Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families

Raphaël Leplae[1], Damien Geeraerts[2], Régis Hallez[2], Julien Guglielmini[2], Pierre Drèze[2] and Laurence Van Melderen[2,*]

[1]Laboratoire de Bioinformatique des Génomes et des Réseaux (BiGRe), Faculté des Sciences, Université Libre de Bruxelles, Bld du Triomphe, 1050 Bruxelles, Belgium and [2]Laboratoire de Génétique et Physiologie Bactérienne, Institut de Biologie et de Médecine Moléculaires, Faculté des Sciences, Université Libre de Bruxelles, 12 rue des Professeurs Jeener et Brachet, 6041 Gosselies, Belgium

ABSTRACT

Type II toxin–antitoxin (TA) systems are generally composed of two genes organized in an operon, encoding a labile antitoxin and a stable toxin. They were first discovered on plasmids where they contribute to plasmid stability by a phenomenon denoted as 'addiction', and subsequently in bacterial chromosomes. To discover novel families of antitoxins and toxins, we developed a bioinformatics approach based on the 'guilt by association' principle. Extensive experimental validation in *Escherichia coli* of predicted antitoxins and toxins increased significantly the number of validated systems and defined novel toxin and antitoxin families. Our data suggest that toxin families as well as antitoxin families originate from distinct ancestors that were assembled multiple times during evolution. Toxin and antitoxin families found on plasmids tend to be promiscuous and widespread, indicating that TA systems move through horizontal gene transfer. We propose that due to their addictive properties, TA systems are likely to be maintained in chromosomes even though they do not necessarily confer an advantage to their bacterial hosts. Therefore, addiction might play a major role in the evolutionary success of TA systems both on mobile genetic elements and in bacterial chromosomes.

INTRODUCTION

Toxin–antitoxin (TA) systems are small genetic modules found on bacterial mobile genetic elements as well as in bacterial chromosomes. They appear to be specific of the eu-bacterial and archae-bacterial worlds as no homologous sequences are detected in eukaryotic genomes (except for the PIN toxin domain which is present in eukaryotes) (1). Toxins are always proteins but, based on the nature of the antitoxin and its mode of action, TA systems are currently divided into three classes. Antitoxins of type I and III systems are small RNAs that inhibit either toxin expression (type I) or activity (type III) (2,3). Antitoxins of type II systems are proteins that inactivate toxins by protein–protein complex formation (4). Type I and II systems were discovered in the mid-1980s on bacterial plasmids while type III was discovered only recently. For type I and II systems, it became rapidly clear that these systems were dedicated to plasmid maintenance. In addition to several mechanisms that prevent plasmid-free cells production, plasmids have evolved subtle molecular systems that act by killing plasmid-free daughter bacteria after plasmid loss. Type I and II systems are responsible for this phenomenon denoted as post-segregational killing or addiction (5,6). Addiction relies on the differential stability of the two components, antitoxins being less stable than toxins. In daughter bacteria that do not inherit a plasmid copy, the antitoxin and toxin pool is not replenished. Since the antitoxin is labile, the toxin is released from inhibition, leading to plasmid-free cell death. The cell is thus addicted to antitoxin production and *in*

extenso to the presence of the TA genes (5). Therefore, type I and II systems contribute to an apparent plasmid stabilization at the population level. An extension of this function for type II systems is the capacity of plasmid-encoded systems to outcompete plasmids of the same incompatibility group devoid of TA systems (7).

Bacterial genome sequencing and database mining revealed that homologs of type I and II systems are found in bacterial chromosomes; the occurrence of type II systems being surprisingly high (8–12). Analysis of their distribution provided information regarding their evolution. While type I systems appear to have evolved by lineage specific-duplication (13), type II systems are thought to move from one genome to another by horizontal gene transfer (9,14). The number of type II systems varies greatly not only from one bacterial species to another, but also between isolates from the same species (9,15). The function(s) of chromosomally encoded type I and II systems remain(s) unclear (2,16,17). Although it has been proposed that type II systems might serve as stress response modules, convincing data are still lacking (18). Recent studies have shown that type II systems are involved in the stabilization of large genomic fragments (19) and of integrative conjugative elements (20), indicating that some TA systems might have retained their addictive properties.

Genetic organization of known type II TA systems is quite conserved. Typically, these systems comprise two small genes organized in an operon, the upstream gene encoding the antitoxin. General features such as small size of both components (31–204 amino acid for antitoxins and 41–206 amino acid for toxins) and short intergenic regions separating the two genes (-20 to $+30$ nt) are also well conserved (9). In general, antitoxins are composed of two domains, an amino-terminal domain responsible for DNA binding and a carboxy-terminal domain responsible for toxin interaction. The antitoxin–toxin protein complex, in which the toxin is inactive, is also responsible for negative autoregulation. Toxins typically are small, stable proteins that inhibit either replication by interacting with DNA-gyrase, or translation by cleaving messenger RNAs (mRNAs) or inhibiting elongation.

Classification of the type II TA systems is based on the amino acid sequence similarity of the toxins, each toxin family being associated with a specific antitoxin family. Thus, type II systems are currently divided into 10 families (8,9). However, a few examples of 'hybrid' associations of a toxin from one family and an antitoxin belonging to another family have recently been characterized (21,22). Furthermore, novel putative families of toxins and antitoxins (i.e. not homologous to the 10 families) have been predicted by bioinformatics approaches recently and some were found to be associated with known toxin or antitoxin families, indicating that 'hybrid' systems might be more common than originally thought (12,23,24). In fact, one such prediction in *Escherichia coli* K-12 was recently validated experimentally (8). Thus, predictions are that the type II TA systems are much more abundant and diversified than what is currently described. To evaluate this diversity, a bioinformatics approach was developed to explore prokaryotic genomes and to identify

novel toxins and antitoxins. From our predictions, 18 antitoxin and 23 toxin sequences originating from different bacterial species were validated experimentally in *E. coli*, significantly increasing the number of families of type II antitoxins and toxins. Therefore, we propose to refer to antitoxin and toxin families independently instead of referring to TA system families.

MATERIALS AND METHODS

Bioinformatics approach

This approach is described in details in Supplementary 'Materials and Methods' section and summarized in Figure 1.

Bacterial strains, plasmids and media

Bacterial strains. The following *E. coli* strains were used: MC1061 (F^- *araD$_{139}$*, *Δ(ara-leu)$_{7696}$*, *galE15*, *galK16*, *Δ(lac)X74*, *rpsL (Str^r)*, *hsdR2 (rK$^-$mK$^+$)*, *mcrA mcrB1*) (25), MG1655 (*rph1 ilvG rfb-50*) (26), DJ624 *Δara* (MG1655 *lacX74 mal::lacI^q*) (27) and DJ624λ*sfiA::lacZ* (this work).

Plasmids. The following plasmids were used and their relevant characteristics are indicated. The pBAD33 vector (p15A, Cm^r, pBAD promoter) (28), the pBAD33-*yoeB* plasmid (pBAD33 derivative containing the *yoeB* toxin gene under the control of the pBAD promoter) (29), the pKK223-3 vector (ColE1, Amp^r, pTac promoter) (30) and the pLac-staby vector (DelphiGenetics, Belgium) were used in this work.

Media. Luria-Bertani liquid and agar medium (LB) (Invitrogen) as well as M9 minimal liquid and agar medium (KH$_2$PO$_4$ (22 mM), Na$_2$HPO$_4$ (42 mM), NH$_4$Cl (19 mM), MgSO$_4$ (1 mM), CaCl$_2$ (0.1 mM), NaCl (9 mM) supplemented with casamino acids [0.2% except for methionine incorporation (0.05%)] and carbon sources [glycerol, glucose or arabinose (1%)].

Antibiotics. Antibiotics were added at the following concentrations: chloramphenicol, 20 µg/ml; ampicillin, 100 µg/ml.

Experimental validation of novel toxins and antitoxins sequences

The experimental validation of the putative novel toxins and antitoxins families relies on a simple assay: expression of the putative toxins should cause growth inhibition while co-expression with their putative cognate antitoxins should restore normal growth. To that end, antitoxin and toxin candidate genes were cloned in compatible expression vectors under the control of inducible promoters.

Toxins cloning. Predicted toxin genes were cloned in the pBAD33 vector, under the control of the arabinose-inducible pBAD promoter. The predicted toxin coding sequences (CDS) were amplified from bacterial genomic DNA using the appropriate start and stop primers (Supplementary Table S1). Start primers carry a canonical

Shine–Dalgarno (SD) sequence. Polymerase chain reaction (PCR) products were digested by XbaI and PstI, except the $sauT3_{COL}$, $nspT5_{PC7120}$, $speT2_{TIGR4}$, $speT1_{TIGR4}$ amplification products that were digested by SacI and KpnI; and $ccrT3_{CB15}$ that was digested by XbaI and SalI. Digested products were ligated into the pBAD33 vector cleaved with the appropriate restriction enzymes. Toxins that did not confer an unambiguous growth inhibition phenotype when expressed from the pBAD33 vector were cloned in the pKK223-3 vector under the control of the pTac promoter. The PCR products were digested by EcoRI and PstI and cloned in the pKK223-3 vector cleaved with the same restriction enzymes.

Antitoxins cloning. Predicted antitoxin genes were cloned in the pLac-staby vector and/or in the pKK223-3 vector, which places the antitoxin genes under the control of the pLac and pTac promoters, respectively. Expression from both promoters is induced by isopropyl β-D-1-thiogalacto-pyranoside (IPTG) addition. Predicted CDSs were amplified from genomic DNA using the appropriate start and stop primers (Supplementary Table S1) and PCR products were ligated in the pLac-staby vector. The predicted antitoxins that failed to counteract the growth inhibition of their cognate toxins were cloned in the pKK223-3 vector under pTac control. The PCR products were digested by EcoRI and PstI and cloned in the pKK223-3 vector cleaved with the same restriction enzymes. To test the activity of antitoxins associated with toxins cloned in the pKK223-3 vector, the cognate antitoxin genes were cloned in the pBAD33 vector. PCR products were digested by XbaI and PstI and cloned in the pBAD33 digested with the same restriction enzymes.

Killing/rescue assay. The DJ624Δara strain was transformed by toxin-encoded plasmids and/or by antitoxin-encoded plasmids and the control vectors. Transformants were grown overnight in M9 liquid medium containing glucose (1%) and the appropriate antibiotics. Overnight cultures were diluted 100-fold in M9 medium containing glucose (1%) and the appropriate antibiotics and grown at 37°C to an OD 600 nm of 0.2. Dilutions (10^0–10^6) were plated on M9 plates containing either glucose (1%) or arabinose (1%) and IPTG (1 mM) and the appropriate antibiotics. Plates were incubated overnight at 37°C.

SOS induction

The DJ624λsfiA::lacZ strain containing the pBAD33, pBAD-$parE_{RK2}$, pBAD-$ccrT4_{CB15}$ or pBAD-$atuT1_{c58}$ plasmids was grown in M9 medium supplemented with glucose (1%) and chloramphenicol to an OD 600 nm of 0.1. Cultures were centrifuged and pellets resuspended in M9 medium containing arabinose (1%) and chloramphenicol. After 120 min of induction, aliquots were taken to perform β-galactosidase assays as described in ref. 31.

^{35}S-methionine incorporation

The DJ624Δara strain containing the pBAD33 vector and its derivates containing the experimentally validated toxin genes and appropriate controls were grown in M9 medium supplemented with glycerol (1%) at 37°C. Toxin over-expression was induced at an OD 600 nm of 0.1 by adding arabinose (1%). After 180 min of induction, 1 ml aliquots were taken and 5 μCi of ^{35}S-methionine was added. After 2 min of incubation at 37°C, 500 μl aliquots were removed and added to tubes containing 5 ml of cold 10% TCA and left 20 min on ice. Samples were then filtered on nitrocellulose filters (0.45 μm) saturated with the non-labeled precursor using a glass funnel. Filters were then washed in 10% TCA and air-dried. Filter-retained material was counted in 10 ml of scintillation liquid (Ready Protein +) in a liquid scintillation counter (Beckman). Translation rate was normalized to the ^{35}S-methionine incorporated by the control strain (pBAD33 vector).

RESULTS

Detecting 'similar' and potentially novel toxins and antitoxins

We collected the antitoxin and toxin sequences of type II systems that were experimentally validated and characterized at the time we started this work. In addition to the *relBE*, *parDE*, *vapBC*, *mazEF*, *phd-doc*, *higAB*, *hipBA*, *omega-epsilon-zeta* and *ccd* systems and their homologs, sequences of the *vapXD* system were added to our data set even though this system is poorly documented (32). We also included sequences of the $parD_{EDL933}$-*parE3*, *paaA2-parE2* (33) and *yafNO* systems that were identified by our procedure and tested as 'proof of concept'. Note that *yafNO* characterization was published during the course of this work (34). The *hicAB* system was not included since it was not validated at the time we started this analysis (8). Our query data set was therefore composed of 24 sequences of antitoxin and 24 sequences of toxin that were denoted as 'original' sequences (Supplementary Table S2). We performed a comprehensive PSI-BLAST search (see Supplementary 'Materials and Methods' section) on 2181 prokaryotic genomes and detected more than 10 000 sequences. These sequences were grouped in eight toxin super-families and nine of antitoxin (Supplementary Tables S3 and S4).

Based on the conservation of type II systems organization, we developed an exploratory bioinformatics approach based on the 'association by guilt' of orphan toxins or antitoxins (i.e. not paired with 'original' antitoxins or toxins) to discover novel sequences potentially showing antitoxin or toxin activities (Figure 1 and see Supplementary 'Materials and Methods' section for details). These sequences were grouped into protein families based on their similarities (see Supplementary 'Materials and Methods' section for the detailed procedure and parameters). As a result, we distinguished two types of predicted antitoxin or toxin sequences: those belonging to a family composed of proteins of unknown function which were called AG (for 'associated by guilt'), and those 'associated by guilt' but present in a family containing at least one 'original' sequence or one sequence presenting similarity with 'original' sequences (denoted as 'similar', see Supplementary 'Materials and
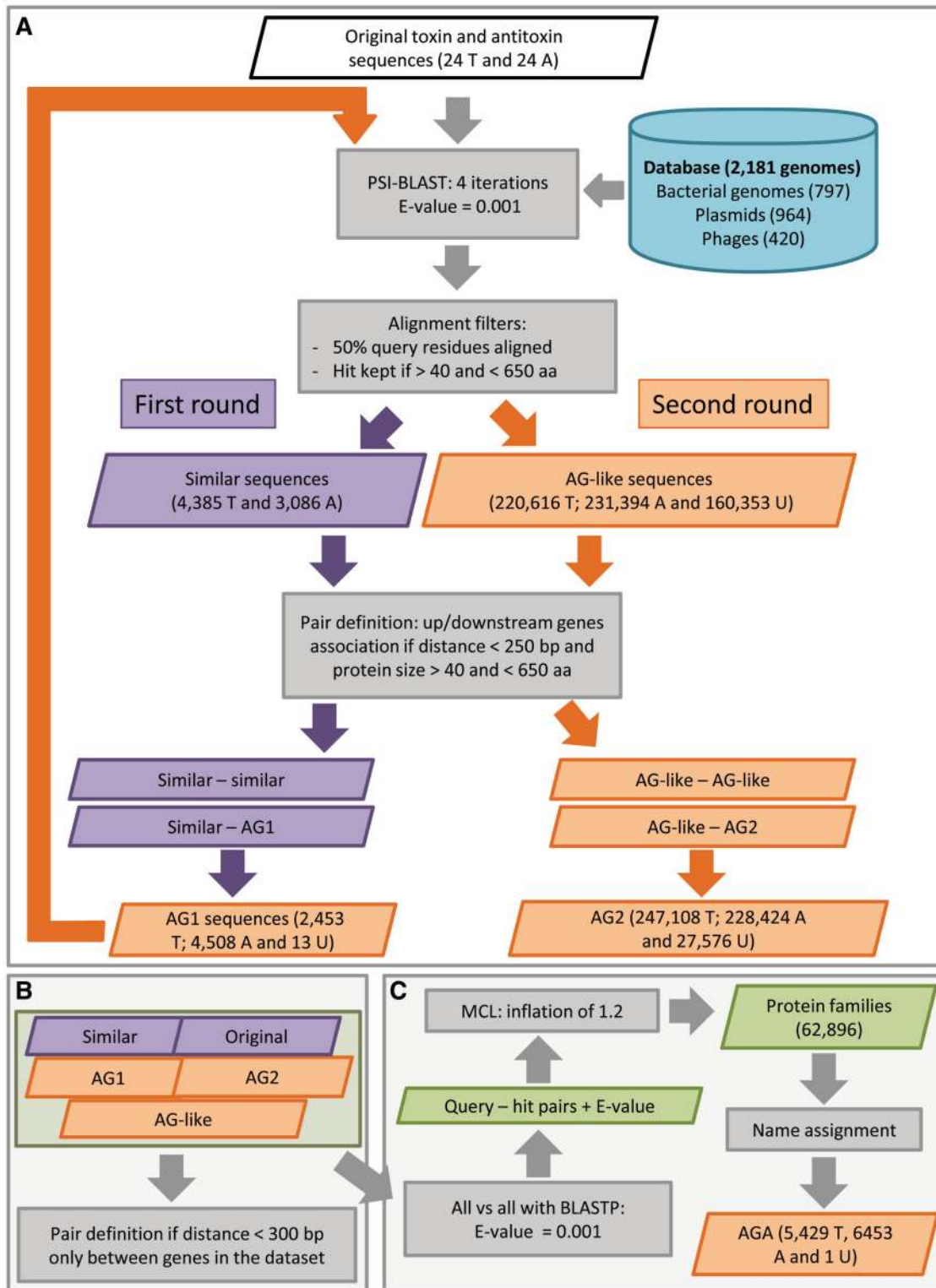
Figure 1. Association by guilt bioinformatics approach. Detection of 'associated by guilt' (AG) sequences (**A**). 'Similar' sequences (purple box) to the 48 'original' sequences (white box) were detected by PSI-BLAST searches using the indicated parameters (gray boxes) and a database composed of 2181 genome sequences (blue cylinder). First round: AG toxins and antitoxins denoted as AG1 (orange box) were identified by paring of the 'similar' sequences using the indicated parameters (gray box). Number and type of sequences are indicated (T for toxin, A for antitoxin and U for unassigned). Second round: AG1 sequences were used as query for a second round of detection using the same parameters as in the initial steps to identify AG-like sequences (orange box). AG2 sequences were identified by pairing the AG-like sequences. Number and type of sequences are indicated. Pair definition (**B**). An additional round of pair definition was performed using sequences detected in (A) using the indicated parameters (gray box). Protein families (**C**). Protein families (green box) were generated by grouping all the proteins present in the dataset using the Markov clustering algorithm (MCL) with the indicated parameters (gray boxes). Names of 'original' or 'similar' sequences were propagated to the AG proteins within the same protein family. These sequences were then defined as 'associated by guilt and annotated' (AGA).

Methods' section for details). Super-family names of these 'original' and/or 'similar' sequences were transferred to these AG sequences and they were denoted consequently 'associated by guilt and annotated' (AGA).

Experimental validation of predicted toxins and antitoxins

Our bioinformatics exploratory procedure led to the prediction of a large number of novel AG and AGA toxins and antitoxins located upstream and/or downstream of orphan 'similar' and originating from numerous bacterial species.

Experimental validation of these AG and AGA sequences was carried out in *E. coli* using a simple Kill/ Rescue assay, in which expression of the toxin gene is expected to inhibit cell growth while co-expression of the toxin and antitoxin genes restores growth.

The toxin and antitoxin sequences were cloned in compatible vectors, under the control of different inducers (see 'Materials and Methods' section). These sequences were placed in the same genetic context, i.e. RBS and ATG as start codon, to achieve a comparable level of expression, although we cannot rule out that some genes are less expressed than others as we did not measure expression levels.

Predicted toxins

Twenty-three AG toxins were experimentally tested. Eight of them are paired with sequences 'similar' to 'original' antitoxins (Supplementary Table S5). Expression of the putative $BceT5_{E33L}$, $SpyT2_{10270}$ and $LmoT1_{EGD-e}$ toxins did inhibit *E. coli* growth. Among the upstream open reading frames (ORFs) predicted to be antitoxins, only $BceA5_{E33L}$, which is associated with the $BceT5_{E33L}$ toxin, turned out to be a functional antitoxin belonging to the HigA super-family of antitoxins. For the $SpyT2_{10270}$ toxin, the antitoxin activity of the downstream $SpyA2_{10270}$ ORF, although not predicted by our approach, was tested. Co-expression of $SpyA2_{10270}$ relieved $SpyT2_{10270}$-mediated cell growth inhibition. In this case, the antitoxin gene is located downstream of the toxin gene. For the third toxin that inhibits *E. coli* growth, $LmoT1_{EGD-e}$, none of the flanking ORFs showed antitoxin activity.

Seven AG toxins associated with AGA antitoxins belonging to the HigA super-family were tested (Supplementary Table S6). Expression of four of them ($SmeT1_{1021}$, $MavT1_{K10}$, $SpyT1_{9429}$ and $SpyT3_{10270}$) inhibited *E. coli* growth. However, only two of the predicted antitoxins ($SmeA1_{1021}$ and $MavA1_{K10}$) relieved the growth inhibition mediated by their cognate toxins. For the $SpyT1_{9429}$ and $SpyT3_{10270}$ toxins, we were unable to detect antitoxin activity for their flanking ORFs.

Eight AG toxins associated with eight AG antitoxins were experimentally tested (Supplementary Table S7). Expression of $BceT1_{E33L}$, $SpyT1_{10270}$, $SpyT1_{M1}$ and $EcoT1_{EDL933}$ resulted in *E. coli* growth inhibition and co-expression of their paired antitoxins relieved growth inhibition (Figure 2).

One predicted pair (MGAS10270 Spy0568 and Spy0569) presented a particular phenotype. This pair is composed of sequences similar to the *E. coli hicAB*
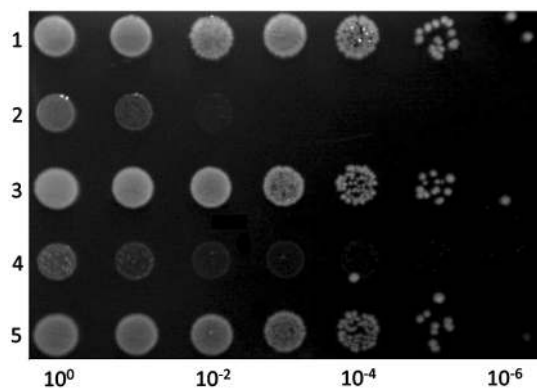

Figure 2. The $ecoA1$-$ecoT1_{EDL933}$ and the $nspA5$-$nspT5_{PC7120}$ pairs constitute TA systems. The DJ624Δara strain containing the pBAD33 and pLac-staby plasmids (1), the pBAD33-$ecoT1_{EDL933}$ and pLac-staby plasmids (2), the pBAD33-$ecoT1_{EDL933}$ and pLac-staby-$ecoA1_{EDL933}$ plasmids (3), the pBAD33-$nspT5_{PC7120}$ and pLac-staby plasmids (4) or the pBAD33-$nspT5_{PC7120}$ and pLac-staby- $nspA5_{PC7120}$ plasmids (5) were grown in log phase in M9 medium supplemented with glucose (1%) and appropriate antibiotics. Serial dilutions (as indicated) were spotted on LB plates containing arabinose (1%) and IPTG (1 mM). Plates were incubated overnight at 37°C.

system (8). HicA was shown to be a toxin cleaving mRNAs and HicB was shown to be its antitoxin. Our prediction was opposite: Spy0568, the *S. pyogenes* HicB homolog was predicted to be a toxin and Spy0569, the HicA homolog, the antitoxin. Surprisingly, expression of both proteins was toxic for *E. coli* (data not shown). We therefore renamed Spy0568 as $SpyT4_{10270}$ and Spy0569 as $SpyT5_{10270}$. The flanking ORFs being located on the complementary strand, we did not check their antitoxic activity.

Predicted antitoxins

Eleven AG antitoxins were experimentally tested. Eight predicted antitoxins are associated with sequences 'similar' to 'original' toxins (Supplementary Table S8). Expression of these toxins inhibited *E. coli* growth except for the RPC4130 ORF. The seven cognate antitoxins relieved the toxicity mediated by their associated toxins.

Three of these antitoxins ($CcrA1_{CB15}$, $SpeA2_{TIGR4}$ and $CcrA4_{CB15}$) were validated by other groups during the course of this work, although their annotation turned out to be incorrect (9,35). The $CcrA1_{CB15}$ antitoxin is associated with a sequence similar to ParE and was described as being a $ParD_{RK2}$-like antitoxin in *Caulobacter crescentus* (36). We were unable to detect any sequence similarity between this sequence and the $ParD_{RK2}$ antitoxin. The second case concerns the $SpeA2_{TIGR4}$ antitoxin, which is associated with a toxin from the RelE family in *Streptococcus pneumoniae*. This ORF was named RelB2*spn* in a previous work (35). However, again, we were unable to detect any sequence similarity between this sequence and RelB$_{K-12}$. The last case concerns the $ccrA4_{CB15}$-$ccrT4_{CB15}$ system that was annotated as *relBE2* in *C. crescentus* (9) and experimentally analyzed (36). The $CcrT4_{CB15}$ toxin belongs to the ParE/RelE super-family. We showed that expression of

this toxin induces the SOS response, although less efficiently than $ParE_{RK2}$ (Figure 3), confirming our predictions and indicating that $CcrT4_{CB15}$ is a ParE-like toxin. The associated $CcrA4_{CB15}$ antitoxin, however, does not show any similarities to 'original' antitoxins. Thus, these three antitoxins represent novel sequences associated with 'similar 'toxins that present a growth inhibition phenotype do not show any similarities with 'original' antitoxins and were not described previously. The $AtuA1_{c58}$ antitoxin is associated with $AtuT1_{c58}$, which belongs to the ParE/RelE toxin super-family. Expression of this toxin induces the SOS system, indicating that $AtuT1_{c58}$ is a ParE-like toxin (Figure 3). The $NspA5_{PC7120}$ and $AfuA2_{DSM4304}$ antitoxins are paired with toxins from the VapC super-family and the $SpeA3_{TIGR4}$ antitoxin is associated with a toxin belonging to the Doc super-family.

Three AG antitoxins associated with AGA toxins were experimentally tested (Supplementary Table S6). Expression of these three toxins ($NspT1_{PC7120}$, $NspT2_{PC7120}$ and $NeuT1_{C91}$) resulted in *E. coli* growth inhibition and co-expression of their associated antitoxins relieved this inhibition (Figure 2). Two of these toxins belong to the VapC super-family and the other to the ParE/RelE super-family.

Thus, this approach led to the validation of 23 toxins and 18 antitoxins. For 5 of the 23 toxins, we were unable to identify the cognate antitoxin, suggesting that they might be solitary toxins or that expression levels of some of the antitoxins might be too low to rescue the toxic phenotype. Eleven of the 34 predicted toxins did not affect *E. coli* growth; it is possible that, as mentioned above, expression levels are too low; these sequences might be false-positives (not related to toxins) or they might be 'true' toxins but only for their natural host and not for *E. coli*.

Toxin activities

To gain insights into the mode of toxicity of the novel validated toxins, several tests were performed. Cell morphology was observed by microscopy and SOS induction,
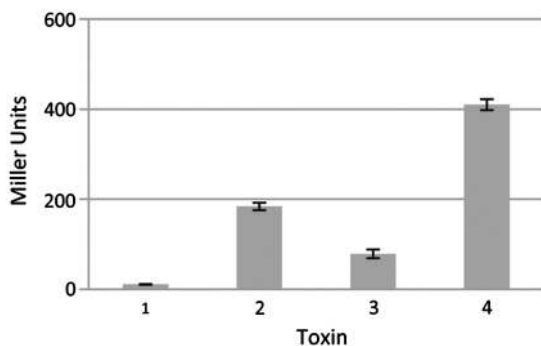

Figure 3. Overexpression of the $CcrT4_{CB15}$ and $AtuT1_{C58}$ toxins induce the SOS system in *E. coli*. The DJ624λ*sfiA::lacZ* strain containing the pBAD33 (1), pBAD-$parE_{RK2}$ (2), pBAD-$ccrT4_{CB15}$ (3) or pBAD-$atuT1_{c58}$ (4) plasmids were grown in M9 medium. After induction of toxin expression by arabinose addition, samples were taken to perform β-galactosidase assays as described in 'Materials and Methods' section.

transcription as well as translation rates were measured in *E. coli* overexpressing the active toxins. In addition to the novel toxins, we also tested the AGA $NspT1_{PC7120}$ and $NspT5_{PC7120}$ toxins belonging to the VapC super-family in order to confirm this annotation. Neither cell morphology, SOS induction nor transcription rates were affected by toxin overexpression (data not shown and Supplementary Figure S1). However, expression of these toxins drastically reduced ^{35}S-methionine incorporation (from 10% to 30% as compared to the control vector) (Figure 4). These data show that all of the tested toxins affect translation. In addition, preliminary data seem to indicate that expression of some of these toxins induces mRNA cleavage in *E. coli* (data not shown).

Evolutionary relationships between novel and 'original' super-families

Six of the 23 toxins validated experimentally define four novel super-families (Table 1). These super-families were denoted as Gin (for growth inhibition).

Three of the novel toxins constitute the GinA super-family ($SpyT1_{10270}$, $SpyT2_{10270}$ and $BceT1_{E33L}$). The GinB super-family is composed of the $SmeT1_{1021}$ novel toxin as well as the YgjN toxin, which was identified in *E. coli* during the course of this work (37). This toxin was considered as a homolog of RelE (37). However, while we detected sequence similarity between $SmeT1_{1021}$ and YgjN (data not shown, E-value: 10e-7), we were unable to detect any sequence similarity between $SmeT1_{1021}$ and the members of the ParE/RelE super-family. The GinC and GinD super-families are composed of $SpyT1_{M1}$ and $BceT5_{E33L}$, respectively. Regarding the solitary toxins, they define four novel super-families (data not shown).

Twelve of the 18 antitoxins that were experimentally validated define 11 novel antitoxin super-families that
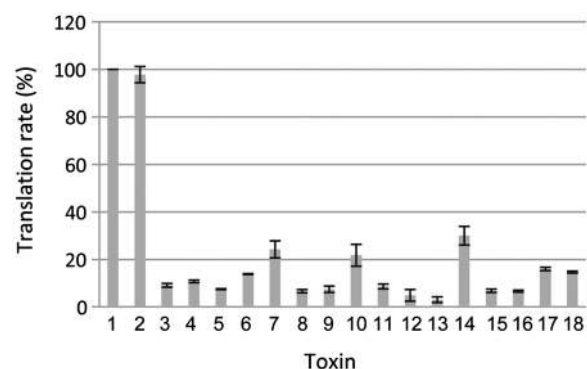

Figure 4. Overexpression of the novel toxins inhibits translation in *E. coli*. The DJ624Δ*ara* strain containing the pBAD33 (1), pBAD-$parE_{RK2}$ (2), pBAD-$yoeB$ (3), pBAD-$ecoT1_{EDL933}$ (4), pBAD-$mavT1_{K10}$ (5), pBAD-$spyT1_{10270}$ (6), pBAD-$spyT2_{10270}$ (7), pBAD-$bceT1_{E33L}$ (8), pBAD-$smeT1_{1021}$ (9), pBAD-$spyT1_{M1}$ (10), pBAD-$bceT5_{E33L}$ (11), pBAD-$nspT1_{PC7120}$ (12), pBAD-$nspT2_{PC7120}$ (13), pBAD-$spyT4_{10270}$ (14), pBAD-$spyT5_{10270}$ (15), pBAD-$spyT3_{10270}$ (16), pBAD-$spyT1_{9429}$ (17) and pBAD-$lmoT1_{EGD-e}$ (18) were grown in M9 medium. After induction of toxin expression by arabinose addition, cultures were labeled with ^{35}S-methionine. Translation rate was measured as described in 'Materials and Methods' section.

Table 1. Occurrence of 12 toxin super-families in 2181 prokaryotic genomes

	Toxin super-families (n = 7034)												Total
	RelE/ParE (n = 2057)	CcdB/MazF (n = 645)	Zeta (n = 1434)	Doc (n = 859)	YafO (n = 25)	VapD (n = 29)	HipA (n = 421)	VapC (n = 1267)	GinA (n = 101)	GinB (n = 98)	GinC (n = 89)	GinD (n = 9)	12
'Original' sequences	$RelE_{K-12}$, $RelE_{307}$, HigB, PasB, StbE, YoeB, Txe, YafQ, $ParE_{RK2}$, ParE1, ParE3	$CcdB_F$, $CcdB_{O157}$, MazF, YdcE, PemK, ChpBK	Zeta	Doc	YafO	VapD	HipA	VapC, MvpT	–	–	–	–	24
Validated 'similar' sequences	$CcrT1_{CB15}$, $AtuT1C_{58}$, $CcrT4_{CB15}$, $SpeT2_{TIGR4}$	NT	NT	$SpeT3_{TIGR4}$	NT	NT	NT	$AfuT2_{DSM4304}$, $NspT5_{PC7120}$	–	$YgjN$[a]	–	–	7
Validated AGA sequences	$NeuT1_{C91}$	NT	NT	NT	NT	NT	NT	$NspT2_{PC7120}$, $NspT1_{PC7120}$	–	–	–	–	3
Validated AG sequences	$EcoT1_{EDL933}$, $MavT1_{K10}$	NT	NT	NT	NT	NT	NT	NT	$SpyT1_{10270}$, $SpyT2_{10270}$, $BceT1_{E33L}$	$SmeT1_{1021}$, $SpyT1_{M1}$	–	$BceT5_{E33L}$	8

The Zeta super-family might contain a significant number of false-positive sequences that were selected on the basis of ATP-binding domain found in the 'original' Zeta sequence.
n represents the number of sequences; NT for not tested.
[a]Indicates that the YgjN toxin was validated during the course of this work (37).

are unrelated to the 'original' super-families. They were named Fiz for full toxin neutralization (Table 2).

Association between antitoxin and toxin super-families

In general, a large number of predicted toxin sequences are associated with antitoxin AG sequences (Figure 5). Three main categories can be defined: toxin super-families that have multiple partners (>4) such as the large ParE/RelE, VapC and CcdB/MazF super-families and the smaller GinA super-family (Figure 5A). The second category is composed of Doc, GinC and Zeta super-families. Sequences belonging to these super-families are associated with three different antitoxin super-families (Figure 5B). In the last category, toxin super-families are associated with AG sequences and a second super-family, Phd for the YafO super-family and HigA for the HipA, GinB and GinD super-families.

As observed for the predicted toxins, a large number of predicted antitoxins are associated with AG sequences (Figure 6). Categories similar to those defined for toxin super-families can be defined. Three super-families of antitoxin are promiscuous and associate with five or more toxin families i.e. the large PhD and HigA super-families (Figure 6A). The second category groups super-families that associated with three or four different toxin super-families such as the RelB and FizD super-families (Figure 6B). Most of the antitoxin super-families are quite restrictive in their associations (Figure 6C). They are associated with AG sequences and a second toxin super-family. Five of the novel super-families associate exclusively with a specific toxin super-family (FizB, FizF, FizI, FizJ and FizK) (Figure 6D).

Phyletic and genomic distribution of the toxin and antitoxin super-families

Phyletic distribution varies from one super-family to the other. The ParE/RelE, Zeta, VapC, Doc, CcdB/MazF, HipA and GinB super-families of toxins are widely distributed in bacterial phyla, although GinB is not detected in Firmicutes (Figure 7A). At least 70% of their members are detected in Proteobacteria, Firmicutes (except GinB), Cyanobacteria and Actinobacteria. The remaining sequences are detected in various phyla, with all of them except HipA and GinB being present in Archae. The GinA and VapD super-families, on the other hand, are restricted to a smaller number of phyla. Note that the VapD sequences are neither detected in Firmicutes nor in Archae. The GinC, YafO and GinD are specific to a single phylum. Most of the toxin sequences are located in chromosomes (Supplementary Figure S2A), except for the VapD sequences that are detected on plasmids (40%) and GinA sequences on phages (30%). The GinC, YafO and GinD sequences are detected neither on plasmids nor on phages.

As observed for the toxin super-families, phyletic distribution of the antitoxin super-families varies from one super-family to the other (Figure 7B). The HigA, FizA, FizB, FizC, Phd, VapB, PasA and RelB super-families appear widely distributed in bacterial phyla. At least 70 % of the members of these super-families are found in

Table 2. Occurrence of 20 antitoxin super-families in 2181 prokaryotic genomes

	Antitoxin superfamilies ($n = 10\,829$)									
	Phd ($n = 955$)	RelB ($n = 304$)	PasA ($n = 656$)	VapB ($n = 665$)	HigA ($n = 3351$)	VapX ($n = 10$)	CcdA ($n = 74$)	Epsilon ($n = 11$)	ParD ($n = 10$)	FizA ($n = 1962$)
'Original' sequences	Phd, YefM, Axe, PasB, StbD, YafN, RelB$_{307}$	RelB$_{K-12}$, DinJ, Paa1	PasA, YdcD, ParD$_{EDL933}$	VapB, MvpA, MazE, ChpBI, PemI	HigA, HipB	VapX	CcdA$_F$	Epsilon	ParD$_{RK2}$	–
Validated 'similar' sequences	NT	NT	NT	NT	YgjM[a] BceA5$_{E33L}$	NT	NT	NT	NT	–
Validated AGA sequences	NT	NT	NT	NT	SmeA1$_{1021}$ MvA1$_{K10}$	NT	NT	NT	NT	–
Validated AG sequences	NT	NT	SpeA2$_{TIGR4}$	SpeA3$_{TIGR4}$	EcoA1$_{EDL933}$	NT	NT	NT	NT	SpyA2$_{1070}$

	Antitoxin superfamilies ($n = 10\,829$)										Total
	FizB ($n = 1542$)	FizC ($n = 1121$)	FizD ($n = 103$)	FizE ($n = 26$)	FizF ($n = 10$)	FizG ($n = 9$)	FizH ($n = 8$)	FizI ($n = 6$)	FizJ ($n = 5$)	FizK ($n = 1$)	20
'Original' sequences	–	–	–	–	–	–	–	–	–	–	24
Validated 'similar' sequences	–	–	–	–	–	–	–	–	–	–	1
Validated AGA sequences	–	–	–	–	–	–	–	–	–	–	2
Validated AG sequences	CcrA1$_{CB15}$	SpyA1$_{M1}$	NeuA1$_{C91}$	AtuA1$_{C58}$	SpyA1$_{10270}$	NspA2$_{PC7120}$, NspA5$_{PC7120}$	AfuA2$_{DSM4304}$	CcrA4$_{CB15}$	BceA1$_{33L}$	NspA1$_{PC7120}$	15

The HigA super-family might contain a significant number of false-positive sequences that were selected on the basis of the HTH-XRE domain found in the 'original' HigA sequence.
n represents the number of sequences; NT for not tested.
[a]Indicates that the YgjM antitoxin was validated during the course of this work (37).

Proteobacteria, Firmicutes and Cyanobacteria (except for RelB sequences). The remaining 30% varies depending on the family, with members of HigA, FizA, Phd, VapB and PasA super-families being detected in Archae. The other super-families appear to be much more restricted. A majority of the FizG sequences are detected in Cyanobacteria. Sequences of FizE and FizF super-families are distributed in two phyla: Proteobacteria and Green sulfur bacteria. Members of the CcdA, ParD and FizI super-families are exclusively found in Proteobacteria, while those of VapX and Epsilon are restricted to Firmicutes and FizH to Archae. The only representative of FizK super-family is found in Cyanobacteria.

Regarding their genomic location (Supplementary Figure S2B), more than 85% of the sequences are located on chromosomes except for the RelB, FizF, CcdA and ParD super-families with 20–40% of the sequences being located on plasmids. The Epsilon super-family represents an exception with its 11 representative sequences being found only on plasmids.

Genome size and number of type II TA systems are not correlated

Figure 8 shows that there is no significant correlation between the number of predicted toxin and antitoxin sequences and the total number of ORFs predicted in a given genome. Even the smallest genomes of some of the intracellular obligate species seem to contain predicted toxin and antitoxin sequences (Table 3). While *Buchnera* and *Chlamydia/Chlamydophyla* genomes are devoid of predicted toxin and antitoxin sequences, some *Rickettsias* (such as *R. bellii* OSU 85-389, *R. bellii* RML369-C, *R. akari* or *R. felis*) contain more than 20 predicted sequences, representing between 1.8% and 2.6% of the total number of ORFs. Interestingly, the number of predicted toxins and antitoxins in the *Rickettsia* genus is quite variable (0–36). Variability is also found within the *Wolbachia* genus (0–6 predicted sequences). For the 20 bacterial isolates with the highest content of toxin and antitoxin predicted sequences, it varies from 37 to 97 per genome, representing from 0.8% to 2.5% with respect to the total number of ORFs (Supplementary Table S9). These bacteria belong to four different phyla (*Proteobacteria*, *Cyanobacteria*, *Green sulfur* and *Actinobacteria*). High genome plasticity appears to be a common theme for most of these strains (described for 12 over 20, e.g. *Microcystis aeruginosa*, *Gleobacter violaceum* and *Nitrosomanas europeae*). They also tend to have a versatile metabolism (6 over 20, e.g. *Nitrosomanas europeae*, *Rhodopseudomas palustris* and *Azoarcus* sp.) or a particular metabolism (4 over 20, e.g. *Geobacter uraniireducens*, *Pelodictyon phaeoclathratiforme* and *Caulobacter* sp.). Two isolates are symbionts (*Verminephrobacter eiseniae* and *Photorabdus luminescens*) and another is part of phototrophic bacterial consortia (*C. chlorochromatii*). Five of the strains belong to the
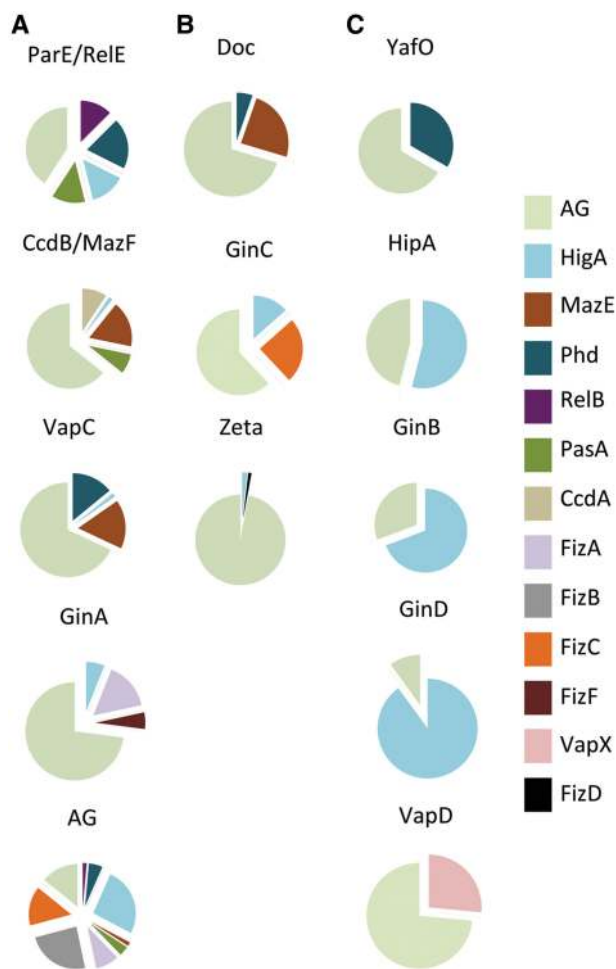
Figure 5. Associations of toxin super-families. The 12 toxin super-families are indicated above the pie chart. Each section of the pie chart represents the relative abundance of antitoxin sequences belonging to a given super-family associated with toxin sequences. Each antitoxin super-family is represented by a specific color. In (**A**), toxin super-families that are associated with multiple antitoxin super-families (>4). In (**B**), toxin super-families that are associated with three different antitoxin super-families. In (**C**), toxin super-families that are associated with AG sequences and another antitoxin super-family. Associations occurring at less than 1% were not considered for clarity.

Mycobacterium tuberculosis complex and are pathogenic (or deriving from a pathogenic strain in the case of *M. tuberculosis H37Ra*).

DISCUSSION

Discovering novel toxin and antitoxin sequences

The 'association by guilt' approach was already used to predict novel TA systems by Makarova *et al.* (24). The approach we used in this work allowed the identification of sequence pairs presenting unusual features such as a total absence of similarity with the 'original' sequences, a larger size for both the antitoxins and toxins (>200 amino acid) and larger intergenic distances (up to 300 bp). Comparison between their data and ours

indicates that our approach predicted their sequences (data not shown) but, more importantly, confirmed that there are many more toxins and antitoxin families not related to 'original' ones or simply uncharacterized yet, than previously thought and that remain to be experimentally studied. The successful experimental validation of some of our predictions is a direct proof of the results presented in the two works.

Activities of type II toxins

Quite unexpectedly, the novel toxins identified in this work are all translation inhibitors. We cannot exclude that a bias might have been introduced by the experimental validation of foreign genes in *E. coli*. Toxins targeting specific pathways might not be toxic for *E. coli* and therefore not validated by our Killing/rescue assay. The counter-argument is that if these genes are involved in stable maintenance of mobile genetic elements and move through horizontal transfer, toxins should target conserved mechanisms to ensure their function and maintenance in different bacterial hosts. 'Original' toxins inhibiting translation use a variety of mechanisms from cleavage of free-RNAs, cleavage of mRNAs during translation, inhibition of elongation by 30S binding and by EF-Tu phosphorylation (38–40), and it remains to be shown whether novel toxins will extend this list of activities. Mechanisms of action of these toxins are currently being investigated in our laboratory.

Origin and evolution of type II TA systems

Evolution of the TA systems remains largely unknown. While some authors proposed that these systems might derive from a common ancestor (41), other groups are in favor of several ancestors (10,12). Our data indicate that the validated toxins (both 'original' and those discovered in this analysis) constitute 12 unrelated super-families. No structural homologs could be detected for the novel toxins in the PDBsum structure database except for SpyT1$_{M1}$. The structure of its homolog is unrelated to type II toxins. This indicates that the novel toxins do not share any evolutionary relationship with 'original' super-families, although they are functionally related. However, only the resolution of their three-dimensional structures will unambiguously answer this question since similar structures have been detected for toxins that do not show significant sequence similarity (42). Subsequent divergence is observed for toxins within the ParE/RelE and CcdB/MazF super-families as they acquired different activities (DNA-gyrase inhibitors and mRNA-interferases).

For the antitoxins, our data indicate that they form 20 unrelated super-families. Further analyses will be required to define whether or not these antitoxins have a DNA-binding domain. Retracing antitoxin evolution might be complex since antitoxins are composed in general of two independent domains, an amino-terminal DNA-binding domain and a carboxy-terminal toxin-binding domain. DNA-binding domains might be recruited from or by other types of activities totally unrelated to antitoxins. A variation on the theme is observed in
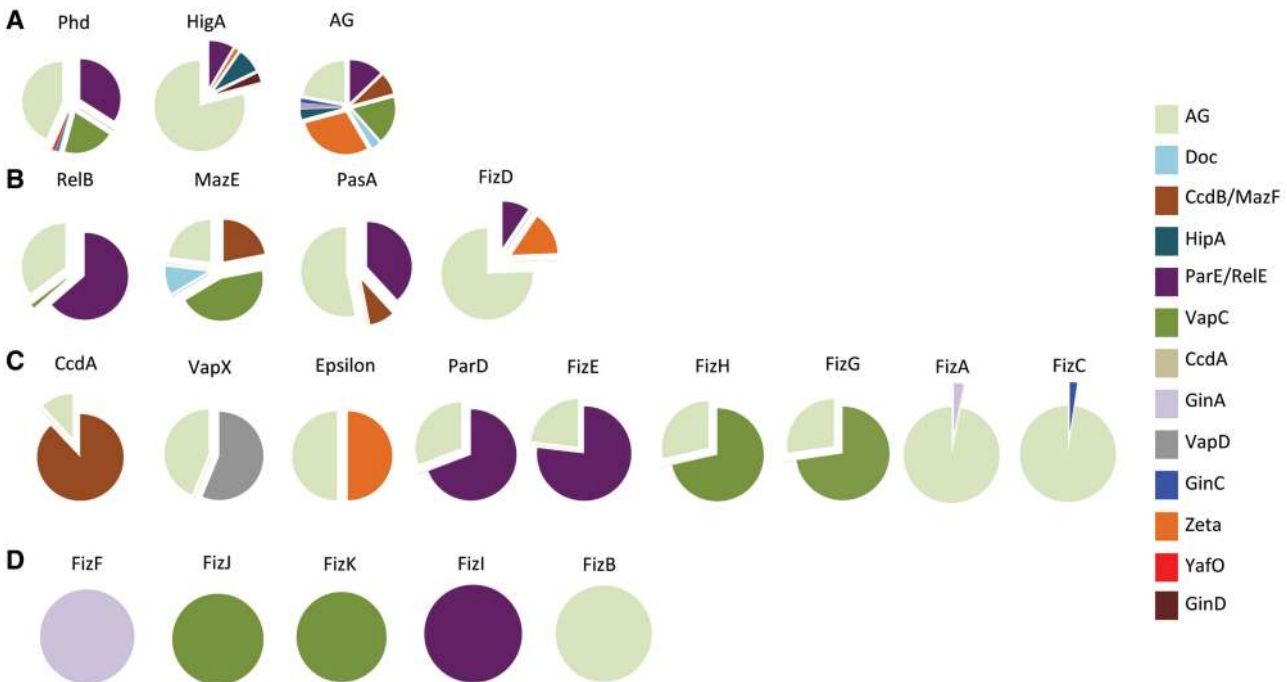
Figure 6. Associations of antitoxin super-families. The 20 antitoxin super-families are indicated above the pie chart. Each section of the pie chart represents the relative abundance of toxin sequences belonging to a given super-family associated with antitoxin sequences. Each toxin super-family is represented by a specific color. In (**A**), antitoxin super-families that are associated with multiple toxin super-families (five or more). In (**B**), antitoxin super-families that are associated with three or four different toxin super-families. In (**C**), antitoxin super-families that are associated with AG sequences and another toxin super-family. In (**D**), antitoxin super-families that are associated with a specific toxin super-family or to AG sequences. Associations occurring at <1% were not considered for clarity.
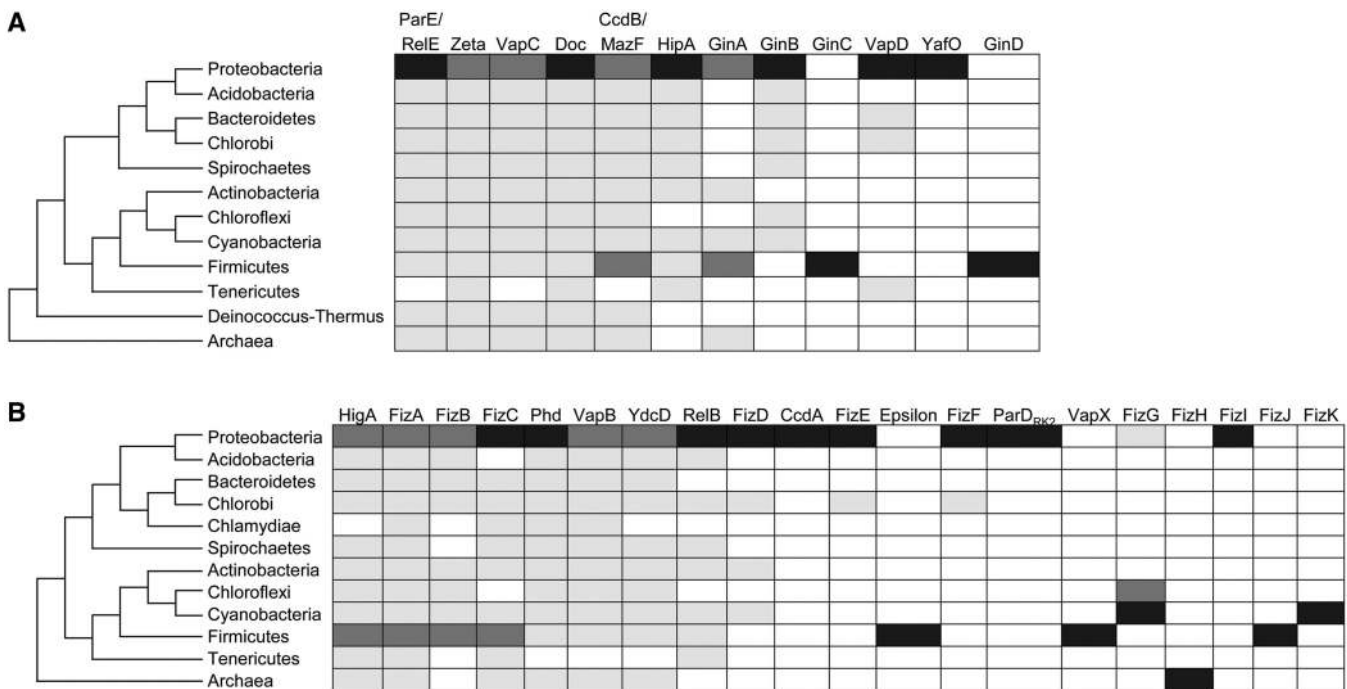

Figure 7. Phyletic distribution of toxin and antitoxin super-families. Sequences of toxin super-families (**A**) and antitoxin super-families (**B**) are detected in the bacterial phyla indicated at the left of the figure in different proportions: not detected (white), 0.1–20% (gray), 20.1–60% (dark gray), above 60.1% (black).

the case of three-component systems. In these systems [*epsilon-omega-zeta* (43) and *paaR1-paaA1-parE3* (33)], DNA-binding and antitoxin activities are encoded by two separated polypeptides. Molecular mechanisms underlying transcriptional regulation of three-component systems remain to be determined.
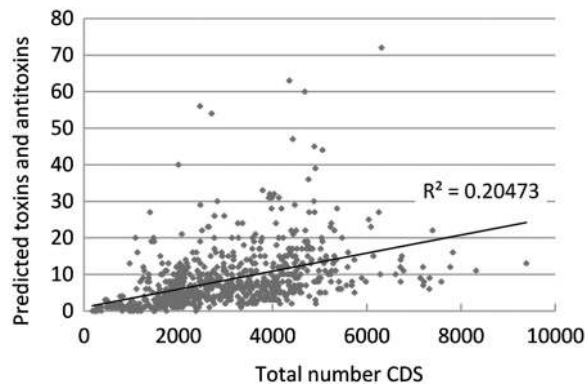

Figure 8. No correlation between the total number of CDS and the number of predicted antitoxin and toxin sequences. Only 'original' and 'similar' antitoxin and toxin sequences were considered.

Abundance and diversity

Another level of diversity is observed for antitoxin and toxin associations. Sequences of the RelE/ParE, CcdB/MazF and VapC super-families are paired with antitoxin sequences of the RelB, Phd, MazE and PasA super-families, supporting the idea that antitoxins and toxins were assembled multiple times as suggested previously (12). Sequences belonging to these super-families are abundant and present on plasmids. Other super-families are less promiscuous in their association. The reason is unclear, although super-families such as GinC, GinD, YafO, FizH or FizI are less represented and tend to be phyla- and chromosome-specific (although we cannot exclude that some are located on genomic islands), which might reduce their possibilities of association. However, although CcdA, Epsilon and ParD antitoxin sequences are present on plasmids, and therefore should be prone to horizontal gene transfer, they are phyla-specific and not abundant. Thus, what makes the evolutionary success of a given antitoxin or toxin super-family is still largely unknown. We have to keep in mind the biases introduced by the fact that a majority of the fully sequenced genomes available at NCBI are of Proteobacteria (745/1566 genomes) and that genomic islands within chromosomes remain difficult to identify.

Nevertheless, in general, these entities are abundant in bacterial chromosomes. Some genomes contain as much as 97 predicted antitoxin and toxin sequences (Supplementary Table S9), representing 1.5% with respect to the total number of CDSs. An even higher percentage is observed in some obligate intracellular bacterial genomes (Table 3). Predicted antitoxin and toxin sequences in *R. belii* and *R. felis* represent as much as 2.2% to 2.6%, respectively. Interestingly, these strains

Table 3. Predicted antitoxin and toxin sequences in intracellular obligate species

Bacterial species	Phyla	CDS in genome	Predicted A and T (%)
Buchnera aphidicola str. APS	Proteobacteria	564	0
Buchnera aphidicola str. Bp	Proteobacteria	504	0
Buchnera aphidicola str. Cc	Proteobacteria	357	0
Buchnera aphidicola str. Sg	Proteobacteria	546	0
Chlamydia muridarum Nigg	Chlamydiae	904	0
Chlamydia trachomatis 434/Bu	Chlamydiae	874	0
Chlamydia trachomatis A/HAR-13	Chlamydiae	911	0
Chlamydia trachomatis D/UW-3/CX	Chlamydiae	895	0
Chlamydia trachomatis L2b/UCH-1/proctitis	Chlamydiae	874	0
Chlamydophila abortus S26/3	Chlamydiae	932	0
Chlamydophila caviae GPIC	Chlamydiae	998	0
Chlamydophila felis Fe/C-56	Chlamydiae	1005	0
Chlamydophila pneumoniae AR39	Chlamydiae	1112	0
Chlamydophila pneumoniae CWL029	Chlamydiae	1052	0
Chlamydophila pneumoniae J138	Chlamydiae	1069	0
Chlamydophila pneumoniae TW-183	Chlamydiae	1113	0
Rickettsia akari str. Hartford	Proteobacteria	1259	23 (1.8)
Rickettsia bellii OSU 85-389	Proteobacteria	1476	32 (2.2)
Rickettsia bellii RML369-C	Proteobacteria	1429	26 (1.8)
Rickettsia canadensis str. McKiel	Proteobacteria	1093	1 (0.09)
Rickettsia conorii str. Malish 7	Proteobacteria	1374	13 (0.9)
Rickettsia felis URRWXCal2	Proteobacteria	1400	36 (2.6)
Rickettsia massiliae MTU5	Proteobacteria	968	16 (1.7)
Rickettsia prowazekii str. Madrid E	Proteobacteria	835	0
Rickettsia rickettsii str. 'Sheila Smith'	Proteobacteria	1345	14 (1.0)
Rickettsia rickettsii str. Iowa	Proteobacteria	1384	13 (0.9)
Rickettsia typhi str. Wilmington	Proteobacteria	838	0
Mycobacterium leprae TN	Actinobacteria	1605	1 (0.06)
Wigglesworthia glossinidia	Proteobacteria	611	0
Wolbachia endosymbiont of Drosophila melanogaster	Proteobacteria	1195	6 (0.5)
Wolbachia endosymbiont strain TRS of Brugia malayi	Proteobacteria	805	0
Wolbachia pipientis	Proteobacteria	1275	2 (0.2)

For each genome, the total number of CDS is indicated.
The total number of 'similar' and AGA sequences belonging to the super-families described in Supplementary Tables S2 and S3 were considered with the exception of those belonging to the Zeta and HigA super-families as many false-positive sequences are suspected.
The percentage of predicted sequences with respect to the total number of CDS in the genome is indicated between brackets.
A for antitoxin, T for toxin.

also contain a high number of repeat sequences (3.7% and 4.4%, respectively) and other genes indicative of genome plasticity (e.g. phage-related genes, transposases) (44,45). In contrast, other strains of the *Rickettsia* genus such as *R. typhi* and *R. prowezaki*, appear to be devoid of predicted antitoxin and toxin sequences (Table 3). Interestingly, both strains are closely related in the *Rickettsia* phylogeny and have undergone massive reductive evolution (a general phenomenon observed in obligate intracellular bacteria) as compared to *R. felis* and *R. bellii* (46). The absence of TA systems in some obligate intracellular bacteria might be due to massive gene loss. The comparison of *M. leprae* and its free-living relative *M. tuberculosis* complex strains supports this idea (one predicted sequence versus 43–45) (Table 3 and Supplementary Table S9). *A contrario*, a high content of TA systems might reflect an intense gene flux (Supplementary Table S9). Interestingly, *R. felis* carries a putative conjugative plasmid (44) and might still be

prone to horizontal gene transfer. This could possibly explain the presence of predicted toxin and antitoxin sequences in this strain.

As a conclusion, we propose that toxin and antitoxin genes have emerged and assembled multiple times during evolution. Associated with mobile genetic elements, TA systems promote their stability and propagate efficiently within the bacterial world. Because of their addictive properties, they integrate stably in bacterial chromosomes without providing necessarily selective advantages to their hosts. Interestingly, most of the toxin families are translation inhibitors, suggesting that this activity might have been selected rather than more detrimental ones such as DNA-gyrase inhibitors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Arcus,V.L., Rainey,P.B. and Turner,S.J. (2005) The PIN-domain toxin–antitoxin array in mycobacteria. *Trends Microbiol.*, **13**, 360–365.
2. Fozo,E.M., Hemm,M.R. and Storz,G. (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.*, **72**, 579–589.
3. 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**, 185–90.
4. Gerdes,K., Christensen,S.K. and Lobner-Olesen,A. (2005) Prokaryotic toxin–antitoxin stress response loci. *Nat. Rev. Microbiol.*, **3**, 371–382.
5. Yarmolinsky,M.B. (1995) Programmed cell death in bacterial populations. *Science*, **267**, 836–837.
6. Gerdes,K., Rasmussen,P.B. and Molin,S. (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl Acad. Sci. USA*, **83**, 3116–3120.
7. Cooper,T.F. and Heinemann,J.A. (2000) Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proc. Natl Acad. Sci. USA*, **97**, 12643–12648.
8. Jorgensen,M.G., Pandey,D.P., Jaskolska,M. and Gerdes,K. (2009) HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J. Bacteriol.*, **191**, 1191–1199.
9. Pandey,D.P. and Gerdes,K. (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.*, **33**, 966–976.
10. Guglielmini,J., Szpirer,C. and Milinkovitch,M.C. (2008) Automated discovery and phylogenetic analysis of new toxin–antitoxin systems. *BMC Microbiol.*, **8**, 104.
11. Ramage,H.R., Connolly,L.E. and Cox,J.S. (2009) Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin–antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet.*, **5**, e1000767.
12. Anantharaman,V. and Aravind,L. (2003) New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.*, **4**, R81.
13. Fozo,E.M., Makarova,K.S., Shabalina,S.A., Yutin,N., Koonin,E.V. and Storz,G. (2010) Abundance of type I toxin–antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res.*, **38**, 3743–3759.
14. Koonin,E.V. and Wolf,Y.I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.*, **36**, 6688–6719.
15. Mine,N., Guglielmini,J., Wilbaux,M. and Van Melderen,L. (2009) The decay of the chromosomally encoded ccdO157 toxin–antitoxin system in the *Escherichia coli* species. *Genetics*, **181**, 1557–1566.
16. Magnuson,R.D. (2007) Hypothetical functions of toxin–antitoxin systems. *J. Bacteriol.*, **189**, 6089–6092.
17. Van Melderen,L. and Saavedra De Bast,M. (2009) Bacterial toxin–antitoxin systems: more than selfish entities? *PLoS Genet.*, **5**, e1000437.
18. Van Melderen,L. (2011) Toxin–antitoxin systems: why so many, what for? *Curr. Opin. Microbiol.*, **13**, 781–785.
19. Szekeres,S., Dauti,M., Wilde,C., Mazel,D. and Rowe-Magnus,D.A. (2007) Chromosomal toxin–antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol. Microbiol.*, **63**, 1588–1605.
20. Wozniak,R.A. and Waldor,M.K. (2009) A toxin–antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.*, **5**, e1000439.
21. Grady,R. and Hayes,F. (2003) Axe-Txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium. Mol. Microbiol.*, **47**, 1419–1432.
22. Schmidt,O., Schuenemann,V.J., Hand,N.J., Silhavy,T.J., Martin,J., Lupas,A.N. and Djuranovic,S. (2007) prlF and yhaV encode a new toxin–antitoxin system in *Escherichia coli. J. Mol. Biol.*, **372**, 894–905.
23. Makarova,K.S., Grishin,N.V. and Koonin,E.V. (2006) The HicAB cassette, a putative novel, RNA-targeting toxin–antitoxin system in archaea and bacteria. *Bioinformatics*, **22**, 2581–2584.
24. 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.
25. Casadaban,M.J. and Cohen,S.N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli. J. Mol. Biol.*, **138**, 179–207.
26. 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.
27. Vanderpool,C.K. and Gottesman,S. (2007) The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *J. Bacteriol.*, **189**, 2238–2248.
28. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.

29. Christensen,S.K., Maenhaut-Michel,G., Mine,N., Gottesman,S., Gerdes,K. and Van Melderen,L. (2004) Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the yefM-yoeB toxin-antitoxin system. *Mol. Microbiol.*, **51**, 1705–1717.

30. Brosius,J. and Holy,A. (1984) Regulation of ribosomal RNA promoters with a synthetic lac operator. *Proc. Natl Acad. Sci. USA*, **81**, 6929–6933.

31. Wilbaux,M., Mine,N., Guerout,A.M., Mazel,D. and Van Melderen,L. (2007) Functional interactions between coexisting toxin-antitoxin systems of the ccd family in *Escherichia coli* O157:H7. *J. Bacteriol.*, **189**, 2712–2719.

32. Daines,D.A., Jarisch,J. and Smith,A.L. (2004) Identification and characterization of a nontypeable *Haemophilus influenzae* putative toxin–antitoxin locus. *BMC Microbiol.*, **4**, 30.

33. Hallez,R., Geeraerts,D., Sterckx,Y., Mine,N., Loris,R. and Van Melderen,L. (2010) New toxins homologous to ParE belonging to three-component toxin–antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.*, **76**, 719–732.

34. Singletary,L.A., Gibson,J.L., Tanner,E.J., McKenzie,G.J., Lee,P.L., Gonzalez,C. and Rosenberg,S.M. (2009) An SOS-regulated type 2 toxin–antitoxin system. *J. Bacteriol.*, **191**, 7456–7465.

35. Nieto,C., Pellicer,T., Balsa,D., Christensen,S.K., Gerdes,K. and Espinosa,M. (2006) The chromosomal relBE2 toxin-antitoxin locus of *Streptococcus pneumoniae*: characterization and use of a bioluminescence resonance energy transfer assay to detect toxin-antitoxin interaction. *Mol. Microbiol.*, **59**, 1280–1296.

36. Fiebig,A., Castro Rojas,C.M., Siegal-Gaskins,D. and Crosson,S. (2010) Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin–antitoxin systems. *Mol. Microbiol.*, **77**, 236–251.

37. Christensen-Dalsgaard,M., Jorgensen,M.G. and Gerdes,K. (2010) Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol. Microbiol.*, **75**, 333–348.

38. Schumacher,M.A., Piro,K.M., Xu,W., Hansen,S., Lewis,K. and Brennan,R.G. (2009) Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. *Science*, **323**, 396–401.

39. Liu,M., Zhang,Y., Inouye,M. and Woychik,N.A. (2008) Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proc. Natl Acad. Sci. USA*, **105**, 5885–5890.

40. Condon,C. (2006) Shutdown decay of mRNA. *Mol. Microbiol.*, **61**, 573–583.

41. Gerdes,K. (2000) Toxin–antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bacteriol.*, **182**, 561–572.

42. Brown,B.L., Grigoriu,S., Kim,Y., Arruda,J.M., Davenport,A., Wood,T.K., Peti,W. and Page,R. (2009) Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathog.*, **5**, e1000706.

43. de la Hoz,A.B., Ayora,S., Sitkiewicz,I., Fernandez,S., Pankiewicz,R., Alonso,J.C. and Ceglowski,P. (2000) Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proc. Natl Acad. Sci. USA*, **97**, 728–733.

44. Ogata,H., La Scola,B., Audic,S., Renesto,P., Blanc,G., Robert,C., Fournier,P.E., Claverie,J.M. and Raoult,D. (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.*, **2**, e76.

45. Ogata,H., Renesto,P., Audic,S., Robert,C., Blanc,G., Fournier,P.E., Parinello,H., Claverie,J.M. and Raoult,D. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.*, **3**, e248.

46. Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.M. *et al.* (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii. Science*, **293**, 2093–2098.