 828-839 e825.
4. Milon, P., Tischenko, E., Tomsic, J., Caserta, E., Folkers, G., La Teana, A., Rodnina, M.V., Pon, C.L., Boelens, R. and Gualerzi, C.O. (2006) The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. *Proc Natl Acad Sci U S A*, **103**, 13962-13967.
5. Corrigan, R.M., Bellows, L.E., Wood, A. and Grundling, A. (2016) ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. *Proc Natl Acad Sci U S A*, **113**, E1710-1719.
6. Atkinson, G.C., Tenson, T. and Hauryliuk, V. (2011) The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS ONE*, **6**, e23479.
7. Steinchen, W., Schuhmacher, J.S., Altegoer, F., Fage, C.D., Srinivasan, V., Linne, U., Marahiel, M.A. and Bange, G. (2015) Catalytic mechanism and allosteric regulation of an oligomeric (p)ppGpp synthetase by an alarmone. *Proc Natl Acad Sci U S A*, **112**, 13348-13353.
8. Nanamiya, H., Kasai, K., Nozawa, A., Yun, C.S., Narisawa, T., Murakami, K., Natori, Y., Kawamura, F. and Tozawa, Y. (2008) Identification and functional analysis of novel (p)ppGpp synthetase genes in Bacillus subtilis. *Mol Microbiol*, **67**, 291-304.
9. Lemos, J.A., Lin, V.K., Nascimento, M.M., Abranches, J. and Burne, R.A. (2007) Three gene products govern (p)ppGpp production by Streptococcus mutans. *Mol Microbiol*, **65**, 1568-1581.
10. Hauryliuk, V. and Atkinson, G.C. (2017) Small Alarmone Synthetases as novel bacterial RNA-binding proteins. *RNA Biol*, **14**, 1695-1699.
11. Ruwe, M., Ruckert, C., Kalinowski, J. and Persicke, M. (2018) Functional Characterization of a Small Alarmone Hydrolase in Corynebacterium glutamicum. *Front Microbiol*, **9**, 916.
12. Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*, **23**, 324-328.
13. Harms, A., Brodersen, D.E., Mitarai, N. and Gerdes, K. (2018) Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell*, **70**, 768-784.
14. Hall, A.M., Gollan, B. and Helaine, S. (2017) Toxin-antitoxin systems: reversible toxicity. *Curr Opin Microbiol*, **36**, 102-110.
15. Goeders, N. and Van Melderen, L. (2014) Toxin-antitoxin systems as multilevel interaction systems. *Toxins (Basel)*, **6**, 304-324.
16. Page, R. and Peti, W. (2016) Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat Chem Biol*, **12**, 208-214.
17. Xiao, H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G. and Cashel, M. (1991) Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations. *J Biol Chem*, **266**, 5980-5990.
18. Dedrick, R.M., Jacobs-Sera, D., Bustamante, C.A., Garlena, R.A., Mavrich, T.N., Pope, W.H., Reyes, J.C., Russell, D.A., Adair, T., Alvey, R. *et al.* (2017) Prophage-mediated defence against viral attack and viral counter-defence. *Nature microbiology*, **2**, 16251.
19. Winther, K.S., Brodersen, D.E., Brown, A.K. and Gerdes, K. (2013) VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA. *Nat Commun*, **4**, 2796.
20. Castro-Roa, D., Garcia-Pino, A., De Gieter, S., van Nuland, N.A.J., Loris, R. and Zenkin, N. (2013) The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat Chem Biol*, **9**, 811-817.

21. Winther, K.S. and Gerdes, K. (2011) Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A*, **108**, 7403-7407.
22. Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K. and Ehrenberg, M. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell*, **112**, 131-140.
23. Berghoff, B.A. and Wagner, E.G.H. (2017) RNA-based regulation in type I toxin-antitoxin systems and its implication for bacterial persistence. *Curr Genet*, **63**, 1011-1016.
24. Rocker, A. and Meinhart, A. (2016) Type II toxin: antitoxin systems. More than small selfish entities? *Curr Genet*, **62**, 287-290.
25. Goeders, N., Chai, R., Chen, B., Day, A. and Salmond, G.P. (2016) Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems. *Toxins (Basel)*, **8**.
26. Masuda, H., Tan, Q., Awano, N., Wu, K.P. and Inouye, M. (2012) YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli. *Mol Microbiol*, **84**, 979-989.
27. Loman, N.J. and Pallen, M.J. (2015) Twenty years of bacterial genome sequencing. *Nat Rev Microbiol*, **13**, 787-794.
28. Saha, C.K., Sanchez Pires, R., Brolin, H. and Atkinson, G.C. (2018) Predicting Functional Associations using Flanking Genes (FlaGs). *bioRxiv*, **doi: https://doi.org/10.1101/362095**.
29. Bloomfield, G.A., Whittle, G., McDonagh, M.B., Katz, M.E. and Cheetham, B.F. (1997) Analysis of sequences flanking the vap regions of Dichelobacter nodosus: evidence for multiple integration events, a killer system, and a new genetic element. *Microbiology*, **143 (Pt 2)**, 553-562.
30. Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct*, **4**, 19.
31. Kaldalu, N., Kasari, V., Atkinson, G. and Tenson, T. (2013) In Gerdes, K. (ed.), *Prokaryotic Toxin-Antitoxins*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 93-105.
32. Aakre, C.D., Phung, T.N., Huang, D. and Laub, M.T. (2013) A bacterial toxin inhibits DNA replication elongation through a direct interaction with the beta sliding clamp. *Mol Cell*, **52**, 617-628.
33. Jurėnas, D., Chatterjee, S., Konijnenberg, A., Sobott, F., Droogmans, L., Garcia-Pino, A. and Van Melderen, L. (2017) AtaT blocks translation initiation by N-acetylation of the initiator tRNA(fMet). *Nat Chem Biol*, **13**, 640-646.
34. Manav, M.C., Beljantseva, J., Bojer, M.S., Tenson, T., Ingmer, H., Hauryliuk, V. and Brodersen, D.E. (2018) Structural basis for (p)ppGpp synthesis by the Staphylococcus aureus small alarmone synthetase RelP. *J Biol Chem*, **293**, 3254-3264.
35. Geiger, T., Kastle, B., Gratani, F.L., Goerke, C. and Wolz, C. (2014) Two small (p)ppGpp synthases in Staphylococcus aureus mediate tolerance against cell envelope stress conditions. *J Bacteriol*, **196**, 894-902.
36. Beljantseva, J., Kudrin, P., Andresen, L., Shingler, V., Atkinson, G.C., Tenson, T. and Hauryliuk, V. (2017) Negative allosteric regulation of Enterococcus faecalis small alarmone synthetase RelQ by single-stranded RNA. *Proc Natl Acad Sci U S A*, **114**, 3726-3731.
37. Abranches, J., Martinez, A.R., Kajfasz, J.K., Chavez, V., Garsin, D.A. and Lemos, J.A. (2009) The molecular alarmone (p)ppGpp mediates stress responses, vancomycin tolerance, and virulence in Enterococcus faecalis. *J Bacteriol*, **191**, 2248-2256.
38. Steinchen, W., Vogt, M.S., Altegoer, F., Giammarinaro, P.I., Horvatek, P., Wolz, C. and Bange, G. (2018) Structural and mechanistic divergence of the small (p)ppGpp synthetases RelP and RelQ. *Sci Rep*, **8**, 2195.
39. Hogg, T., Mechold, U., Malke, H., Cashel, M. and Hilgenfeld, R. (2004) Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response [corrected]. *Cell*, **117**, 57-68.

40. Hatfull, G.F. and Hendrix, R.W. (2011) Bacteriophages and their genomes. *Current opinion in virology*, **1**, 298-303.
41. Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*, **44**, W16-21.
42. Santos, S.B., Kropinski, A.M., Ceyssens, P.J., Ackermann, H.W., Villegas, A., Lavigne, R., Krylov, V.N., Carvalho, C.M., Ferreira, E.C. and Azeredo, J. (2011) Genomic and proteomic characterization of the broad-host-range Salmonella phage PVP-SE1: creation of a new phage genus. *J Virol*, **85**, 11265-11273.
43. Kim, M., Kim, S. and Ryu, S. (2012) Complete genome sequence of bacteriophage SSU5 specific for Salmonella enterica serovar Typhimurium rough strains. *J Virol*, **86**, 10894.
44. Horesh, G., Harms, A., Fino, C., Parts, L., Gerdes, K., Heinz, E. and Thomson, N.R. (2018) SLING: a tool to search for linked genes in bacterial datasets. *Nucleic Acids Res*.
45. Wei, Y.Q., Bi, D.X., Wei, D.Q. and Ou, H.Y. (2016) Prediction of Type II Toxin-Antitoxin Loci in Klebsiella pneumoniae Genome Sequences. *Interdisciplinary sciences, computational life sciences*, **8**, 143-149.
46. Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Dreze, P. and Van Melderen, L. (2011) Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res*, **39**, 5513-5525.
47. Song, S. and Wood, T.K. (2018) Post-segregational Killing and Phage Inhibition Are Not Mediated by Cell Death Through Toxin/Antitoxin Systems. *Frontiers in microbiology*, **9**, 814.
48. Fernandez, L., Gonzalez, S., Campelo, A.B., Martinez, B., Rodriguez, A. and Garcia, P. (2017) Low-level predation by lytic phage phiIPLA-RODI promotes biofilm formation and triggers the stringent response in Staphylococcus aureus. *Sci Rep*, **7**, 40965.
49. Slominska, M., Neubauer, P. and Wegrzyn, G. (1999) Regulation of bacteriophage lambda development by guanosine 5'-diphosphate-3'-diphosphate. *Virology*, **262**, 431-441.
50. Tabib-Salazar, A., Liu, B., Barker, D., Burchell, L., Qimron, U., Matthews, S.J. and Wigneshweraraj, S. (2018) T7 phage factor required for managing RpoS in Escherichia coli. *Proc Natl Acad Sci U S A*, **115**, E5353-E5362.
51. Blower, T.R., Pei, X.Y., Short, F.L., Fineran, P.C., Humphreys, D.P., Luisi, B.F. and Salmond, G.P. (2011) A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat Struct Mol Biol*, **18**, 185-190.
52. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, **30**, 772-780.
53. Eddy, S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol*, **7**, e1002195.
54. Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, **26**, 1641-1650.
55. Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276-3278.
56. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312-1313.
57. Minh, B.Q., Nguyen, M.A. and von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*, **30**, 1188-1195.
58. Miller, M.A., Pfeiffer, W. and Schwartz, T. (2010), *Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, pp. 1-8.