SCIENTIFIC REPORTS

OPEN

Genomic analysis of *Acinetobacter baumannii* prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness

Ana Rita Costa, Rodrigo Monteiro ⓘ & Joana Azeredo

The recent nomination by the World Health Organization of *Acinetobacter baumannii* as the number one priority pathogen for the development of new antibiotics is a direct consequence of its fast evolution of pathogenicity, and in particular of multidrug resistance. While the development of new antibiotics is critical, understanding the mechanisms behind the crescent bacterial antibiotic resistance is equally relevant. Often, resistance and other bacterial virulence elements are contained on highly mobile pieces of DNA that can easily spread to other bacteria. Prophages are one of the mediators of this form of gene transfer, and have been frequently found in bacterial genomes, often offering advantageous features to the host. Here we assess the contribution of prophages for the evolution of *A. baumannii* pathogenicity. We found prophages to be notably diverse and widely disseminated in *A. baumannii* genomes. Also remarkably, *A. baumannii* prophages encode for multiple putative virulence factors that may be implicated in the bacterium's capacity to colonize host niches, evade the host immune system, subsist in unfavorable environments, and tolerate antibiotics. Overall our results point towards a significant contribution of prophages for the dissemination and evolution of pathogenicity in *A. baumannii*, and highlight their clinical relevance.

Acinetobacter baumannii was recently indicated by the World Health Organization (WHO) as the number one priority pathogen for research and development of new antibiotics (http://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/). This human opportunistic pathogen has been gradually evolving towards clinical success since the 1970s, due to an increasing overall pathogenicity mostly related to a growing multidrug resistance.

Genomically, *A. baumannii* is characterized by a relatively small core genome and a large and diversified accessory genome[1]. This indicates gene acquisition and loss as important contributors to *A. baumannii* evolution and adaptation towards pathogenicity. For example, genes associated with antibiotic resistance have been found in both core and accessory genomes of *A. baumannii*[1]. In the accessory genome, these genes were found often flanked by integrases, transposases, or insertion sequences, suggesting a possible acquisition by horizontal gene transfer (HGT) from other strains or bacterial species. HGT may thus be a major force in the evolution of *A. baumannii* pathogenicity.

Among mediators of HGT we find bacteriophages (phages), viruses of bacteria thought to be the most abundant biological entities on Earth[2]. When infecting a bacterial host, phages may follow distinct life cycles: virulent phages follow a lytic path in which they replicate inside the bacteria and cause cell lysis for progeny release; temperate phages may also follow the lytic cycle or opt for a lysogenic cycle where they integrate into the host genome and replicate passively with the bacterial genome. When integrated in the bacterial genome, temperate phages are known as prophages.

CEB – Centre of Biological Engineering, University of Minho, 4710-057, Braga, Portugal. Ana Rita Costa and Rodrigo Monteiro contributed equally. Correspondence and requests for materials should be addressed to J.A. (email: jazeredo@deb.uminho.pt)

1
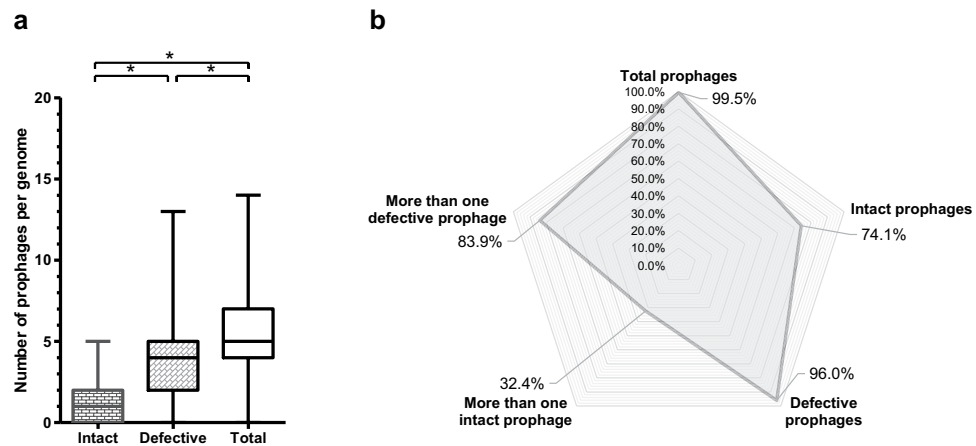
Figure 1. Prevalence of prophages in *Acinetobacter baumannii* genomes. (**a**) Whiskers plot of prophage frequency per bacterial genome. The horizontal line at the center of the whiskers plot represents the median. The bottom and top of the plot represent the first and third quartiles. The external edges of the whiskers represent the minimum and maximum number of prophages per genome. Significant differences (Tukey's test) of $P < 0.05$ are represented by*. (**b**) Prevalence of total prophages, "intact prophages", defective prophages, more than one "intact prophage", and more than one defective prophage. Prevalence was determined considering a dataset of 795 *A. baumannii* genomes.

Prophages and their bacterial hosts have partly aligned evolutionary interests, since proliferation of the host results in increased prophage population[3]. This is possibly the reason why some prophages provide the host bacterium beneficial traits such as protection from infection by other phages (superinfection exclusion)[4,5], increased host fitness[6], and encoding of virulence factors (VF) exploited for bacterial pathogenesis (e.g. antibiotic tolerance[7] or toxins[8]).

Under certain stimuli, prophages can excise from the host genome, entering the lytic cycle with the release of phage progeny. During excision, a process of specialized transduction may occur, where parts of the bacterial genome adjacent to the prophage may be erroneously excised with the prophage genome and introduced with the virion into a new host[9]. Temperate phages thus contribute to host evolution by a constant transfer of genes between host genomes[9,10]. Still, prophage genes are under selection for rapid deletion from bacterial genomes, with studies suggesting that most prophages in bacterial genomes are to some extent defective[11,12]. Even so, defective prophages can lead to bacterial evolution, with a few bacterial molecular systems thought to derive from the process of prophage inactivation, e.g. gene transfer agents[13], bacteriocins, and type VI secretion systems (T6SS)[14,15].

The number of prophages in bacterial genomes and their contribution to bacterial evolution differ among species. Here we aimed at evaluating the prevalence of prophages in *A. baumannii* genomes, and at understanding the contribution of these elements to the rapid evolution of pathogenicity in this bacterial pathogen.

Results

Prevalence of "intact" and defective prophages in *A. baumannii* strains.

We analyzed 795 genomes of *A. baumannii* of a total of 1,614 genomes deposited on GenBank at the date of March 2016. Selection of genomes was random, but restricted to bacteria sequenced by Illumina with a coverage above 45x. Prophages were identified using PHAST and manually curated. A total of 4,122 prophages were found, of which 943 were "intact" and 3,179 were defective. The significantly higher prevalence of defective prophages (Fig. 1a) was expected since "intact prophages" are usually under strong selection by bacteria for mutations causing prophage inactivation[3]. Still, 74.1% of the *A. baumannii* strains contained "intact prophages" (Fig. 1b), suggesting a recent integration in the bacterial genome. To note that the analysis here performed included a large number of draft genomes (97.9%), which implies that PHAST may under-estimate the number of intact prophages (these may be split into different contigs) and over-estimate the number of defective prophages (intact prophages split into different contigs may be identified as several defective prophages). We compared the numbers of both intact and defective prophages obtained for draft (778) and complete (17) genomes, which indicate averages of both intact and defective prophages significantly higher in draft genomes (5.2 ± 2.4 total prophages, 1.2 ± 1.0 "intact prophages", 4.0 ± 2.3 defective prophages) than in complete genomes (2.9 ± 1.0 total prophages, 0.6 ± 0.7 "intact prophages", 2.4 ± 1.1 defective prophages). The differences however may be just a consequence of the small representation of complete genomes in our analysis, and not a problem related to PHAST analysis of draft genomes; if that were the case, a low number of "intact prophages" was expected in draft genomes.

Distribution of prophages by bacterial genome size.

Like Touchon *et al.*[16] and Bobay *et al.*[17] did for other bacterial species[16,17], we question if *A. baumannii* with larger genomes will allow for the integration of more prophages. The existence of more neutral targets for phage integration in larger bacterial genomes may facilitate prophage accumulation without interference with the vital functions of the bacteria[17]. To evaluate this hypothesis,
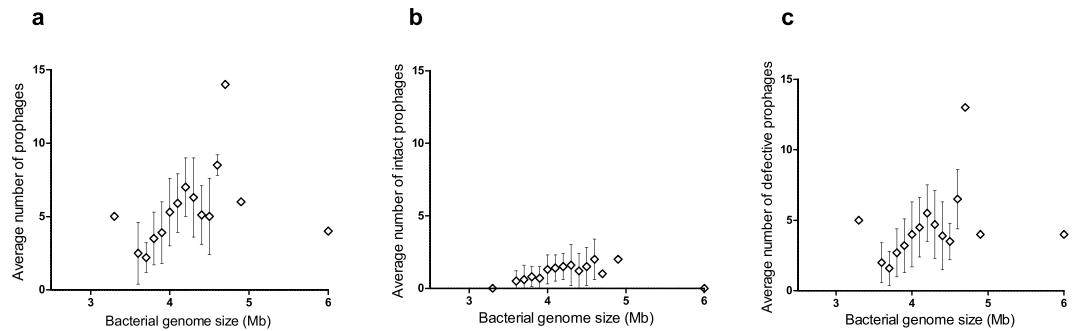
Figure 2. Distribution of prophages among *Acinetobacter baumannii* strains considering the size of the bacterial genome. (**a**) Average number of prophages ("intact" and defective) per bacterial genome size. (**b**) Average number of "intact prophages" per bacterial genome size. (**c**) Average number of defective prophages per bacterial genome size.
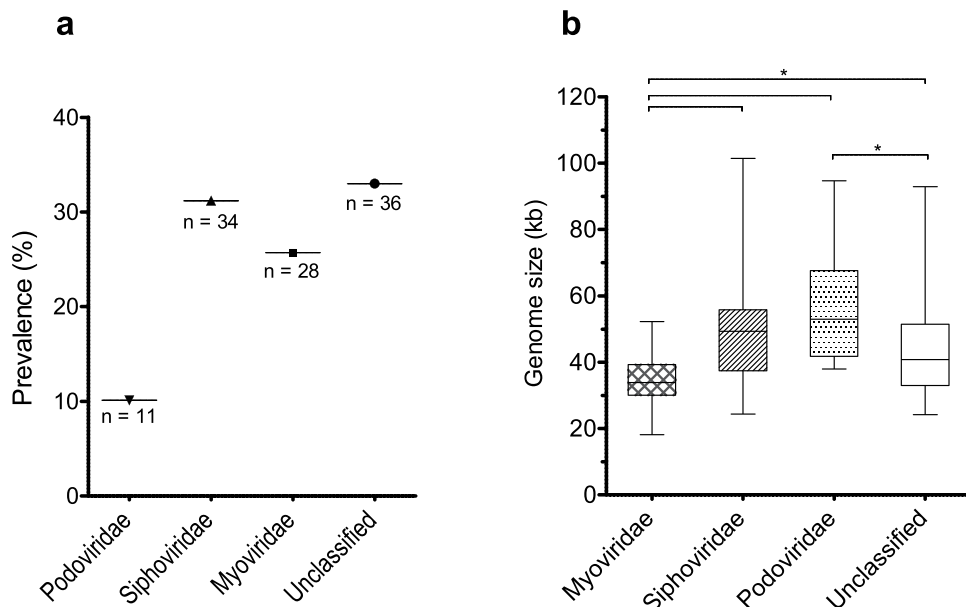
Figure 3. Distribution of genome size of the 109 "intact prophages" integrating *Acinetobacter baumannii* genomes. (**a**) Prevalence of prophages in *A. baumannii* genomes by family; (**b**) Whiskers plot of average genome size of prophages according to family. Horizontal line at the center represents the median, bottom and top of the plot represent the first and third quartiles, and external edges of the whiskers represent the minimum and maximum genome size of prophages per family. Significant differences (Tukey's test) of $P < 0.05$ are represented by*.

we determined the distribution of prophages considering the size of the *A. baumannii* genomes (Fig. 2). A tendency of larger bacteria to harbor increased numbers of prophages can be observed, although the number of prophages appears to stabilize for larger (>4.2 Mbp) genomes (for statistical analysis refer to Supplementary Table S1).

Distribution of prophages by family and genome size. We have classed the 109 "intact prophages" in family taxa based on homology to known classed phages. Classification relied on genes considered the most indicative of family: major capsid protein, large terminase subunit, tail tape measure protein and tail sheath protein. Approximately 67% of the prophages could be assigned a family, with the majority identified as *Siphoviridae*, followed by *Myoviridae* and *Podoviridae* (Fig. 3a). This is in accordance with the estimated distribution in nature[18]. On the contrary, the average genome size per prophage family goes against the trends described in the literature (http://viralzone.expasy.org/) (Fig. 3b). *Myoviridae* are typically the largest phages, sizing between 33 to 244 kb. However, here *Myoviridae* have the smallest genomes (34 kb, $P < 0.001$) of the prophages with predicted family. Moreover, *Siphoviridae* in *A. baumannii* have the widest size range (24–101 kb) and the largest genomes, when they usually size around 50 kb. Still, in general, the average size of all prophages was 44.7 kb, agreeing with values previously reported for other bacterial species[3,16,19].

Prophages found integrated in *A. baumannii* mobile genetic elements. We analyzed a set of plasmids associated with the *A. baumannii* strains for the presence of prophages. One (pAB386, Accession no. CP010780) of 26 plasmids were found to possibly harbor an "intact prophage" (Supplementary Table S2). Intactness of the prophage is suggested by the presence of proteins related to phage morphogenesis (capsid and tail elements), packaging (terminase), and host lysis (lysozyme), as well as a potential protein involved in phage-host interaction (putative host specificity protein J) (Supplementary Table S3 and Supplementary Fig. S1). Curiously, plasmid pAB386 and its prophage are highly similar (≥55% genome homology) to a few other plasmids deposited in GenBank (Supplementary Table S3). Some of the *A. baumannii* strains harboring these prophage-containing plasmids have distinct geographical origins (Supplementary Table S3), indicating a possible global dissemination of these elements. In fact, since plasmids are much less specific than phages, prophages integrated in these mobile genetic elements may reach a higher diversity of bacteria, and perhaps cross bacterial species.

Whole genome and proteome comparison of "intact prophages" reveals remarkable diversity. To determine the relationship and diversity of *A. baumannii* prophages we performed whole genome and proteome dot plot analysis of their sequences. Whole genome analysis revealed 19 small clusters of prophages with genome identity above 50%, indicating strong evolutionary relationships (Fig. 4a). Still, the majority of prophages (about 89% of the comparisons) have genome identities below 20%, suggesting an enormous genomic diversity. Interestingly, whole proteome analysis revealed an even higher diversity in the amino acid sequences, with about 97% of the comparisons giving an identity below 10%, and only 8 very small clusters of highly similar prophages (50% identity, Fig. 4b).

To understand if the clusters formed were related to prophage family, we constructed genomic and proteomic phylogenetic trees and inserted family information, as shown in Figs 5 and 6. Genomic clusters 1, 7, 9, 11–13, 15–17, and 19 (identified in Fig. 4a) are comprised of sub-clusters containing highly related phages (more than 90% identity). These sub-clusters are identified in Fig. 5 and are composed of prophages of the same family (when determined). Even for areas of lower identities prophages tend to cluster according to family, although a few singletons are observed. Nevertheless, clusters of the same family are scattered in the tree demonstrating that prophages of the same family can have significantly divergent genomes. A similar analysis is made when observing the proteomics tree (Fig. 6) where sub-clusters of highly related prophages (>90% identity, clusters 2, 4, 5, and 6) tend to group prophages of the same family. All nine clusters of high proteome identity are also clusters with high (>50%) genomic identity. Moreover, only two of the 10 highly (>90%) similar genome sub-clusters are not identified as clusters in the proteomic analysis. Overall, this demonstrates a strong agreement between both analyses.

Prophages encode a multitude of potential virulence factors. The establishment of stable and long relationships between prophages and the bacterial host has profound implications on both bacterial fitness and virulence[10,20]. Here we hypothesize the rapid spread of pathogenicity in *A. baumannii* to be linked with prophages. We have searched for putative virulence genes encoded by the 109 "intact prophages" in study. For this purpose, we considered virulence genes as those that might influence bacterial capacity to colonize a niche in the host, evade or inhibit the host immune defense, resist antibiotics, obtain nutrition from the host, and survive and proliferate in different environmental conditions. We found that 78% of the *A. baumannii* "intact prophages" encode putative virulence genes, with an average of 1.75 VF in their genomes (Fig. 7a). By grouping the virulence genes in classes we were able to analyze those most prevalent, as shown in Fig. 7b. A complete list of VF (and fitness factors) identified per prophage can be found at Supplementary Table S4. To note that some of the putative VF found may also simply be genes involved in the prophage life cycle, or have a dual function in the phage life cycle and providing the bacterial host with a beneficial trait. We attempt to indicate this fact where relevant.

The most prevalent putative VF found were membrane-associated factors (18.4%), e.g. outer membrane proteins/adhesins, lipoproteins, and fimbrial usher protein. These may interfere with bacterial motility, interaction with host cells and phages, and evasion of host immune defenses[21–23].

Importantly, antibiotic resistance genes were identified in the prophages and can be separated into efflux pumps (4.6%) and enzymes (9.2%). Bacterial efflux systems able to export antibiotics found in the prophages include the major facilitator superfamily, the ATP-binding cassette family, and the resistance-nodulation-division family.

Resistance to antibiotics also occurs as a result of drug inactivation, drug modification, or target alteration by enzymes[24,25]. Here, the following enzymes were found: beta-lactamase OXA-23 which confers resistance to carbapenems[26]; pmr and MCR phosphoethanolamine transferases which provide resistance to cationic antimicrobial peptides (e.g. colistin)[27]; chloramphenicol phosphotransferase which prevents chloramphenicol binding to ribosomes[28]; and acetyltransferases conferring resistance to streptogramines.

Several other factors were found that may confer advantages to the bacteria harboring the prophage. Transcriptional regulators were the most prevalent (11.9%), such as TraR/DksA family transcription regulators, proposed to regulate a diverse set of genes including those involved in virulence, activation of stress response, and motility[29,30]; and IcIR family transcriptional regulators that control genes involved in e.g. multidrug resistance and inactivation of QS signals[31]. Some of these transcriptional regulators may also play a role in the regulation of prophage propagation by interfering with the bacterial mechanisms of regulation.

Transporters (7.3%) were found with several putative functions, including tolerance/resistance to toxic compounds (e.g. chromate transporter), siderophore export for iron acquisition (TonB-dependent receptor), and interaction with host cells (glutamine transport system permease)[32].

A few molecular chaperones (5.5%) were also identified. These have important functions in the assembly and replication of phage particles, but may also be involved in fimbriae biosynthesis and thus bacterial motility and adhesion to the host (e.g. fimbrial chaperone protein), or response to stress conditions by protecting newly synthesized or stress-denatured polypeptides from misfolding and aggregation (e.g. GroES, GroEL, DnaJ[33]).
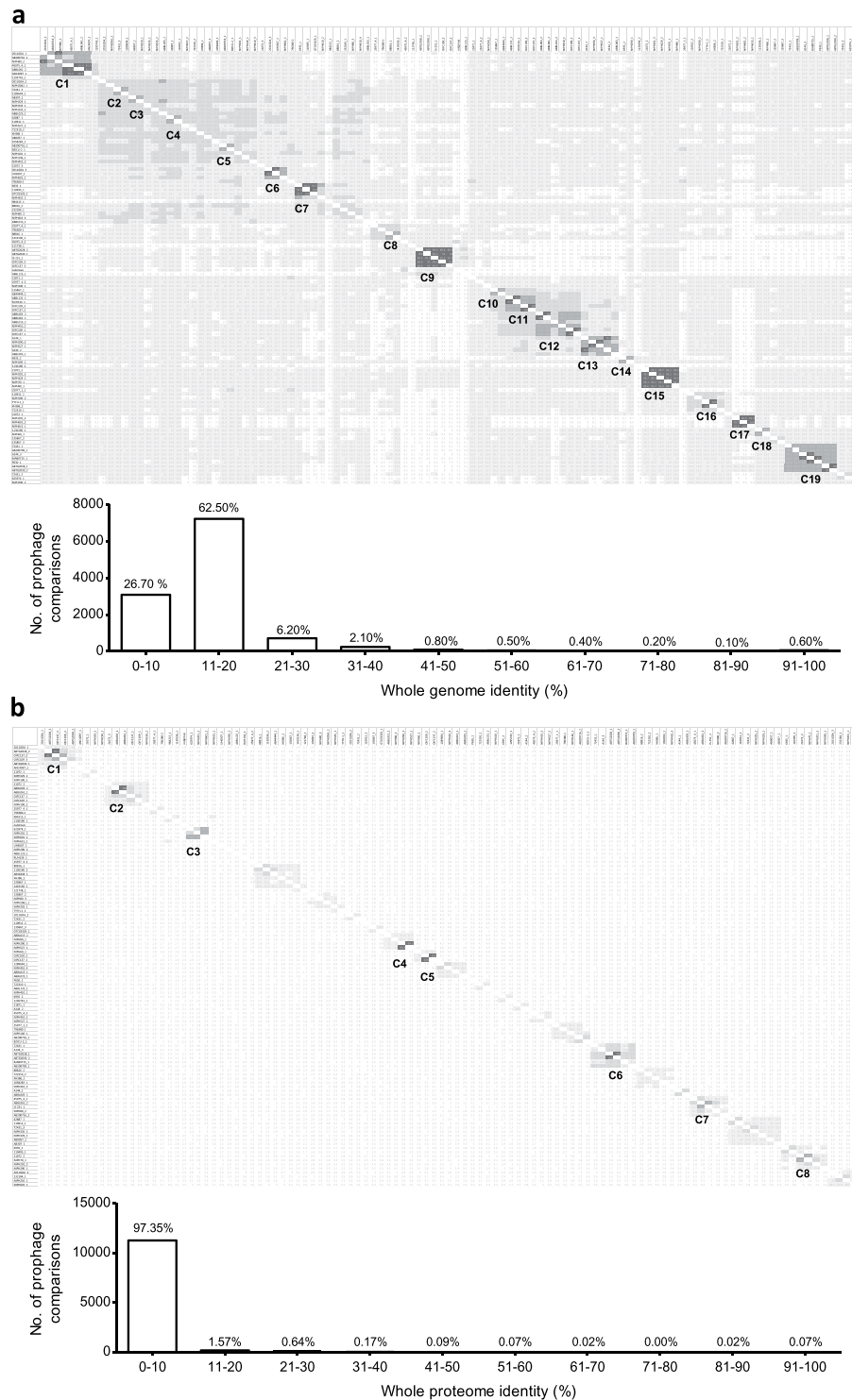
Figure 4. Dot plot matrices of whole sequences of 109 prophages from *Acinetobacter baumannii*. (**a**) Whole genome analysis; and (**b**) Whole proteome analysis. Darker zones indicate higher identity. Clusters of prophages with identities higher than 50% are indicated and numbered. Graphics summarize the frequency of genome identity levels found in the analysis. For dot plot matrices with values of identity see Supplementary Tables S5 and S6. Matrices were adapted from the identity matrices retrieved from the phylogenetic trees constructed using Geneious Tree Builder.

A variety of other putative VF were also found, among which proteins involved in red blood cell degradation (e.g. hemolysin activator protein), manipulation of host functions (e.g. Ankyrin repeat protein), promotion of bacterial survival and persistence under stress conditions (e.g. protein umuD), and targeting of host cells (e.g. RTX toxin).
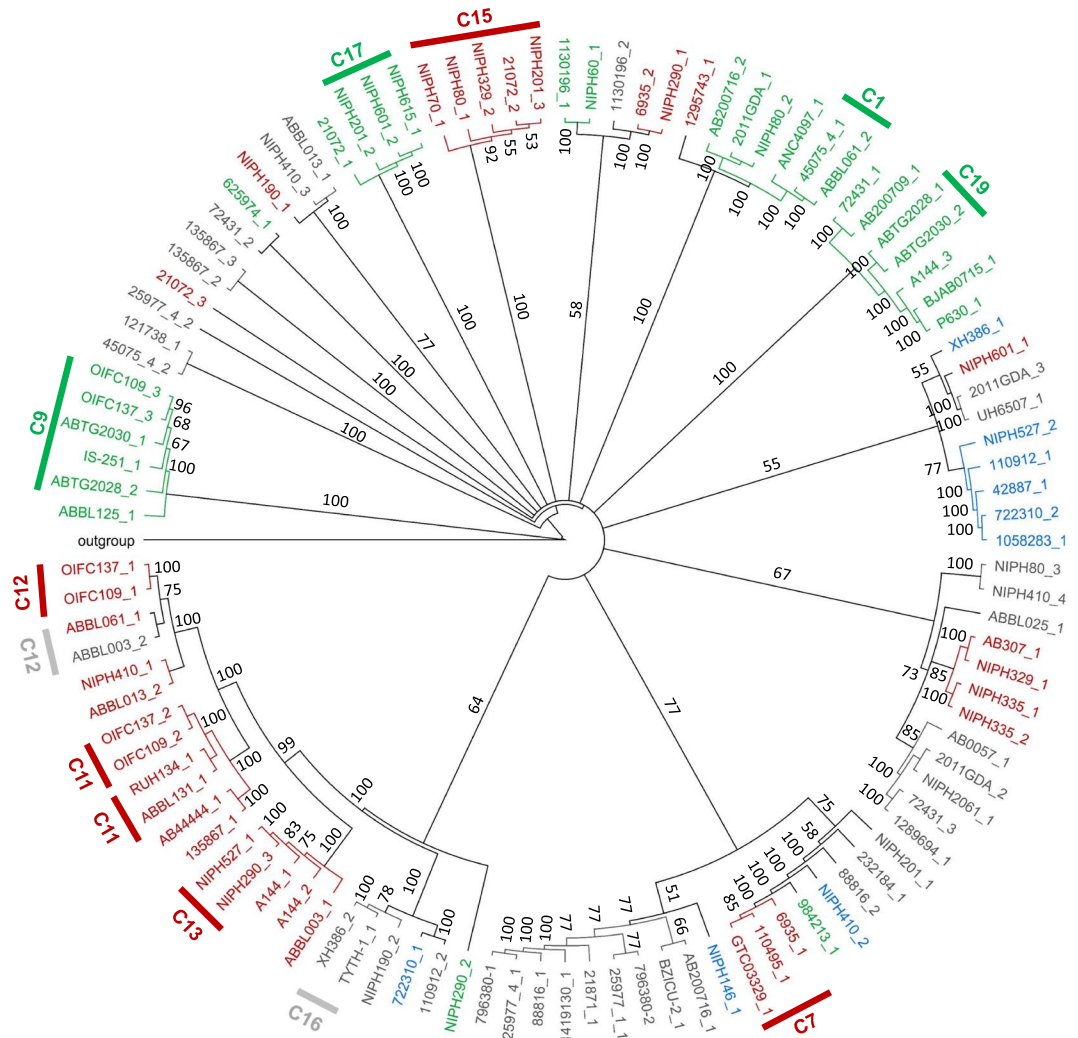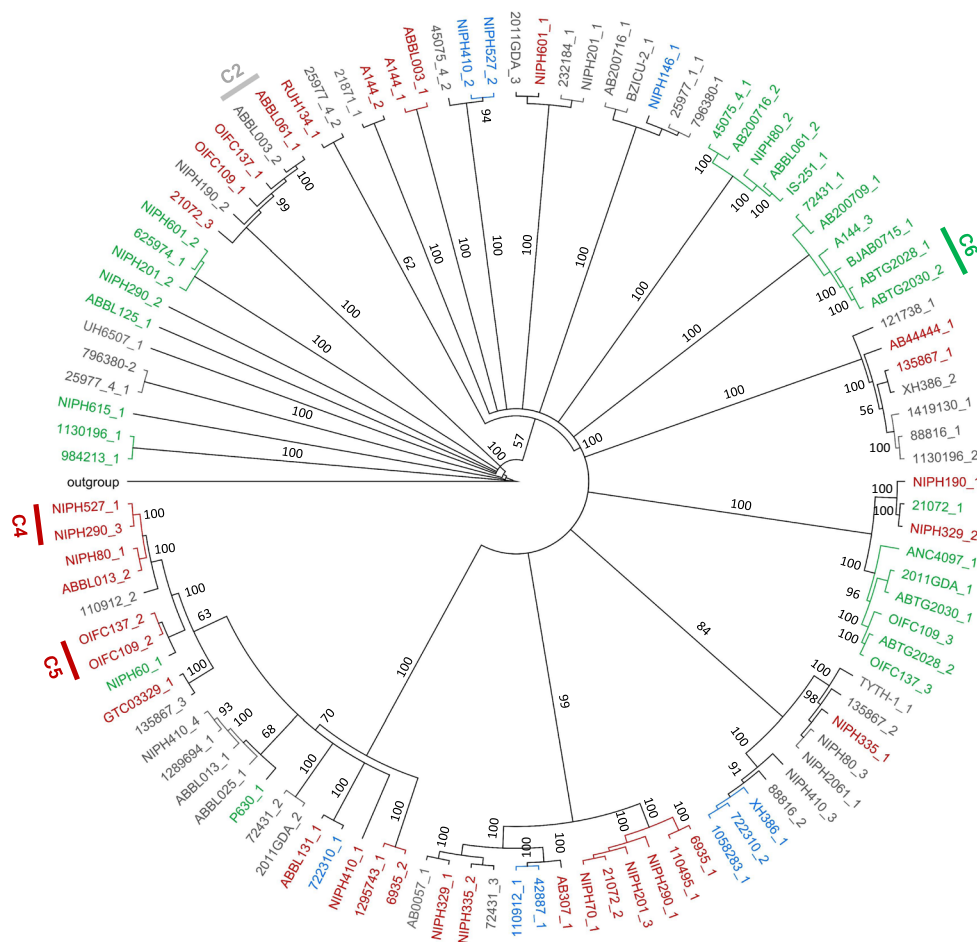
Figure 5. Phylogenetic tree of prophage genomic sequences. Tree was constructed using the Tamura-Nei genetic distance model and the neighbor-joining tree building method in Geneious Tree Builder (Geneious version 9.1.8), with boostrapping set to 100 and tree rooted using *Acinetobacter baumannii* plasmid pNaval18-231 as the outgroup. Tree branches are proportional to branch length, and branch labels represent bootstrap percentages. Clusters of prophages with genome identities above 90% are indicated in the tree. Red: *Siphoviridae*; Green: *Myoviridae*; Blue: *Podoviridae*; Grey: family unknown.

Additionally, several metabolic enzymes were identified in *A. baumannii* prophages (90.8% prevalence), which may improve survival or proliferation of the host and phage. Among these we highlight: enzymes involved in iron acquisition (e.g. porphyrin biosynthetic protein), which provide an advantage to bacteria in the battle for iron with eukaryotic hosts, especially in nutrient-limited niches[34]; enzymes involved in the regulation of bacterial survival under conditions of nutritional (e.g. nucleotide pyrophosphohydrolases[35]) or oxidative stress (e.g. pho-tolyase[36]); enzymes sensing and responding to environmental signals as those resulting from entering the host (e.g. serine/threonine phosphatase[37]); enzymes indirectly involved in antibiotic and xenobiotic resistance (e.g. acetyltransferase family protein[38]), or in rhamnolipid production, i.e. glycolipids with biosurfactant properties involved in bacterial virulence (anthranilate phosphoribosyltransferase)[39].

Discussion

As vehicles for HGT, prophages have been linked to bacterial diversification[40] and evolution[20], and may have strong repercussions on bacterial fitness and virulence[9,10]. Only a few studies have characterized the prevalence of prophages in bacterial species and evaluated their role in virulence. Herein we report the analysis of prophage prevalence in *A. baumannii*, and discuss their possible contribute to the evolution of pathogenicity of this human nosocomial pathogen.

We found *A. baumannii* to harbor prophages in most 795 genomes analyzed. While the majority were defective, a high amount of "intact prophages" were still detected indicating their recent integration. Previous reports[16,19,41–43] have estimated lysogen (including "intact" and defective prophages) prevalence lower than that reported here for *A. baumannii* (99.5%). Still, species as *Streptococcus pyogens* have been reported to have

Figure 6. Phylogenetic tree of prophage proteomic sequences. Tree was constructed using the Jukes-Cantor genetic distance model and the neighbor-joining tree building method in Geneious Tree Builder (Geneious version 9.1.8), with boostrapping set to 100 and tree rooted using *Acinetobacter baumannii* plasmid pNaval18-231 as the outgroup. Tree branches are proportional to branch length, and branch labels represent bootstrap percentages. Clusters of prophages with proteome identities above 90% are indicated in the tree. Red: *Siphoviridae*; Green: *Myoviridae*; Blue: *Podoviridae*; Grey: family unknown.
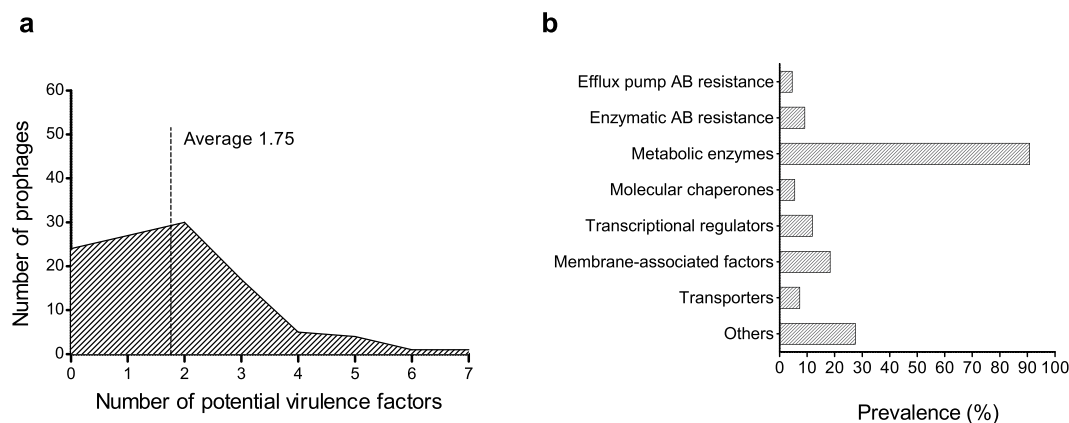
Figure 7. Putative virulence genes identified in the genomic sequences of 109 "intact" prophages of *Acinetobacter baumannii*. (**a**) Distribution of putative virulence genes per prophage; (**b**) Prevalence of potential virulence factors grouped by class.

similar high levels of lysogens (90%)[19]. It appears that some species are more prone to be lysogenized than others, although the variables associated to the process remain largely unknown. Touchon *et al.*[16] found minimal doubling time and genome size to be the variables most correlated with lysogeny[16]. Fast growing bacteria (with minimal doubling times <2.5 h) were shown to be more lysogenized than slow growing bacteria. Doubling times of *A. baumannii* have been reported to be around 0.5 h[44]; as a fast grower a higher percentage of lysogens is therefore expected. An explanation for this phenomenon has been suggested; fast growing bacteria grow weakly under poor environmental conditions[45]. In such circumstances phages tend to assume a lysogenic life cycle to preserve their genome while waiting for more propitious conditions for lytic propagation[46]. The prevalence of *A. baumannii* in hospital environments, where growth conditions are not ideal, may play a fundamental role in the high prevalence of prophages in *A. baumannii* genomes. Our analysis also suggests that *A. baumannii* strains of larger genomes harbor more prophages, but only for genomes up to 4.2 Mbp. Touchon *et al.*[16] had similar observations in a set of prophages of mixed species, but stabilization of the number of prophages occurred only above 6 Mbp[16]. They have suggested two hypothesis that may also apply here. First, larger genomes may result from selection for functional diversification by HGT, thus facilitating prophage integration. After a certain moment, it is possible that further integration of prophages will not result in the acquisition of novel functions and thus bacteria may become less prone for accepting this type of HGT. Second, superinfection exclusion may be more effective in bacteria with multiple prophages, leading to saturation of prophages in larger genomes. Still, future work is needed to understand the correlation of bacterial genome size and number of integrated prophages.

Among a subset of 109 "intact prophages" we found *Siphoviridae* to be the most prevalent family, followed by *Myoviridae* and *Podoviridae*, in agreement with the assumed distribution of tailed phages in nature[18]. However, the average genome sizes of each prophage family diverged from common descriptions. More strikingly, different trends were observed for each family. *Siphoviridae* had sizes above average and we hypothesize these differences to result from the acquisition of bacterial genes adjacent to the prophage during repeated excision and integration cycles. Conversely, prophages of the *Myoviridae* family have a genome much smaller than the average *Myoviridae* deposited on GenBank. In fact, this family had the smallest average genome size, when it is commonly characterized by the largest phages. A similar observation was made by Bobay *et al.*[3], although for prophages of different taxa[3]. As they have done, we also suggest these phages might have endured some form of genetic degradation that caused a significant reduction of genome size. The reasons why distinct prophage families seem to have evolved differently in the bacterial genomes are unknown. We hypothesized that the analysis of draft genomes using PHAST could erroneously delimit prophages (e.g. if these are distributed through different contigs) and thus give rise to unexpected sizes. However, family analysis considered only phages identified as "intact" and whose ends were manually curated, so we expect bias caused by PHAST to be majorly reduced. We also discarded the hypothesis of inaccurate classifications given by our prophage classification method for two reasons. First, we only attributed prophages with a taxa when comparative analysis of three genes gave concordant classifications. Second, the phylogenetic tree constructed clearly indicates the clustering of phages from the same taxa, supporting our classification. Finally, although we attempted to eliminate method-related bias from our analysis by choosing only genomes sequenced by Illumina and with coverage higher than 50%, we cannot discard assembly problems as a possible reason for some size discrepancies. Further studies are necessary to reveal if this is a common trend among prophages of all bacterial species, if it is specific of *A. baumannii*, or if it is simply a methodology-related bias.

On an interesting note, we found that for a (not so) few prophages, different proteins (e.g. capsid and large terminase proteins) indicated a distinct family, e.g. the prophage of strain NIPH 2061 (1272028-1318975 bp) identified as *Siphoviridae* or *Podoviridae*. We believe this to reflect the mosaic nature of phages, and to suggest the exchange of genetic information by homologous recombination between prophages and infecting phage genomes or other prophages in the same cell. Among other genetic trades, structural genes may be exchanged leading to a difficult interpretation of phage family when exclusively based on genomic information. For example, we found that many prophages having a tail tape measure protein, perceived as characteristic of long tailed phages[47,48] were classified as *Podoviridae*, e.g. prophages of strains NIPH 146 (3227804-3269564 bp) and NIPH 527 (1517738-1585337 bp). The presence of this gene is therefore non synonymous with long tailed phages, as recently suggested by Ma *et al.*[49]).

Also interestingly, we found some "intact prophages" integrated in *A. baumannii* plasmids. It is possible that plasmids have acquired the prophages via homologous recombination with the bacterial chromosome, or perhaps by direct integration of the phage into the plasmid. Prophage integration in plasmids may have important implications for the genetic trades occurring within and among bacterial species, resulting in an extremely rich, available gene bank.

Prophages of *A. baumannii* were found to be greatly diversified. Our comparison of 109 "intact prophages" revealed less than 20% of genome identity and less than 10% proteome identity among the majority of prophages. This may indicate one or more of the following: a diversification of prophages into different lineages in ancient times; the constant and intensive diversification of prophage genomes by genetic trades; or a distinct origin of *A. baumannii* prophages (e.g. derived from different bacterial species). Still we could identify a few small clusters of prophages with genomic and proteomic identities suggesting stronger evolutionary relationships. These may be related to phage taxa, since clusters of high identity tended to group prophages of the same family.

Some of the genes expressed from prophages can alter the properties of the host, ranging from increased protection against further phage infection, to increased virulence[9]. Many cases have been reported linking pathogen virulence to the acquisition of prophages, among which are the well-known *E. coli* O157:H7 whose virulence is correlated with two Shiga-toxin-encoding prophages[50,51], or *Vibrio cholerae*, producer of the cholera toxin encoded by phage CTXφ[52]. Here we found prophages to frequently encode genes of putative function related to bacterial virulence and fitness. The prophage-encoded genes may be contributing to the high levels of multidrug resistance found in *A. baumannii*. We identified both drug-specific and multidrug efflux pumps, as well as enzymes able to inactivate/modify the antibiotic or its bacterial target. Among these we highlight the presence

of enzymes conferring resistance to colistin, one of the very few last resource antibiotics. Spread of resistance is therefore fostered by prophages, especially under stress conditions that induce prophage excision, as those encountered by bacteria when entering the host environment.

Several other putative VF were found, such as membrane-associated factors, transcriptional regulators, transporters, chaperones, and other proteins, with functions in protection from nutritional and oxidative stress, bacterial motility, interaction with host cells, evasion of host immune defenses, iron acquisition, and regulation of virulence gene expression.

Overall, our results suggest a significant contribution of prophages for the evolution and spread of *A. baumannii* pathogenicity and highlight the clinical relevance of these virions. This study centered the analysis on virulence genes of "intact prophages" only. However, defective prophages, which are the vast majority, most probably also codify for genes of relevance to *A. baumannii* pathogenicity.

Methods

Data collection.
A data set of 795 complete genomes of *A. baumannii* were retrieved from GenBank (last accessed March 2016). All the genomes selected were sequenced by Illumina and had a genome coverage above 45x. Other than these two features, strain selection was random. The vast majority of the genomes deposited were at scaffold assembly level, and were used in the analysis as unassembled contigs (Supplementary Table S2).

Detection of prophages in *A. baumannii* strains.
Prophages were detected using the PHAge Search Tool (PHAST) webserver[53], using the GenBank accession number for complete genomes, and the nucleotide sequence file in FASTA format for draft genomes, selecting the option of contigs file for concatenation of all sequences together prior to analysis. PHAST separates the identified prophages into intact, questionable and incomplete according to criteria that consider the number of coding DNA sequences (CDSs) of a region attributable to prophage CDSs, and the presence of phage-related genes. For the purposes of our analysis, questionable and incomplete prophages were grouped as defective prophages (Supplementary Table S2). Still, an analysis of prophage prevalence considering questionable and incomplete prophages individually can be seen in Supplementary Fig. S2. Prophages identified by PHAST were manually curated for increased stringency; only prophages with identified integrase and/or at least one structural gene (e.g. capsid, tail, tail fiber) were considered. "Intact prophages" were identified as such when all elements required for phage infection were present. Prophages smaller than 10 kb were excluded because these may be difficult to distinguish from other integrative elements[17,19]. The number of "intact" and defective (questionable and incomplete) prophages identified for each strain are detailed in Supplementary Table S2.

Classification of prophages.
The prophages of a set of 99 strains were selected for further assessment. Strains were chosen based on the GenBank dendrogram at the date of April 11, 2016 (see Supplementary Fig. S3). At least one strain from each branch of the dendrogram was chosen. Branch size was considered, i.e. more strains were selected from larger branches. Strains selected harbor a total of 109 "intact prophages", whose limits were manually curated using gene annotation and PFAM 30.0 protein functions. Prophages were classed by comparison of the major capsid protein and large terminal subunit[54–56] to those of previously deposited classed phages, using BLASTp version 2.8.0+ with default parameters. The presence of elements characteristic of specific families, as the tail sheath of *Myoviridae* or the tail tape measure protein of long tailed phages[9], was also considered for prophage classification. Prophages were attributed a family only when the results of all comparisons gave identical classifications, with E-values lower than 1×10^{-7}, identity higher than 35% and coverage higher than 50%. Prophage classification can be seen in Supplementary Table S4.

Whole genome and proteome comparisons.
Prophage genomic and proteomic sequences were aligned using MAFFT version 7.304[57], using the Phylip output format, sorted, strategy "—auto". The genome phylogenetic tree was constructed using the Tamura-Nei genetic distance model and the neighbor-joining tree building method in Geneious Tree Builder (Geneious version 9.1.8[58]). The proteome phylogenetic tree was constructed using the Jukes-Cantor genetic distance model. Boostrapping was set to 100 and the trees were rooted using *A. baumannii* plasmid pNaval18-131 as the outgroup. The identity matrix generated during construction of the phylogenetic trees was used to infer on whole genome and whole proteome identity.

Identification of potential virulence factors encoded by prophages.
The subset of 109 "intact prophages" was analyzed for the encoding of putative VF. For this, prophage proteins with assigned function (by myRAST and sequence comparison using PFAM 30.0 and HMMER webserver version 2.25.0, with default parameters) were correlated to putative VF using PubMed search. Protein sequences identified as related to antibiotic resistance were further analyzed using the Resistance Gene Identifier (RGI) of CARD, selecting criteria of perfect, strict and loose hits. A full list of the identified putative VF can be found at Supplementary Table S4, and the results of RGI analysis can be seen in Supplementary Table S7.

Statistical analysis.
Statistical analysis of the data was performed using the independent samples t-test for comparisons of two samples, or one-way analysis of variance (ANOVA) with post-hoc Tukey HSD test for comparing multiple variables, using the software GraphPad Prism 5 version 5.03, and considering a significance level of 95%.

Data Availability
All data generated or analyzed during this study are included in this published article and its Supplementary Information files.

References

1. Imperi, F. *et al*. The genomics of Acinetobacter baumannii: Insights into genome plasticity, antimicrobial resistance and pathogenicity. *IUBMB Life* **63**, 1068–1074 (2011).
2. Thurber, R. V. Current insights into phage biodiversity and biogeography. *Curr Opin Microbiol* **12**, 582–587 (2009).
3. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci USA* **111**, 12127–12132 (2014).
4. Mahony, J., McGrath, S., Fitzgerald, G. F. & van Sinderen, D. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. *Appl Environ Microbiol* **74**, 6206–6215 (2008).
5. Hofer, B., Ruge, M. & Dreiseikelmann, B. The superinfection exclusion gene (sieA) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. *J Bacteriol* **177**, 3080–3086 (1995).
6. Edlin, G., Lin, L. & Bitner, R. Reproductive fitness of P1, P2, and Mu lysogens of *Escherichia coli*. *J Virol* **21**, 560–564 (1977).
7. Wang, X. *et al*. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**, 147 (2010).
8. Shaikh, N. & Tarr, P. I. *Escherichia coli* O157:H7 shiga toxin-encoding bacteriophages: Integrations, excisions, truncations, and evolutionary implications. *J Bacteriol* **185**, 3596–3605 (2003).
9. Casjens, S. Prophages and bacterial genomic: what have we learned so far? *Mol Microbiol* **49**, 277–300 (2003).
10. Canchaya, C., Fournous, G. & Brussow, H. The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53**, 9–18 (2004).
11. Matos, R. C. *et al*. Enterococcus faecalis prophage dynamics and contributions to pathogenic traits. *PLOS Gen* **9**, e1003539 (2013).
12. Asadulghani, M. *et al*. The defective prophage pool of *Escherichia coli* O157: Prophage–prophage interactions potentiate horizontal transfer of virulence determinants. *Plos Pathog* **5**, e1000408 (2009).
13. Lang, A. S., Zhaxybayeva, O. & Beatty, J. T. Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Microbiol* **10**, 472–482 (2012).
14. Michel-Briand, Y. & Baysse, C. The pyocins of Pseudomonas aeruginosa. *Biochimie* **84**, 499–510 (2002).
15. Leiman, P. G. *et al*. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci* **106**, 4154–4159 (2009).
16. Touchon, M., Bernheim, A. & Rocha, E. P. C. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J* **10**, 2744–2754 (2016).
17. Bobay, L. M., Rocha, E. P. C. & Touchon, M. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol* **30**, 737–751 (2013).
18. Ackermann, H.-W. 5500 Phages examined in the electron microscope. *Arch Virol* **152**, 227–243 (2007).
19. Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. Prophage genomics. *Microbiol Mol Biol Rev* **67**, 238–276 (2003).
20. Brussow, H., Canchaya, C. & Hardt, W. D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560–602 (2004).
21. Barondess, J. J. & Beckwith, J. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J Bacteriol* **177**, 1247–1253 (1995).
22. Wang, H. *et al*. Influence of fimbriae on bacterial adhesion and viscoelasticity and correlations of the two properties with biofilm formation. *Langmuir* **33**, 100–106 (2017).
23. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E. V. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**, 913–923 (2002).
24. Dermott, P. F. M., Walker, R. D. & White, D. G. Antimicrobials: Modes of action and mechanisms of resistance. *Int J Toxicol* **22**, 135–143 (2003).
25. Wright, G. D. Mechanisms of resistance to antibiotics. *Curr Opin Chem Biol* **7**, 563–569 (2003).
26. Evans, B. A. & Amyes, S. G. B. OXA β-Lactamases. *Clin Microbiol Rev* **27**, 241–263 (2014).
27. Anandan, A. *et al*. Structure of a lipid A phosphoethanolamine transferase suggests how conformational changes govern substrate binding. *Proc Natl Acad Sci* **114**, 2218–2223 (2017).
28. Shaw, W. V. *et al*. Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. *Nature* **282**, 870–872 (1979).
29. Pal, R. R., Bag, S., Dasgupta, S., Das, B. & Bhadra, R. K. Functional characterization of the stringent response regulatory gene dksA of Vibrio cholerae and its role in modulation of virulence phenotypes. *J Bacteriol* **194**, 5638–5648 (2012).
30. Girard, M. E. *et al*. DksA and ppGpp regulate the σS stress response by activating promoters for the small RNA DsrA and the anti-adapter protein IraP. *J Bacteriol* **200**, e00463–17 (2017).
31. Molina-Henares, A. J., Krell, T., Eugenia Guazzaroni, M., Segura, A. & Ramos, J. L. Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol Rev* **30**, 157–186 (2006).
32. Clark, T. J., Momany, C. & Neidle, E. L. The benPK operon, proposed to play a role in transport, is part of a regulon for benzoate catabolism in Acinetobacter sp. strain ADP1. *Microbiology* **148**, 1213–1223 (2002).
33. Susin, M. F., Baldini, R. L., Gueiros-Filho, F. & Gomes, S. L. GroES/GroEL and DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in Caulobacter crescentus. *J Bacteriol* **188**, 8044–8053 (2006).
34. Skaar, E. P. The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS Pathog* **6**, e1000949 (2010).
35. Lu, L.-d *et al*. Mycobacterial MazG is a novel NTP pyrophosphohydrolase involved in oxidative stress response. *J Biol Chem* **285**, 28076–28085 (2010).
36. Ziegelhoffer, E. C. & Donohue, T. J. Bacterial responses to photo-oxidative stress. *Nat Rev Micro* **7**, 856–863 (2009).
37. Canova, M. J. & Molle, V. Bacterial serine/threonine protein kinases in host-pathogen interactions. *J Biol Chem* **289**, 9473–9479 (2014).
38. Llano-Sotelo, B., Azucena, E. F. Jr., Kotra, L. P., Mobashery, S. & Chow, C. S. Aminoglycosides modified by resistance enzymes display diminished binding to the bacterial ribosomal aminoacyl-tRNA site. *Chem Biol* **9**, 455–463 (2002).
39. Zulianello, L. *et al*. Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by Pseudomonas aeruginosa. *Infect Immun* **74**, 3134–3147 (2006).
40. Desiere, F., McShan, W. M., van Sinderen, D., Ferretti, J. J. & Brüssow, H. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: Evolutionary implications for prophage-host interactions. *Virology* **288**, 325–341 (2001).
41. Casjens S, Hendrix RW. Bacteriophages and the bacterial genome in *Bacterial chromosome* (ed. Higgins, P. N.) 39–53 (ASM Press, 2015).
42. Paul, J. H. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**, 579–589 (2008).
43. Ghosh, D. *et al*. Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and trzN genes in viral-community DNA. *Appl Environ Microbiol* **74**, 495–502 (2008).
44. Moffatt, J. H. *et al*. Colistin resistance in Acinetobacter baumannii is mediated by complete loss of lipopolysaccharide production. *Antimicrob Agents Chemother* **54**, 4971–4977 (2010).
45. Koch, A. L. Oligotrophs versus copiotrophs. *Bioessays* **23**, 657–661 (2001).
46. Stewart, F. M. & Levin, B. R. The population biology of bacterial viruses: why be temperate? *Theor Popul Biol* **26**, 93–117 (1984).
47. Pell, L. G., Kanelis, V., Donaldson, L. W., Lynne Howell, P. & Davidson, A. R. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci USA* **106**, 4160–4165 (2009).
48. Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses* **3**, 172–203 (2011).

49. Ma, Y. *et al*. Isolation and molecular characterisation of Achromobacter phage phiAxp-3, an N4-like bacteriophage. *Sci Rep* **6**, 24776 (2016).
50. Ohnishi, M., Kurokawa, K. & Hayashi, T. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol* **9**, 481–485 (2001).
51. Hayashi, T. *et al*. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11–22 (2001).
52. Waldor, J. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
53. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: A Fast Phage Search Tool. *Nucleic Acid Res* **39**, W347–W352 (2011).
54. Comeau, A. M. & Krisch, H. M. The capsid of the T4 phage superfamily: The evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol Biol Evol* **25**, 1321–1332 (2008).
55. Proux, C. *et al*. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. *J Bacteriol* **184**, 6026–6036 (2002).
56. Serwer, P. *et al*. Improved isolation of undersampled bacteriophages: finding of distant terminase genes. *Virology* **329**, 412–424 (2004).
57. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
58. Kearse, M. *et al*. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

Acknowledgements

Author Contributions

J.A. and A.R.C. designed and conceived the original idea for the study. A.R.C. and R.M. collected, analyzed, and interpreted the data. A.R.C. and R.M. wrote the manuscript. J.A. and A.R.C. critically revised the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-33800-5.

Competing Interests: The authors declare no competing interests.

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