Comparative genomics evidence that only protein toxins are tagging bad bugs

Kalliopi Georgiades and Didier Raoult *

Unité de Recherche en Maladies Infectieuses Tropical Emergentes (URMITE), CNRS–IRD UMR 6236-198, Université de la Méditerranée, Marseille, France

INTRODUCTION

The term toxin was introduced by Roux and Yersin and describes macromolecular substances that, when produced during infection or when introduced parenterally or orally, cause an impairment of physiological functions that lead to disease or to the death of the infected organism. Long after the discovery of toxins, early genetic studies on bacterial virulence demonstrated that removing a certain number of genes from pathogenic bacteria decreases their capacity to infect hosts. Each of the removed factors was therefore referred to as a “virulence factor,” and it was speculated that non-pathogenic bacteria lack such supplementary factors. However, many recent comparative studies demonstrate that the specialization of bacteria to eukaryotic hosts is associated with massive gene loss. We recently demonstrated that the only features that seem to characterize 12 epidemic bacteria are toxin–antitoxin (TA) modules, which are addiction molecules in host bacteria. In this study, we investigated if protein toxins are indeed the only molecules specific to pathogenic bacteria by comparing 14 epidemic bacterial killers (“bad bugs”) with their 14 closest non-epidemic relatives (“controls”). We found protein toxins in significantly more elevated numbers in all of the “bad bugs.” For the first time, statistical principal components analysis, including genome size, GC%, TA modules, restriction enzymes, and toxins, revealed that toxins are the only proteins other than TA modules that are correlated with the pathogenic character of bacteria. Moreover, intracellular toxins appear to be more correlated with the pathogenic character of bacteria than secreted toxins. In conclusion, we hypothesize that the only truly identifiable phenomena, witnessing the convergent evolution of the most pathogenic bacteria for humans are the loss of metabolic activities, i.e., the outcome of the loss of regulatory and transcription factors and the presence of protein toxins, alone, or coupled as TA modules.

Keywords: protein toxins, toxin–antitoxin modules, bad bugs, restriction enzymes, comparative genomics
In this study, we investigated if toxins are indeed the only molecules specific to pathogenic bacteria. We supplemented the 12 “bad bugs” from our previous study with two additional “bad bugs” (Neisseria meningitidis and Staphylococcus aureus) and compared all 14 with their 14 closest non-epidemic relatives (“controls”). These 14 “bad bugs” are the most dangerous epidemic bacteria of all times and we chose to limit our study to them in order to achieve a neutral and not biased approach.

**MATERIALS AND METHODS**

The following “bad bugs” were used: Mycobacterium leprae TN (NC_002677), Mycobacterium tuberculosis H37Rv (NC_000962), Rickettsia prowazekii Madrid E (NC_000963), Corynebacterium diphtheriae NCTC 13129 (NC_002935), Treponema pallidum SS14 (NC_010741), Yersinia pestis KIM (NC_004088), Bordetella pertussis Tohama I (NC_002929), Streptococcus pneumoniae G54 (NC_011072), Streptococcus pyogenes M1 GAS (NC_002737), Salmonella typhi CT18 (NC_003198), Shigella dysenteriae Sd197 (NC_007606), Vibrio cholerae O395 (NC_009457), Staphylococcus aureus MC58 (NC_003112). For the “controls,” we constructed a 16s rRNA phylogenetic tree for each group of species and the non-pathogenic, or at least, non-epidemic species found as the closest relative to the “bad bug” was chosen as “control.” The following 14 related bacterial species were used: Mycobacterium avium 104 (NC_008595), Mycobacterium smegmatis MC2 155 (NC_008596), Rickettsia africae ESF-5 (NC_012633), Corynebacterium glutamicum R (NC_009342), Yersinia pseudotuberculosis IP 3295 (NC_006155), Bordetella bronchiseptica RB50 (NC_002927), Streptococcus agalactiae 2603V/R (NC_004116), Streptococcus suis 05ZYH33 (NC_009442), Salmonella Schwarzengrund CVM19633 (NC_011094), Escherichia coli HS (NC_004316), Vibrio parahaemolyticus RIMD 2210633 (NC_004603), Staphylococcus haemolyticus JCSC1435 (NC_007168), and Neisseria cinerea (NZ_ACDY00000000).

**GENOMIC CHARACTERISTICS**

All genomic characteristics used in this study [i.e., genome size, GC%, and number of open reading frames (ORFs)] were retrieved from the NCBI database. A graph was plotted and a χ² statistical test was performed for each feature to determine if there were statistically significant differences between the “bad bugs” and the “controls.”

**RESTRICTION ENZYMES**

Restriction enzymes for each of the 28 bacterial species were retrieved from the REBASE database (Roberts et al., 2010). A graph was designed, and a χ² statistical test was performed to conclude whether there were statistically significant differences between the “bad bugs” and the “controls.”

**TA MODULES**

Text-mining searches were conducted in the GenBank protein database for the following seven type II TA families: VapB/C, RelE/B, ParE/D, MazE/F, phd/doc, ccdA/B, and higA/B. Each protein was used in a tBLASTN query, and hits were defined based...
on an e-value threshold of 10e-5 with >30% identity and at least 70% coverage. A graph was designed and a χ² statistical test was performed to conclude whether there were statistically significant differences between the “bad bugs” and the “controls.”

**PROTEIN TOXINS**
Protein toxins for each of the 28 bacterial species were retrieved from the MvirDB database (Zhou et al., 2007). A graph was designed, and a χ² statistical test was performed to conclude whether there were statistically significant differences between the “bad bugs” and the “controls.”

**PHYLOGENIES**
Phylogenomic trees were constructed for all restriction enzymes, protein toxins, and all toxins and TA modules by generating a matrix of binary discrete characters (“0” and “1” for absence and presence, respectively) and implementing the neighbor-joining (NJ) method in Phylogeny Inference Package (PHYLIP; Felsenstein, 1993). Phylogenetic trees were also separately constructed for each restriction enzyme family gene using three methods: NJ, maximum-parsimony (MP), and maximum-likelihood (ML). Alignments were performed with ClustalX2 (Larkin et al., 2007), and trees were constructed using Mega4 (Tamura et al., 2007).

**ORIGINS OF THE TOXINS AND TA MODULES**
A literature analysis allowed us to determine which of the toxins were transported to the bacterium by various mobile elements. Additionally, using phylogenies, we were able to determine whether toxins and TA modules were acquired by horizontal gene transfer (HGT) for each of the bacteria in our study. A tBLASTN query was performed for each toxin, and hits were defined based on an e-value threshold of 10e-5 with >30% identity and at least 70% coverage. Alignments were performed with ClustalX2 (Larkin et al., 2007), and trees were constructed using Mega4 (Tamura et al., 2007) via three methods: NJ, MP, and ML.

**PRINCIPAL COMPONENTS ANALYSES**
A first principal components analysis (PCA) was performed, including genome size, GC%, number of ORFs, protein toxins, restriction enzymes, and TA modules. The second PCA included protein toxins and TA modules, taking into consideration the lifestyle of the toxins (intracellular or extracellular). The PCAs were performed using R software (http://www.r-project.org/version 2.11.0).

**RESULTS**
Confirming our previous study (Georgiades and Raoult, 2011b), our current results reveal significant differences in genome size and the number of ORFs between the “bad bugs” and “controls.” Indeed, “bad bugs” have smaller genomes and less ORFs than “controls” (Table 1; Figures A1A,B in Appendix). However, while most “bad bugs” have a lower GC%, the differences are not statistically significant (Table 1; Figure A1C in Appendix).

Furthermore, only five bad bugs have more restriction enzymes than their controls (M. tuberculosis, C. diphtheriae, S. pneumoniae, S. typhi, and N. meningitidis), and the differences are not statistically significant (p = 0.2857; Table 1; Figure A1D and Table A1 in Appendix). The phylogenomic tree for restriction enzymes displays a small cluster containing six species that do not encode restriction enzymes, of which four are “bad bugs” (Y. pestis, B. pertussis, T. pallidum, and R. prowazekii; Figure A2 in Appendix). The phylogenetic tree for type I restriction enzymes displays a cluster containing five “controls” (S. suis, V. parahaemolyticus, M. smegmatis, E. coli HS, and T. denticola) and only one “bad bug” (S. pneumoniae). The rest of the trees generally present topologies that are different from the topologies given by a 16S rRNA phylogenetic tree, with closely related species not clustering together. This result holds true for all of the methods used in this study (NJ, MP, and ML; Figure A3 in Appendix).

Concerning the protein toxins, all of the “bad bugs” contain more proteins, and the difference is statistically significant (p = 0.0002; Table 1; Figure A1E and Table A2 in Appendix). The phylogenomic tree of all toxins includes a cluster containing 12 species, of which nine are “controls” missing toxins (T. denticola, S. haemolyticus, V. parahaemolyticus, E. coli HS, S. schwarzenberg, C. glutamicum, R. africae, M. smegmatis, and M. avium). Moreover, a cluster of seven “bad bugs” containing toxins is also present. “Bad bugs” or “controls” tend to cluster together and not with their phylogenetically closest relatives (Figure A4 in Appendix). Finally, we searched the 28 genomes for members of the seven known TA modules and found that eight “bad bugs” (M. tuberculosis, Y. pestis, S. pneumoniae, S. pyogenes, S. typhi, S. dysenteriae, V. cholerae, and N. meningitidis) contain significantly more TA modules than their controls (p = 0.0445; Tables 1 and 2, Figure A1F in Appendix). Our results agree with the findings of previous studies (Pandey and Gerdes, 2005; Goulard et al., 2016; Georgiades and Raoult, 2011b) and the data in the toxin–antitoxin database (TADB; Shao et al., 2011).

We constructed a phylogenomic tree based on the presence/absence of toxins and TA modules that presents three clusters. One of these clusters contains seven “controls” with no toxins or TA modules (M. avium, R. africae, C. glutamicum, S. schwarzenberg, E. coli, V. parahaemolyticus, and S. haemolyticus), whereas “bad bugs” that encode many toxic elements are grouped together (M. tuberculosis, Y. pestis, S. pneumoniae, S. pyogenes, S. typhi, S. dysenteriae, V. cholerae, S. aureus, and N. meningitidis; Figure 2).

It has been hypothesized and widely accepted that bacterial toxins are encoded by mobile and variable elements (Novick, 2003). Surveying the literature, we reviewed the origins of some of these protein toxins (Table 3). Most bacterial toxins, including Cholera toxin, Streptococcal pyrogenic exotoxin (SpeC), Shiga (and Shiga-like) toxins, Diphtheria toxin, most staphylococcal and pathogenic E. coli Enterotoxins, and the Leukocidin toxin of S. aureus, are phase-associated, encoded on bacteriophages and inserted into the bacterial chromosomes (Waldor and Mekalanos, 1996; Nataro and Kaper, 1998; Boudy et al., 2001; Muniesca et al., 2003; Novick, 2003; Tinsley et al., 2006). Similarly, other mobile elements encode bacterial toxins, such as transposons for some E. coli enterotoxins (Nataro and Kaper, 1998), plasmids that carry staphylococcal enterotoxins and exfoliatins (Novick, 2003), and staphylococcal pathogenicity islands (SaPIs) carrying enterotoxins B, C, and K, and the S. aureus toxin shock syndrome toxin (TSST-1; Novick, 2003).
We also searched for possible HGT events, which would be located in the genomic origins of toxin and TA module sequences present in the genomes of the bacteria of interest. We did not find strong evidence for such events, except in three cases: cholera toxin seems to have been gained by *V. cholerae* from *E. coli*, and Zeta toxin of *N. cinerea*, and Shiga toxin also grouped with *E. coli* strains (the commensal strain O8 IAI1 and strain O103, respectively). As for TA modules, they appear to have different origins; HigA/B were likely transferred by *Firmicutes*, MazE/F likely came from different proteobacteria, RelB/E may have been transferred by *Gammaproteobacteria*, and VapB/C were most likely transferred from *Betaproteobacteria* (Figure A5 in Appendix).

Our initial PCA revealed that other than TA modules, only toxins appear to be correlated with pathogenicity (Figure 3). “Bad bugs” and “controls” are separated on the analysis plot when these two features are taken into consideration. Furthermore, as discovered in the second PCA, when we considered the lifestyles of the toxins (i.e., whether they are intracellular or extracellular), the intracellular toxins, including TA modules, are most related to the pathogenic capacity of the bacteria studied and not the extracellular toxins, as would be expected (Figure 4). Other than TA modules, which are intracellular toxins, the toxins of *Salmonella*, *Shigella*, *V. cholerae*, *Yersinia*, and *B. pertussis* are also intracellular. This is why their pathogenic capacities were so difficult to understand and why the TA modules were never considered important for bacterial pathogenesis.

**DISCUSSION**

In this study, we investigated whether restriction enzymes and protein toxins are the only features, other than TA modules (Georgiades and Raoult, 2011b), that can be considered natural virulence factors for bacteria. To maintain an unbiased analysis, we assessed the 14 most dangerous epidemic bacteria for humans and compared them to their closest non-epidemic related species.

In a previous study, we demonstrated that 12 epidemic bacteria have reduced-size genomes and more selfish elements, i.e., TA modules, which have evolved such that the host cells become addicted to them (Lawrence, 1999b; Makarova et al., 2009; Georgiades and Raoult, 2011b). Other studies also report a high number of TA modules in pathogenic bacteria, such as *Y. pestis* (Goulard et al., 2010), but their role in pathogenesis is not evident and was never considered previously. TA modules are mostly encoded on plasmids or within prophages and were initially identified as plasmid stabilization factors. When they were also found in multiple copies on chromosomes, it was hypothesized that they might also have a role in the stabilization of integrons in bacterial chromosomes (Szekeres et al., 2007). Other systems that possess properties of selfish elements are restriction enzymes,
### Table 2 | Number of TA modules in each of the TA families. “Bad bugs” are in red and “controls” in blue.

| TA families | VapB/C | RelB/E | ParE/D | MazE/F | phd/doc | ccdA/B | higA/B | Unclassified | Σ     |
|-------------|--------|--------|--------|--------|---------|--------|--------|--------------|-------|
| Species     |        |        |        |        |         |        |        |              |       |
| M. leprae   | –      | –      | –      | –      | –       | –      | –      | –            | 0     |
| M. avium    | –      | –      | –      | –      | –       | –      | –      | –            | 1     |
| M. tuberculosis* | 38    | –      | –      | –      | –       | –      | –      | –            | 38    |
| M. smegmatis | 1     | –      | –      | 1      | –       | –      | –      | 2            | 4     |
| R. prowazekii | –    | –      | –      | –      | –       | –      | –      | –            | 0     |
| R. africae  | 2      | –      | –      | 4      | 2       | –      | 2      | –            | 10    |
| C. diphtheriae | –   | –      | –      | –      | –       | –      | –      | –            | 0     |
| C. glutamicum | –    | –      | –      | –      | 1       | –      | –      | –            | 1     |
| T. pallidum | –      | –      | –      | –      | –       | –      | –      | –            | 0     |
| T. denticola | 1     | –      | –      | –      | –       | –      | –      | –            | 1     |
| Y. pestis*  | –      | –      | –      | –      | –       | –      | 2      | 3            | 5     |
| Y. pseudotuberculosis | – | –      | –      | –      | –       | –      | –      | –            | 0     |
| B. pertussis | –     | –      | –      | –      | –       | –      | –      | –            | 0     |
| B. bronchiseptica | – | –      | –      | –      | –       | –      | –      | –            | 0     |
| C. pneumoniae* | – | –      | –      | –      | –       | –      | –      | –            | 0     |
| S. agalactiae | –    | 1      | 1      | –      | –       | –      | –      | –            | 2     |
| S. pyogenes* | –     | –      | 1      | 1      | –       | –      | –      | –            | 2     |
| S. suis     | –      | –      | –      | –      | –       | –      | –      | –            | 0     |
| S. typhi*  | 2      | 2      | –      | –      | –       | –      | 2      | –            | 6     |
| S. schwarzengrund | – | –      | –      | –      | –       | –      | –      | 1            | 1     |
| S. dysenteriae | –   | –      | –      | 2      | –       | –      | –      | –            | 2     |
| E. coli    | –      | –      | –      | –      | 1       | –      | –      | –            | 1     |
| V. cholerae | –      | 2      | 4      | –      | –       | –      | 1      | 6            | 13    |
| V. parahaemolyticus | – | 2      | 2      | –      | –       | –      | –      | –            | 4     |
| S. aureus  | –      | 1      | –      | 2      | –       | –      | –      | –            | 3     |
| S. haemolyticus | –  | –      | –      | 1      | –       | –      | –      | 2            | 3     |
| N. meningitidis | 1 | 1      | –      | 2      | 1       | –      | –      | 1            | 6     |
| N. cinerea | –      | –      | –      | –      | –       | –      | 2      | –            | 2     |

* Asterisk shows bad bugs with more TA than their controls.

which may also constitute a toxic danger for bacterial host cells. It has been demonstrated that restriction enzymes can be toxic for mammalian cells by promoting DNA mutations (Kinashi et al., 1993; Price et al., 1995). However, our study did not reveal a significant difference in the numbers of restriction enzymes in “bad bugs.”

Since Diphtheria toxin was isolated by Roux and Yersin in 1888, bacterial toxins have been recognized as the primary virulence factors for pathogenic bacteria. These toxins are defined as “soluble substances that alter the normal metabolism of host cells with deleterious effects on the host,” are considered as one of the most powerful human poisons known, and can retain high activity when very dilute. Indeed, the major symptoms associated with diseases caused by *C. diphtheriae*, *B. pertussis*, *V. cholerae*, *Clostridium anthrax*, *Clostridium botulinum*, and enterohemorrhagic *E. coli* are all related to the activities of the toxins produced by these organisms (Alouf, 2000; Merrell and Falkow, 2004). Thus, we hypothesize that toxins have a direct measurable effect on cells. Injecting toxins in an animal or a cell leads to its death, whereas administration of antibodies against the toxin in question offers protection against future infections. This is what constitutes the principle of vaccinations against Diphtheria and Tetanus and the principle of preventive passive immunization (Stiehm, 1998). Recent experimental and cellular models testing different individual genes from bacterial genomic repertoires demonstrate that some glycoproteins have similar toxic roles (Merrell and Falkow, 2004). These endotoxins may be released in a soluble form by young cultures grown in the laboratory; however they act in a way that does not reflect bacterial virulence in experimental models (Wesselingh et al., 1978). For the most part, endotoxins remain associated with the cell wall until disintegration of the organisms. In vivo, this is the result of autolysis, external lysis mediated by complement and lysozyme, and phagocytic digestion of bacterial cells. Both the toxic component of LPS and the immunogenic portion of LPS act as determinants of pathogenicity for bacteria (Merrell and Falkow, 2004). Furthermore, bacterial pathogenicity in humans is also associated with the epidemic capacity of the bacteria, about which there are few hypotheses, except in the case of vector-borne diseases (Gubler, 1997), whose multiplication in the blood and levels of bacterial load are critical. Such a considerable multiplication causes deleterious effects to the host that may lead to its death, and events such as high multiplication results in a de-regulation associated
with the disappearance of transcription regulators observed in pathogenic species (Merhej et al., 2009; Georgiades and Raoult, 2011b). This likely causes the persistence of multiplication in the blood, even if nutrients decrease, and the appearance of a less favorable condition that breaks the balance with the host. In our study, we found significantly elevated numbers of protein toxins in all of the “bad bugs,” while most of the “controls” only possess two or less toxins. The statistical PCA, including genome size, GC%, TA modules, restriction enzymes, and toxins, revealed that other than TA modules, toxins are the only proteins linked to the pathogenicity of bacteria (Figure 3). Moreover, intracellular toxins appear to be more related to the pathogenicity of bacteria than secreted toxins (Figure 4). Of course, this hypothesis needs to be confirmed by experimentations and possible mechanisms should be proposed. Phylogenomic trees, constructed based on the presence/absence of restriction enzyme genes, toxins, and TA modules, display different topologies compared to 16S rRNA phylogenetic trees, meaning that closely related species are not found in sister taxa. “Bad bugs” or “controls” tend to cluster together rather than with their phylogenetic neighbors. This demonstrates that specialized epidemic bacteria have congruent evolutionary histories, resulting in a virulent gene repertoire defined by both present

Table 3 | Toxins and variable genetic elements.

| Toxin          | Organism         | Element               |
|----------------|------------------|-----------------------|
| Diphtheria toxin | C. diphtheriae  | Phage                 |
| Cholera toxin   | V. cholerae      | Phage                 |
| Shiga toxin     | S. dysenteriae  | Phage                 |
| Enterotoxins    | E. coli          | Phage, plasmids, transposons |
| TSST-1          | S. aureus        | Pathogenicity island  |
| Enterotoxin A   | S. aureus        | Phage                 |
| Enterotoxin D   | S. aureus        | Plasmid               |
| Enterotoxins B, C, K | S. aureus     | Pathogenicity island  |
| Exfoliatin      | S. aureus        | Phage, plasmid        |
| Leukocidin      | S. aureus        | Phage                 |
| Sea enterotoxin A | S. aureus     | Phage                 |
| SpeC            | S. pyogenes      | Phage                 |

FIGURE 2 | Phylogenomic tree based on presence/absence of toxin-antitoxin modules and toxins. Closely related species do not cluster together while totally unrelated species do, meaning that highly divergent species may have common evolutionary histories. “Bad bugs” are in red, and “controls” are in blue.

FIGURE 3 | Principal components analysis. TA modules and toxins characterize epidemic bacteria. “Bad bugs” are in red, and “controls” are in blue.
Only protein toxins are tagging bad bugs

FIGURE 4 | Principal components analysis. Intracellular toxins and TA modules characterize epidemic bacteria. “Bad bugs” are in red, and “controls” are in blue.

FIGURE 5 | The “bad bug” creation scenario. Initially, bacteria contain metabolic functions and recombination machinery. TA modules are gained by gene transfer, and toxin genes arrive in bacterial genomes by various mobile elements, such as plasmids. After specialization, the bacterial recombination systems are degraded, and metabolic functions are lost. “Bad bugs” are characterized by toxins and TA modules that stabilize neighboring genes and limit massive gene loss.

and absent genes. The gene repertoires of closely related species can possess different histories, while distant species can have a similar gene repertoire due to related evolutionary events. In particular, TA modules are found in “bad bugs” as a result of HGT events from other bacteria, including Firmicutes, Gammaproteobacteria, Betaproteobacteria, and other proteobacteria (Figure A5 in
Appendix). Furthermore, most toxins arrive in bacterial genomes by various mobile elements, especially bacteriophages. This is the case for the streptococcal and staphylococcal toxins, Cholera toxin, Shiga toxin, and *E. coli* Enterotoxins. Modifications of bacterial virulence by bacteriophages were initially brought to light in 1951 (Freeman, 1951). This first demonstration showed that virulent strains of *C. diphtheriae* infected with a bacteriophage yielded virulent lysogens producing Diphtheria toxin (Freeman, 1951). The outbreak of bloody diarrhea and hemolytic uremic syndrome (HUS) due to an *E. coli* O104:H4 strain in Germany in May and June 2011 illustrates the capacity of bacterial species to produce new combinations of genes, leading to the emergence of highly aggressive strains. The O104:H4 strain acquired the phage-mediated Shiga toxin and resistance to numerous antibiotics. However, antibiotic selective pressure has nothing to do with specialization (Werner et al., 2004; Furuya and Lowy, 2006). As in the HUS case, antibiotics are not always indicated in the treatment of the human disease because they may worsen the symptoms when the toxin is released (Denamur, 2011).

In conclusion, the only truly identifiable phenomena, witnessing a convergent evolution in the most pathogenic bacteria for humans, are the loss of metabolic activities (Georgiades and Raoult, 2011b) that occur via the loss of regulatory and transcription factors and the acquisition of protein toxins, alone, or coupled as TA modules (Figure 5). For example, in the case of *Shigella*, the loss of the lysine decarboxylase activity is a major event in the acquisition of its virulence capacity (Maurelli et al., 1998). TA modules stabilize and protect neighboring genes, which explains the fact that massive gene loss in some specialized bacteria seems to be decreased, while in some others (e.g., *R. prowazekii*) that do not contain TA modules, gene loss is much more extreme and on-going.

**REFERENCES**

Alouf, E. J. (2000). Bacterial toxins: methods and protocols. *Methods Mol. Biol.* 145, 1–26.

Andersson, J. O., and Andersson, S. G. E. (1999). Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* 16, 1178–1191.

Andersson, S. G. E., and Kurland, C. G. (2011). *Autobiography*. [online]. Available at: <http://nobelprize.org/nobel_prizes/medicine/laureates/1978-arber-auto-biography.html>

Audic, S., Robert, C., Campagna, B., Parinello, H., Claverie, J. M., Raoult, D., and Drancourt, M. (2007). Genome analysis of *Minibacterium massiliense* highlights the convergent evolution of water-living bacteria. *PLoS Genet.* 3, e138. doi: 10.1371/journal.pgen.0030138

Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestrís, G., Claverie, J. M., and Raoult, D. (2007). Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.* 3, e14. doi: 10.1371/journal.pgen.0030014

Boudy, T. B., Pancholi, V., and Fischetti, V. A. (2001). Induction of lysogenic bacteriophage and phage-associated toxin from Group A Streptococci during co-culture with human pharyngeal cells. *Infect. Immun.* 69, 1440–1443.

Bourniquel, A. A., and Bickle, T. A. (2002). Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 84, 1074–1059.

Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J., and Andersson, S. G. E. (2007). Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* 23, 511–520.

Denamur, E. (2011). The 2011 Shiga toxin-producing *Escherichia coli* O104:H4 German outbreak: a lesson of genomic plasticity. *Clin. Microbiol. Infect.* doi: 10.1111/j.1469-0691.2011.03620.x

Denamur, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 214–242.

Felsenstein, J. (1993). *PHYLIPI (Phylogeny Inference Package)*. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Fournier, P. E., Karkouri, El K., Leroy, Q., Robert, C., Giunelli, B., Renesto, P., Socolovschi, C., Parola, P., Audic, S., and Raoult, D. (2009). Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. * BMC Genomics* 10, 166. doi: 10.1186/1471-2164-10-166

Fuzzo, 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.

Freeman, V. J. (1951). Studies on the virulence of bacteriophages-infected strains of *Corynebacterium diphtheriae*. *J. Bacteriol.* 61, 675–688.

Furuya, E. Y., and Lowy, F. D. (2006). Antimicrobial-resistance bacteria in the community setting. *Nat. Rev. Microbiol.* 4, 36–45.

Georgiades, K., and Raoult, D. (2011a). Defining pathogenic bacterial species in the genomic era. *Front Microbiol.* 1:151. doi: 10.3389/fmicb.2010.00151

Georgiades, K., and Raoult, D. (2011b). Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS ONE* 6, e17962. doi: 10.1371/journal.pone.0017962

Goulard, C., Langrand, S., Carniel, E., and Chauvaux, S. (2010). The Yersinia pestis chromosome encodes active addiction toxins. *J. Bacteriol.* 192, 3669–3677.

Gubler, D. J. (1997). Resurgent vector-borne diseases as a global health problem. *Emerging Infect. Dis.* 4, 442–450.

Jensen, R. B., and Gerdes, K. (1995). Programmed cell death in bacteria: prokaryotic plasmid stabilization systems. *Mol. Microbiol.* 17, 205–210.

Kinnamon, P., Nagasama, H., and Little, B. J. (1993). Mutagenic effects of restriction enzymes in Chinese hamster cells: evidence for high mutagenic potential of Sau3AI and the hpr locus. *Mutat. Res.* 285, 251–257.

Kristoffersen, P., Jensen, G. B., Gerdes, K., and Piskur, J. (2000). Bacterial *antisense* enzymes: NTP-driven molecular motors. *Biochimie* 82, 365–375.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.

Lawrence, J. (1999a). Transfer of regulatory factors and the acquisition of protein toxins, alone, or coupled as TA modules during co-culture with human pharyngeal cells. *Infect. Immun.* 69, 1440–1443.

Merhej, V., and Raoult, D. (2010). Rickettsial evolution in the light of comparative genomics. *Biol. Rev.* 85, 379–405.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586.

Muniesa, M., de Simon, M., Prats, G., Ferrer, D., Panella, H., and Jofre, J. (2003). Shiga toxin 2-converting bacteriophage associated with clonal viability in *Escherichia coli* O157:H7 strains of human origin isolated from a single outbreak. *Infect. Immun.* 71, 4554–4562.

Nataro, J. P., and Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201.

Frontiers in Cellular and Infection Microbiology www.frontiersin.org October 2011 | Volume 1 | Article 7 | 8
Pandey, P. D., and Gerdes, K. (2005). Ochman, H., Lerat, E., and Daubin, V. (2003). Mobile genetic elements and bacterial toxins: the superantigens-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* 49, 93–105.

Ochman, H., Lerat, E., and Daubin, V. (2003). Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6595–6599.

Price, R. C., Margison, G. P., Hendry, J. H., and West, C. M. (1995). Relationships between the cytotoxic effects of restriction endonucleases and radiation on mammalian cells. *Int. J. Radiat. Biol.* 67, 327–334.

Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2010). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 38, 234–236.

Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H. Y., Rajakumar, K., and Deng, Z. (2011). TADRE: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and Archaea. *Nucleic Acids Res.* 39, D606–D611.

Stiehm, E. R. (1998). “Passive immunization,” in *Textbook of Paediatric Infectious Diseases*, 4th Edn, eds R. D. Feigin and J. D. Cherry (Philadelphia, PA: The WB Sanders Co.), 2769–2802.

Szekeres, S., Dauti, M., Wüde, 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.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.

Tinsley, C. R., Bille, E., and Nasif, X. (2006). Bacteriophages and pathogenicity: more than just providing a toxin? *Microbes Infect.* 8, 1365–1371.

Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914.

Werner, C. A., Monnet, L. D., and Harbath, S. (2004). Antibiotic selection pressure and resistance in *Streptococcus pneumoniae* and *Streptococcus pyogenes*. *Emerging Infect. Dis.* 10, 514–517.

Wesselink, P. R., Thoden van Velzen, S. K., and Makkes, P. C. (1978). Release of endotoxins in an experimental model simulating the dental root canal. *Oral Surg. Oral Med. Oral Pathol.* 45, 789–795.

Yamamoto, T. A., Gerdes, K., and Tunnacliffe, A. (2002). Bacterial toxin RelE induces apoptosis in human cells. *FEBS Lett.* 519, 191–194.

Zhou, C. E., Smith, J., Lam, M., Zemla, A., Dyer, M. D., and Slezak, T. (2007). MvrDR—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.* 35, 391–394.
## APPENDIX

### Table A1 | Restriction enzymes.

| Bacteria            | Restriction enzymes | Total |
|---------------------|---------------------|-------|
| M. leprae           | 1 Homing endonuclease | 1     |
| M. avium            | 2 Type I restriction enzyme | 4     |
|                     | 1 Type II restriction enzyme |     |
|                     | 1 Type III restriction enzyme |     |
| M. tuberculosis     | 2 Homing endonuclease | 2     |
| M. smegmatis        | 1 Type I restriction enzyme | 1     |
| R. prowazekii       | –                   | –     |
| R. africae          | 1 Type I restriction enzyme | 1     |
| C. diphtheriae      | 1 Type I restriction enzyme | 4     |
|                     | 1 Type II restriction enzyme |     |
|                     | 2 Type III restriction enzyme |     |
| C. glutamicum       | –                   | –     |
| T. pallidum         | –                   | –     |
| T. denticola        | 2 Type I restriction enzyme | 4     |
|                     | 2 Type II restriction enzyme |     |
| Y. pestis           | –                   | –     |
| Y. pseudotuberculosis | 2 Type I restriction enzyme | 2     |
| B. pertussis        | –                   | –     |
| B. bronchiseptica   | 1 Type II restriction enzyme | 2     |
|                     | 1 Type III restriction enzyme |     |
| S. pneumoniae       | 4 Type I restriction enzyme | 8     |
|                     | 4 Type II restriction enzyme |     |
| S. agalactiae       | –                   | –     |
| S. pyogenes         | 1 Type I restriction enzyme | 1     |
| S. suis             | 3 Type I restriction enzyme | 4     |
|                     | 1 Type III restriction enzyme |     |
| S. typhi            | 2 Type I restriction enzyme | 3     |
|                     | 1 Type III restriction enzyme |     |
| S. schwarzengrund   | 1 Type II restriction enzyme | 2     |
|                     | 1 Type III restriction enzyme |     |
| S. dysenteriae      | 1 Type II restriction enzyme | 1     |
| E. coli             | 2 Type II restriction enzyme | 2     |
| V. cholerae         | 1 Type I restriction enzyme | 1     |
| V. parahaemolyticus | 1 Type I restriction enzyme | 1     |
| S. aureus           | 1 Type I restriction enzyme | 1     |
| S. haemolyticus     | 2 Type I restriction enzyme | 2     |
|                     | 1 Type II restriction enzyme |     |
| N. meningitidis     | 1 Type I restriction enzyme | 7     |
|                     | 4 Type II restriction enzyme |     |
|                     | 2 Type III restriction enzyme |     |
| N. cinerea          | 2 Type I restriction enzyme | 2     |
## Table A2 | Protein toxins.

| Species                  | ID       | Protein Toxins                                                                 |
|--------------------------|----------|-------------------------------------------------------------------------------|
| **M. leprae**             | O3307    | Probable secreted protease                                                    |
| **M. avium**              | –        |                                                                               |
| **M. tuberculosis**       | O05458   | Probable alanine and proline rich membrane-anchored mycosin                   |
|                           | O05461   | Membrane anchored mycp1                                                       |
|                           | O53945   | Probable proline rich membrane-anchored mycosin mycp5                        |
|                           | O53695   | Probable membrane anchored mycosin mycp3                                     |
| **M. smegmatis**          | –        |                                                                               |
| **R. prowazekii**         | –        |                                                                               |
| **R. africae**            | –        |                                                                               |
| **C. diphtheriae**        | NP_938615| Diphtheriae toxin precursor                                                   |
|                           | Q6NK15   | Diphtheria toxin                                                              |
| **C. glutamicum**         | –        |                                                                               |
| **T. pallidum**           | –        |                                                                               |
| **T. denticola**          | –        |                                                                               |
| **Y. pestis**             | –        | YopQ/yopK                                                                     |
|                           |          | YopJ/yopP                                                                     |
|                           |          | YopH; ypkA                                                                    |
|                           |          | yopT; yopM                                                                    |
|                           |          | Rho Gap                                                                       |
|                           |          | Toxin complex subunit TcaA                                                    |
|                           |          | Toxin complex subunit TcaB                                                    |
|                           |          | Toxin complex subunit TcaC                                                    |
|                           | P31493   | Outer membrane virulence protein yopE                                         |
|                           | P17811   | Coagulase/fibrolysin                                                          |
| **Y. pseudotuberculosis**| P08008   | YopE membrane virulence outer protein                                         |
|                           | Q05608   | YpkA kinase                                                                    |
|                           | B2BCZ0   | Toxin complex subunit TcaA                                                    |
|                           | B2BCZ1   | Toxin complex subunit TcaB                                                    |
|                           | B2BCY3   | Toxin complex subunit TcaC                                                    |
|                           | VFDB/VF0393 | Cytotoxic necrotizing factor                                            |
| **B. pertussis**          | POAR5    | Pertussin toxin subunit 4 AND 32 more toxins *****                            |
| **B. bronchiseptica**     | –        |                                                                               |
| **S. pneumoniae**         | O85254   | Pneumolysin                                                                    |
|                           | P0C2J9   | Thiol-activated cyclolysin                                                     |
| **S. agalactiae**         | Q53637   | SCPB                                                                          |
| **S. pyogenes**           | MV1889   | Sea Antitoxin A                                                                |
|                           | CAA45934 | Zeta                                                                          |
|                           | P0A4L0   | Thiol-activated cytolysin                                                      |
|                           | P0C0I3   | Thiol-activated cytolysin                                                      |
|                           | Q57211   | Exotoxin typB/streptococcus peptidase A AND 16 others ***                     |
| **S. suis**               | O85102   | Hemolysin                                                                      |
|                           | Q55996   | Hemolysin (suilysin)                                                           |
| **S. typhi**              | Q9RP6M   | SopE                                                                          |
|                           | STM2878  | SptP                                                                          |
|                           | STM2884  | SipC                                                                          |
|                           | CAA57991 | SipA                                                                          |
|                           | T1477    | Hemolytic toxin                                                                |
|                           | STY3008  | SipB                                                                          |
|                           | P061NO   | Invasion protein invB                                                          |
|                           | P35671   | Invasion protein invE                                                          |
|                           | P06185   | Outer membrane protease E                                                      |

(Continued)
Table A2 | Continued

| Species                        | ID           | Gene name                                                                 |
|--------------------------------|--------------|---------------------------------------------------------------------------|
|                                | Q54044       | Antigen presentation protein spaN                                         |
|                                | Q56019       | Cell invasion protein sipB                                                 |
|                                | Q56020       | Cell invasion protein sipC                                                 |
|                                | Q56052       | Invasion invE                                                              |
|                                | Q56134;       | sipB; sipC, spaN                                                           |
|                                | Q56135;      |                                                                            |
|                                | P40613       |                                                                            |
| S. schwarzenberg               |              |                                                                            |
| S. dysenteriae                | AAF28121     | Shiga toxin subunit A                                                      |
|                                | AAF28122     | Shiga toxin subunit B                                                      |
| E. coli                       |              |                                                                            |
| V. cholera                     | VC1451       | Cholera toxin                                                              |
|                                | NP_230476    | Toxin co-regulated pilin                                                   |
|                                | NP_230477    | Toxin co-regulated pilus biosynthesis protein B                           |
|                                | NP_230479    | Toxin co-regulated pilus outer membrane protein                           |
|                                | NP_230481    | Toxin co-regulated pilus biosynthesis protein D                           |
|                                | NP_230484    | Toxin co-regulated pilus biosynthesis protein E                           |
|                                | NP_230485    | Toxin co-regulated pilus biosynthesis protein F                           |
|                                | NP_230475    | Toxin co-regulated pilus biosynthesis protein H                           |
|                                | NP_230473    | Toxin co-regulated pilus biosynthesis protein I                           |
|                                | NP_230474    | Toxin co-regulated pilus biosynthesis protein P                           |
|                                | NP_230478    | Toxin co-regulated pilus biosynthesis protein Q                           |
| V. parahaemolyticus            |              |                                                                            |
| S. aureus                      | BAB47174     | iukNS                                                                      |
|                                | MW1889       | sea enterotoxin A                                                          |
|                                | PO6886       | Tst                                                                        |
|                                | SACOL2022    | Hld                                                                        |
|                                | Q53691       | HlgC-like                                                                  |
|                                | Q57227       | HlgB-like                                                                  |
|                                | AAA17490     | Exfoliatin toxin                                                           |
|                                | AAA26617     | Enterotoxin D                                                              |
|                                | AAB06195     | Enterotoxin E                                                              |
|                                | NP_646706    | Staphylococcal enterotoxin A                                               |
|                                |              | 3 Staphylococcal enterotoxin C3                                            |
|                                |              | Staphylococcal enterotoxin C1                                              |
|                                |              | Staphylococcal enterotoxin SeG                                             |
|                                |              | Staphylococcal enterotoxin SeK                                             |
|                                |              | Pyrogenic exotoxin                                                          |
|                                |              | 16 Enterotoxins                                                            |
|                                |              | Toxin shock syndrome toxin (TSST-1)                                        |
|                                |              | Alpha-toxin                                                                |
|                                |              | Leukocidin                                                                 |
| S. haemolyticus                |              |                                                                            |
| N. meningitidis                | 15677759     | Biosynthesis glycosyltransferase                                           |
|                                | 15677756     | Biosynthesis glycosyltransferase                                           |
|                                | 15677562     | Beta-1,4-glycosyltransferase                                               |
|                                | 15677969     | Lipopolysaccharide heptosyltransferase                                     |
|                                | 15677379     | ADP-heptose-LPS-heptosyltransferase II                                     |
|                                | 15677553     | Alpha-1,2-N-acetylglicosamine transferase Lipooligosaccharide LOS          |
| N. cinerea                     | ZP_05983406.1| Zeta-toxin                                                                 |
|                                | ZP_05983573.1| 2 Zonula occludens toxin family                                            |
|                                | ZP_05983608.1|                                                                            |
|                                | ptxA ptxS1   | ADP ribosylase                                                             |
|                                | ptxB ptxS2   | Toxin entry                                                                 |

(Continued)
**Table A2 | Continued**

| Species | Protein Toxins |
|---------|----------------|
|         | ID              | Gene name                                                      |
|         | BP3787          | Toxin entry                                                   |
|         | BP3785          | Toxin entry                                                   |
|         | BP3786          | Toxin entry                                                   |
|         | P11091          | CyaD                                                         |
|         | Q956M9          | Petracin                                                     |
|         | Q956N0          | Petracin                                                     |
|         | Q956N1          | Petracin                                                     |
|         | NP_879578       | Bifunctional hemolysin-adenylate cyclase precursor            |
|         | NP_879579       | Cyclolysin secretion ATP-binding protein                     |
|         | NP_879577       | Cyclolysin activating lysine-acetyltransferase                |
|         | P0A3R5          | Pertussis toxin subunit 4                                    |
|         | P04981          | Pertussis toxin subunit 5                                    |
|         | NP_881965       | Dermonecrotic toxin                                          |
|         | NP_882287       | Pertussis toxin transport protein                             |
|         | NP_882288       | Pertussis toxin transport protein                             |
|         | NP_882282       | Pertussis toxin subunit 1 precursor                           |
|         | NP_882283       | Pertussis toxin subunit 2 precursor                           |
|         | NP_882286       | Pertussis toxin subunit 3 precursor                           |
|         | NP_882284       | Pertussis toxin subunit 4 precursor                           |
|         | NP_882285       | Pertussis toxin subunit 5 precursor                           |
|         | P04977          | Pertussis toxin subunit 1                                    |
|         | QPS3M8          | Pertactin                                                    |
|         | P15318          | Bifunctional hemolysin                                       |
|         | Q66135          | Tracheal colonization factor                                  |
|         | P14283          | Petracin auto transporter                                    |
|         | Q88143          | Petracin                                                     |
|         | Q95917          | Beta-lactamase                                               |
|         | O69257          | Petracin                                                     |
|         | O69259          | Petracin                                                     |
|         | P04979          | Pertussis toxin subunit 3                                    |
|         | MV1889          | Sea Antitoxin A                                              |
|         | CAA45934        | Zeta                                                         |
|         | P0A4L0          | Thiol-activated cytolysin                                    |
|         | P0C0I3          | Thiol-activated cytolysin                                    |
|         | Q57211          | Exotoxin typB/streptococcus peptidase A                      |
|         | NP_268546       | Streptolysin O                                               |
|         | NP_268959       | Mitogenic exotoxin Z                                         |
|         | NP_268105       | Exotoxin type A                                              |
|         | NP_268947       | Pyrogenic exotoxin C                                         |
|         | NP_268582       | Pyrogenic exotoxin G                                         |
|         | NP_269186       | Streptococcal exotoxin H                                    |
|         | NP_269185       | Streptococcal exotoxin                                       |
|         | NP_268735       | Putative exotoxin                                            |
|         | NP_265009       | Streptococcal pyrogenic exotoxin SpeK                        |
|         | NP_607349       | Exotoxin SpeL                                               |
|         | NP_607380       | Exotoxin SpeM                                               |
|         | NP_604724       | Streptococcal superantigen SSA-phage associated              |
|         | P0C0I5          | Exotoxin type C                                             |
|         | Q8NKX2          | Exotoxin type C                                             |
|         | Q54738          | SSA                                                         |
|         | Q54739          | SSA                                                         |

* *Bordetella pertussis toxins.
**Streptococcus pyogenes toxins.*
FIGURE A1 | Graphical representations of the genomic characteristics of “bad bugs” and “controls.” (A) Genome sizes in kilobases; (B) numbers of ORFs; (C) numbers of restriction enzymes; (D) GC%; (E) numbers of toxin-antitoxin modules; (F) numbers of protein toxins. All colored dots represent “bad bug”/“control” couples.
Only protein toxins are tagging bad bugs

FIGURE A2 | Phylogenomic tree for restriction enzymes.

FIGURE A3 | Phylogenetic trees for restriction enzymes.
Phylogenomic tree of ALL protein toxins

Bordetella pertussis
Staphylococcus aureus
Salmonella Typhi
Streptococcus pyogenes
Vibrio cholerae
Yersinia pestis
Neisseria meningitidis
Yersinia pseudotuberculosis
Mycobacterium tuberculosis
Shigella dysenteriae
Bordetella bronchiseptica
Rickettsia prowazekii
Corynebacterium diphtheriae
Neisseria cinerea
Streptococcus pneumoniae
Treponema pallidum
Mycobacterium leprae
Streptococcus suis
Streptococcus agalactiae
Treponema denticola
Staphylococcus haemolyticus
Vibrio parahaemolyticus
Escherichia coli HS
Salmonella Schwarzengrund
Corynebacterium glutamicum
Rickettsia africae
Mycobacterium smegmatis
Mycobacterium avium

FIGURE A4 | Phylogenomic tree for toxins.
Only protein toxins are tagging bad bugs

**Cholera toxin**

**Zeta toxin**

**Shiga toxin**

FIGURE A5 | Continued
FIGURE A5 | Phylogenetic trees for HGT events.

vapB /C

Deinococci/Betaproteobacteria/
Gammaproteobacteria/
Actinobacteria/Cyanobacteria/
Terenicutes

beta

alpha