Genome organisation of the *Acinetobacter* lytic phage ZZ1 and comparison with other T4-like *Acinetobacter* phages

Jing Jin1,2*, Zhen-Jiang Li1, Shu-Wei Wang1, Shan-Mei Wang3, Song-Jian Chen1, De-Hai Huang1, Gai Zhang1, Ya-Hui Li1, Xiao-Ting Wang1, Jin Wang1 and Guo-Qiang Zhao2*

**Abstract**

**Background:** Phage ZZ1, which efficiently infects pathogenic *Acinetobacter baumannii* strains, is the fifth completely sequenced T4-like *Acinetobacter* phage to date. To gain a better understanding of the genetic characteristics of ZZ1, bioinformatics and comparative genomic analyses of the T4 phages were performed.

**Results:** The 166,687-bp double-stranded DNA genome of ZZ1 has the lowest GC content (34.4%) of the sequenced T4-like *Acinetobacter* phages. A total of 256 protein-coding genes and 8 tRNA genes were predicted. Forty-three percent of the predicted ZZ1 proteins share up to 73% amino acid identity with T4 proteins, and the homologous genes generally retained the same order and transcriptional direction. Beyond the conserved structural and DNA replication modules, T4 and ZZ1 have diverged substantially by the acquisition and deletion of large blocks of unrelated genes, especially in the first halves of their genomes. In addition, ZZ1 and the four other T4-like *Acinetobacter* phage genomes (Acj9, Acj61, 133, and Ac42) share a well-organised and highly conserved core genome, particularly in the regions encoding DNA replication and virion structural proteins. Of the ZZ1 proteins, 70, 64, 61, and 56% share up to 86, 85, 81, and 83% amino acid identity with Acj9, Acj61, 133, and Ac42 proteins, respectively. ZZ1 has a different number and types of tRNAs than the other 4 *Acinetobacter* phages, although some of the ZZ1-encoded tRNAs share high sequence similarity with the tRNAs from these phages. Over half of ZZ1-encoded tRNAs (5 out of 8) are related to optimal codon usage for ZZ1 proteins. However, this correlation was not present in any of the other 4 *Acinetobacter* phages.

**Conclusions:** The comparative genomic analysis of these phages provided some new insights into the evolution and diversity of *Acinetobacter* phages, which might elucidate the evolutionary origin and host-specific adaptation of these phages.

**Keywords:** Phage genome annotation, Phage genome organisation, Comparative genomic analyses, T4-like phage

**Background**

ZZ1, a novel sequenced lytic phage that can efficiently infect *Acinetobacter baumannii* clinical strains, was identified from fishpond water in Zhengzhou, China [1]. As is typical for T4-like virion morphology, the ZZ1 viral particle contains an isometric head and a contractile tail, which are visible with transmission electron microscopy.

The phage was reported to have different antibacterial activity against three *A. baumannii* clinical strains (AB09V, AB0901, and AB0902). Of the host bacteria, AB09V is the most sensitive, and ZZ1 causes large and distinguishable plaque formation on a lawn of AB09V and is highly infectious with a short latent period (9 min) and a large burst size (200 PFU/cell) [1].

To date, the genomes of 12 *Acinetobacter* phages have been completely sequenced and published in the NCBI genome database. Five belong to the *Myoviridae* phage group, including 133, Acj9, Acj61, and Ac42, and four belong to the *Siphoviridae* phage group, including Abp1...
Table 1 General genome features of ZZ1, T4 and the other completely sequenced T4-like Acinetobacter phages

| Phage name | Bacterial strain used in phage isolation | Accession | Size (bp) | No. of predicted CDSs | Gene density | % coding sequence | Avg gene product size (aa) | GC% |
|------------|-----------------------------------------|-----------|----------|-----------------------|--------------|-------------------|--------------------------|-----|
| ZZ1        | Acinetobacter baumannii                 | NC_018087.2 | 166687   | 256                   | 1.5          | 93.9              | 203                      | 34.41|
| 133        | Acinetobacter johnsonii                 | NC_015250.1 | 159801   | 257                   | 1.6          | 95.5              | 197                      | 39.67|
| Acj9       | Acinetobacter johnsonii                 | NC_014663.1 | 169947   | 253                   | 1.5          | 93.3              | 208                      | 40.03|
| Acj61      | Acinetobacter johnsonii                 | NC_014661.1 | 164093   | 241                   | 1.5          | 92.5              | 209                      | 39.01|
| Ac42       | Acinetobacter sp.                       | NC_014660.1 | 167716   | 255                   | 1.5          | 94.9              | 207                      | 36.37|
| T4         | Escherichia coli                        | NC_000866.4 | 168903   | 278                   | 1.6          | 97.8              | 197                      | 35.30|
new viral proteins were revealed by the characterisation of this phage. However, further Batch CD-Search analyses suggested that 17 of the 97 genes with functional annotations lack protein domain information. Of the 256 CDSs, 156 were conserved hypothetical proteins, and only 23 of the proteins (9%) had protein domain information (see Figure 2). Further functional analysis revealed that the 97 CDSs assigned putative functions could be classified into 12 functional categories according to previous descriptions of phage T4 [3] (see Figure 1 and Additional file 1). Most of the named functional proteins are highly conserved among T4-like phages and are either structural (36 CDSs) or involved in DNA replication, recombination, repair, packaging, and processing (21 CDSs) (see Figure 2 and Additional file 1).

**Comparative genomics of T4 and ZZ1**

The determination of phage relatedness is not based exclusively on sequence similarity but also takes genome organisation into consideration [15]. The phage ZZ1 has similar virion morphology [1], genome size and number of CDSs compared to the coliphage T4 (168903 bp, 278 CDSs) (Figure 3). To further investigate the genomic similarities of T4 and ZZ1, we constructed comparative genome maps of ZZ1 and T4. The putative gene rIIA from ZZ1 was positioned on the minus strand (leftward...
Figure 2 Summary of the 256 ZZ1 genes annotated by BLASTP and Batch CD-Search analyses. (A) Pie chart showing the relative abundance of genes with and without named functions, including 95 KFP (known function proteins), 158 CHP (conserved hypothetical protein), 23 CHP-D (CHP with protein domain), and 3 HP (hypothetical protein) genes. (B) Relative abundance of ZZ1 proteins in 12 functional categories: A, Transcription; B, Translation; C, Nucleotide metabolism; D, DNA replication, recombination, repair, packaging, and processing; E, Virion proteins; F, Chaperons/assembly catalysts; G, Lysis; H, Host or phage interactions; I, Host alteration/shutoff; J, Homing endonucleases and homologs; K, Predicted membrane spanning proteins in unknown function proteins; L, Unknown function.

Figure 3 T4-ZZ1 comparison genomic mapping and ZZ1 orthologous genes from the other 4 Acinetobacter phages. The genomes and coding regions of T4 and ZZ1 are drawn in parallel approximately to scale. Alignment of different genes is from T4 (upper line) and from ZZ1 (lower line). Following convention, the map starts at the top left with the rIIA gene and ends at the bottom right with the rIIB gene. Horizontal arrows indicate transcription direction. The likely function of the genes is indicated by the colour of the arrows as described in Figure 1. Genes sharing protein sequence identity are linked according to the colour key provided at the middle left. ZZ1 genes are annotated randomly with their gene order in the GenBank database. The map was artificially split into upper halves and lower halves. ZZ1 genes that are orthologous to the 4 other Acinetobacter phages are indicated by coloured boxes under the corresponding ZZ1 genes. The amino acid sequence identity of ZZ1 genes to these orthologous genes was also indicated by colour as described in the key provided at the upper right. The % coverage of ZZ1 genes to these orthologous genes is indicated by the grey boxes under each coloured shade or box according to the grey scale key at the upper right. The genes indicated by the grey vertical arrows share significant similarity with proteins from the 4 other Acinetobacter phages, but share no significant similarity with T4 proteins.
transcription) and at the start of the map, following the T4 convention. Homologous genes are projected onto the T4 map, and the projections are colour-coded based on protein sequence similarity (see Figure 3). BLASTP analysis indicated that the ZZ1 CDS140 shows significant similarity to the putative Alt-like gene (<7e-23 E value, > 80% coverage, and >50% maximum identity) from the other T4-like Acinetobacter phages, Acj9, Acj61, 133, and A42 (see Figure 3). However, when we compared the ZZ1 CDS140 to the set of all T4 proteins by aligning two or more sequences from BLASTP using the default parameters, no counterpart was found in the results. Further BLASTP analysis revealed that, of the 4 Acinetobacter phages, only phage Acj61 vs showed a slight similarity to the T4 vs (0.1 E value, 99% coverage, and 32% max identity). Similar results were observed in the annotation of ZZ1 CDS156, which showed significant similarity to the putative 57A (<e-4 E value, >70% coverage, and >37% max identity) from phages Ac42 and Acj61 although no counterpart was found in the BLASTP results when we compared ZZ1 CDS156 to the set of all T4 proteins. Thus, we annotated ZZ1 CDS140 and 156 as hypothetical proteins (see Figure 1), and of the 256 CDSs from ZZ1, only 95 (37.1%) could be assigned a functional annotation, and 158 (61.7%) were conserved hypothetical proteins (see Figure 2). Overall, such gradual evolution of the T4 genes between these closely and distantly related phages may eventually provide some interesting insights into the particular steps of genome divergence.

Notably, 110 of the 256 ZZ1 CDSs (43%) shared up to 73% amino acid sequence identity with T4 proteins (E value < 10^-40) (see Figure 3 and Additional file 1). Long blocks of synteny containing homologous genes are interspersed with stretches that lack homology. The lower halves of the genome, which are closely related, include three separate clusters of rightward-transcribed genes, specifically: (i) a cluster of base plate wedge/ head/tail genes (genes 53 to 24), (ii) a smaller base plate hub gene cluster in its vicinity (genes 51 to 54), and (iii) further away, a tail fibre gene cluster consisting of a few large genes (genes 34 to t). Two clusters of leftward-oriented non-structural genes separate the 3 rightward-oriented clusters and mainly encode proteins involved in nucleotide metabolism. The degree of sequence identity varied substantially. The degree of sequence conservation did not follow structural/non-structural gene divisions.

The region that displays the most significant difference between ZZ1 and T4 is located in the upper halves of the genome that exclusively encode leftward-oriented, non-structural genes (with the exception of one rightward-oriented unknown gene, CDS147), including a large DNA replication module. Large regions of non-aligned genome segments were thought to have resulted from genetic gains, losses, or replacements, i.e., events similar to those that have shaped the evolution of all microbial genomes in nature. For example, ZZ1 rIIA.1, 60.1 and mobA are flanked by 14 unknown genes and 19 genes were deleted between ZZ1 vs 1 (CDS141) and segB (CDS142). Further alignment and comparison of T4 genes that are homologous to ZZ1 revealed that the order and transcriptional direction of T4 genes are broadly maintained in the ZZ1 genome, which helped provide the phylogenetic distance between ZZ1 and T4. There were three major differences: 1) the rnlB gene was at the junction of 24 in the lower genome half in T4 but was adjacent to the CDS012 in the upper genome half of ZZ1 (see Figure 3); 2) the topoisomerase gene (39 + 60) and the DNA polymerase gene (43) were split and intact, respectively, in T4 but intact and split, respectively, in phage ZZ1; and 3) a cluster of Alt genes including three putative Alt-like genes differs between T4 and ZZ1. As shown in Figure 3, the transcriptional direction of the three ZZ1 Alt-like genes was different from the T4 Alt, revealing an unusual genomic organisation for the Acinetobacter phage ZZ1. Moreover, the first Alt counterpart (CDS209) is 695 aa in length and shares 98% coverage and 28% amino acid identity with the T4 Alt (total 682 aa in length). The second Alt-like gene (CDS210) is 229 aa in length and shares 35% amino acid identity with the T4 Alt segment that extends from 1 to 217 amino acids, and the third Alt analogue (CDS211) is 463 aa in length and shares 33% amino acid identity with the T4 Alt segment that extends from 409 to 643 aa. Apparently, the last two ZZ1 Alt-like genes resulted from an even splitting event from the ancestral Alt gene (see Figure 3). The duplication of the Alt gene was also present in the genomes of the Acinetobacter phage Ac42 (see Figure 4), the coliphage JS98 (NC_010105.1), and the Enterobacteria phage Bp7 (NC_019500.1), but the splitting of the Alt gene was not observed in other T4-like phages.

T4-like phages were often shaped by gene duplications. For example, in T4, genes 23 and 24 share 29% amino acid identity [3], and 16 proteins (mostly hypothetical) were identified as likely to have been duplicated in the KVP40 lineage [8]. To further examine duplications, we aligned genes from ZZ1 with other ZZ1 genes using BLASTP at the amino acid level and BLASTN at the DNA level. Those with better matches (based on E value) were considered candidates for lineage-specific duplications. Using this method, 14 proteins were identified as likely to have been duplicated in the ZZ1 lineage (data not shown). We further reasoned that if the homologous genes are the result of duplication of an ancestral gene, and if the different degrees of sequence diversification of these duplicated genes resulted from the accumulation of nucleotide point mutations resulting from differential selection pressure, then the genes with
lineage-specific duplications should have significant nucleotide sequence identities. However, although the duplicated genes have significant amino acid identities, no significant nucleotide sequence identities were observed. Thus, the similarity of the genes is more likely to be the result of convergent evolution than the result of duplication of an ancestral gene. For example, ZZ1 and 24 share 25% amino acid identity but negligible nucleotide sequence identity (data not shown). The 24 protein plays an interesting role in T4 head maturation. The T4 prohead consists of an outer shell (23) and an inner scaffold (22). When 24 is added to the prohead, maturation cleavage occurs: 23 loses its N-terminal portion, whereas 22 is degraded completely [16]. In addition, in phage KVP40, the duplication of three genes encoding proteins associated with the phage tail or the tail fibre (gp12, gp19, and gp37) suggests added flexibility in the range of host adaptation and the infection process [8]. In our study, duplication was observed in three putatively duplicated gp34 (CDS244) homologues (gp37, gp12, and gp36 compared to CDS247, CDS171, and CDS246, respectively), which encode tail fibre-associated proteins and share 29-30% amino acid identity with ZZ1 gp34. Overall, splitting and/or duplication followed by sequence diversification and new gene insertion might be a mode of T4-like phage evolution.

**Similarities and distinctions in the genome features of the* E. coli* phage T4 and the T4-like *Acinetobacter* phages**

Alignment of the ZZ1 genome with the T4-like phages in the Tulane database showed that ZZ1 was closely related to four *Acinetobacter* phages. Based on the total search score, the closest was Acj9, followed by Acj61, Ac42, and 133. No other *Acinetobacter* phages were uncovered. In DNA sequence dot plots, we observed a frequently interrupted but straight diagonal line between ZZ1 and the 4 phages (data not shown). Overall, their genomes are colinear but are frequently interrupted by replacements with unrelated genome segments of comparable lengths, especially in the first halves of their genome. Further comparison of these genome sequences (including T4) at the DNA level using Mauve showed remarkable synteny in five large conserved regions (>10,000 bp) (see Figure 5). The most conserved syntenic region covered the morphogenesis module as
well as the DNA replication and metabolism module, which are consistent with previous analyses [2,14].

Protein-by-protein comparison of the 5 phages using BLASTP and CoreGenes revealed that ZZ1 shares 179 (69.9%) protein homologues with *Acinetobacter* phage Acj9, 164 (64.1%) with Acj61, 157 (61.3%) with 133, and 143 (55.9%) with Ac42, although ZZ1 shares only 110 protein homologues with coliphage T4. In contrast, 101 protein homologues are present in T4 and in all 5 *Acinetobacter* phages, whereas 119 homologous genes are present in all 5 of the *Acinetobacter* phages (46.3 to 49.3% of their protein-coding genes), which is further evidence that ZZ1 and the 4 other *Acinetobacter* phages belong to the same genus (see Figure 3).

The full portrait of the T4 phage superfamily that has emerged from the many reported genomic comparisons [2,9-11] is that the T4 superfamily can be distilled down to core signature genes, which are known as the “Core Genome” of the T4-related phages or T4-like Viruses [2]. The Core Genome primarily includes homologues of essential T4 genes, such as the virion structure and DNA replication genes. In contrast, the hyper plastic regions (HPRs) contain mostly novel genes of unknown function and origin [7,14]. The Core Genome of the T4-related phages has been considered to consist of two genetic components: one highly resistant component was termed the Core genes, which are essential for all known conditions; and the other component, which is somewhat permissive to attrition in evolution, was termed the Quasicore genes and can be substituted or circumvented in certain genetic backgrounds of phage and/or bacterial hosts [2]. The genetic background for the Core Genome can vary considerably between T4 relatives, and thus, the number of the core genes that constitute the Core Genome depends on the precise subset of phages considered, which constitutes an important criterion for distinguishing between close and distant relatives [14]. A closer look at the components and organization of the Core Genome in the 6 T4-like phages (see Figures 3 and 4) shows that all 5 *Acinetobacter* phages share a Core Genome with T4 interrupted by several HPRs, which are where most of their divergence occurs. This finding is consistent with the previous observations of T4-like phages [2,7,11,13,14,17]. All of the Core genes described by Petrov et al. exist in all 5 T4-related *Acinetobacter* phages and are indicated by the red font directly above the coloured genes bars in Figure 4. The topology of the set of T4 phage Core genes is also shared by all 5 T4-related *Acinetobacter* phages. Moreover, we observed that almost all the Quasicore genes described by Petrov et al. (by the blue font directly above coloured genes bars in Figure 4) were shared by all 5 T4-related *Acinetobacter* phages, except for *uvSX* (RecA-like recombination protein), *uvS Y* (*uvS X* helper protein), *nrDC* (thioredoxin), *nrDG* (subunits of an anaerobic ribonucleotide reductase complex), *tk* (thymidine kinase), *49* (Endonuclease VII, required for recombination and DNA packaging), *cd* (dCMP deaminase), and *denA* (Endonuclease II). These genes are permissive to attrition in evolution of all 5 T4-related *Acinetobacter*
phages (see Figure 4). In addition, nearly half of the other T4 signature genes (indicated by the black font directly above coloured genes bars in Figure 4) are shared by all 5 T4-related Acinetobacter phages, including the gene topology. Thus, all 5 T4-related Acinetobacter phages use similar mechanisms to control propagation in their hosts. These phages all shared a common ancestral genome but, during the course of evolution, they have modified it in numerous ways. These conserved T4 genes allowed the 5 T4-related Acinetobacter phages to conserve their highly successful virion design and mode of replication. Notably, several major distinctive genome features in all 5 T4-related Acinetobacter phages (indicated by the green shading in Figure 4) could reflect a dominant feature in the evolution of all 5 T4-related Acinetobacter phages.

The most significant difference between the 5 T4-related Acinetobacter phages and T4 is in a series of HPRs, which are interspersed between the conserved T4 core sequences and especially enriched in the upper genome halves of all 5 T4-related Acinetobacter phages. These regions vary greatly in gene number and content. The HPRs were primarily composed of genes of unknown origin, but they do contain some identifiable sequences from bacteria and unrelated phages. Thus, the HPRs have been predicted to be the result of an evolutionary history of isolation within distinct hosts and extensive lateral gene transfer (LGT), i.e., importing genes or exchanges with diverse biological entities in nature [7,18]. Our observations showed that all of the genes lack T4 homologues in the 5 T4-related Acinetobacter phage genomes and cannot be assigned a putative function, most of which likely appear to be homologous genes from other phages, bacteria or other organisms. Notably, of the 119 ZZ1 genes that have significant homology with genes present in 4 of the Acinetobacter phages, 31 have unknown functions. Of these 31 genes, 18, which are indicated by the vertical grey arrows in Figure 3 (the counterparts of the 18 ZZ1 genes in the 4 other T4-related Acinetobacter phages are linked by dark or light purple shading in Figure 4), share no significant similarity to T4 proteins. Moreover, most of these homologous genes maintain a consistent order and content in the 5 T4-related Acinetobacter phages (see Figure 4). Further, BLASTP analyses suggested that 9 of the 18 ZZ1 genes (CDS042, CDS047, CDS049, CDS060, CDS125, CDS128, CDS138, CDS196, and CDS217), whose counterparts are linked with dark purple shading in Figure 4, lack significant matches to any of the phage sequences in GenBank except for the 4 other Acinetobacter phages (E value < 10^-5) and a few other organisms (E value > 0.1). In particular, 7 (CDS042, CDS047, CDS049, CDS060, CDS138, CDS196, and CDS217) share high similarities (>50%) with the other 4 Acinetobacter phages.

The diversity among phage “Pangenome” (the union of all the different naturally occurring ORFs that exist in T4-related phages [2]) is a reflection of the adaptations of a phage ancestor to a variety of evolutionary challenges, including encountering new host environments. LGT most certainly has played a role in such adaptations, and these adaptations may have been facilitated in part by certain particularities of the T4 phage recombination and gene expression systems [7,14]. However, to date, there are few clues about the agents that might mediate such a transfer. T4 recombines most efficiently early in the infection before the host genome is degraded, and this process could facilitate the acquisition of host genes by the phage [19]. As little as 50 bp of homology (perhaps less) is sufficient for T4 recombination system to recombine at reasonable frequencies [20]. Thus, “semilegitimate” recombination that relies on the small but generally conserved regulatory signals in the intergenic regions (promoters, translation initiation regions, transcription terminators, etc.) could perhaps have mediated the acquisition of such foreign genes [21]. Once the foreign genes are acquired, the T4 expression system could exploit their endogenous bacterial promoters because these sequences are very similar to the early promoter sequences of the T4-type phages [11,22]. Overall, different phages sharing a common ancestor undergo a specific and similar adaptation process for different bacterial strains that belong to the same genera or species. The existence of these highly conserved unknown homologous genes specific to Acinetobacter phages in HPRs could reflect the complex interactions of Acinetobacter phages with conserved cell components that are specific to Acinetobacter bacterial hosts and are distantly related to E. coli or other bacterial hosts. Theoretically, the uniqueness of homologous genes or certain sequences in specific phage genomes or lineages might help distinguish between the different clusters or types of phage and help predict the bacterial host range when treating the corresponding bacterial genera clinical isolates. Additional research is required to elucidate the highly conserved genes that are specific to Acinetobacter phages.

GC skew in the ZZ1 genome

Phage ZZ1 DNA contains only 34.4% GC, which is slightly lower than the value (38.9% to 39.2%) observed in sequenced A. baumannii strains [23-26]. To date, it is the lowest reported GC content for sequenced T4-like Acinetobacter phages (see Table 1). The GC content of ZZ1 is comparable with that of the Enterobacteria phage T4 (35.3%). The mol% GC of the T4 genome is also substantially lower than that of its host (approximately 50% GC) [3]. The difference is in disagreement with previous studies suggesting that the GC content of phage genomes such as the mycobacteriophages [27] and Staphylococcus aureus phages [28] have a GC content similar to their respective hosts.
Although ZZ1 has a lower GC content, 12 of the predicted phage ZZ1 genes contain more than 40% GC, including one major head gene (23 counterpart/CDS187), three tail fibre genes (12 counterpart/CDS171, 37 counterpart/CDS247, and 36 counterpart/CDS246), one nucleotide metabolism gene (cd counterpart/CDS221), and seven unknown genes. Notably, the GC content of the ZZ1 23 gene is 40.5%, which is significant lower than the local GC content of T4 23 (45%) and the T4-like phage JS98 23 (47.6%) [17]. Gene has the highest expression; approximately 1,000 copies of the protein are present in each) were located near the three adjacent genes, CDS150, 151, and 152. The two regions were in this CDS-devoid region (as shown in Figure 1). A previous study revealed that the phage ZZ1 had a very short latent period of 9 min [1]. The large number of tRNA genes and their strategic locations might enable the phage to translate its sequence more efficiently, reducing its latency time and increasing its reproduction rate and thus its infectivity.

**Codon usage and tRNAs**

Eight tRNA genes were predicted in the ZZ1 genome (see Table 2). The T4 genome also has 8 tRNA genes, whereas Ac42 has slightly fewer (6), and others have substantially more including phage 133 (16), Acj9 (19), and Acj61 (14), as shown in Figure 7. The different numbers and types of tRNAs present in the 5 *Acinetobacter* phages are consistent with a previous study of tRNAs in T4-like phages [7,13]. In addition, although ZZ1 has a low mol% GC content (34.4%), 4 (50%) out of the 8 tRNAs recognise codons with A in the third position (see Figure 7).

The exact function of phage-encoded tRNAs in the phage-infected cell is still not clear. They might be involved in the adaptation of the host translation apparatus to the demands of the phage codon usage pattern. In other words, the tRNAs overcome the phage codon usage problem, and the presence of tRNAs in a phage genome has been suggested to compensate for differences in codon usage between the phage and the host, corresponding to codons that are expected to be poorly translated by the host machinery [30]. Previous studies indicated that T4-encoded tRNAs are related to codons that are highly used in T4 genes but rarely used in the host and that the phage tRNAs can enhance the low expression of T4 late-stage protein genes through optimal codon usage in translation [3,31]. Similarly, analysis of the ZZ1-encoded tRNAs indicated that over half of them (5 out of 8) might be related to the optimal codon usage of ZZ1 proteins (Figure 7). However, the codon usage frequencies of the 4 other *Acinetobacter* phages suggested that phage-encoded tRNAs might be unrelated to decoding the relatively more frequent codons in all of the genes (Figure 7). Thus, the functional role of the tRNA genes for these phages remains unclear. Although the types and numbers of phage-encoded tRNAs differ between the 5 phages, some consistency was observed. For example, all 5 phages encode Met-tRNA and Trp-tRNA, which are related to optimal codon usage, but Arg-tRNA, which was not related to optimal codon usage, was also common to all 5 phages (see Figure 7). Of the 8 tRNA genes in ZZ1, 6 (75%) shared up to 89 and 90% nucleotide sequence similarity to tRNAs from phage Acj9 and phage 133, respectively; 4 (50%) shared up to 89% nucleotide sequence similarity to the tRNAs from phage Acj61; and only 1 tRNA showed 83% sequence similarity with the tRNA-Thr...
from phage Ac42 (see Table 2). In addition, the ZZ1-encoded tRNA-Thr gene shared 83% identity with the T4-encoded tRNA-Pro gene (see Table 2). The high degree of conservation of some of the tRNAs suggests an important functional role. The number and type of homologous tRNA appears to confirm the relationship between ZZ1 and the other phages, suggesting some elements of vertical evolution in these tRNAs.

**Mobile elements**

The homing endonuclease genes are not genuine phage DNA; rather, they belong to intron-associated selfish DNA elements [32,33]. There are 7 seg, 5 mob, and 3 intron nuclease (I-TevI to -III) genes in the T4 genome [3], which have efficiently invaded the T4 genome and are also frequently found in the DNA transaction module opposite the tail fibre cluster (I-TevI and segG) and the DNA replication module (segA, mobB, mobC, and I-TevII) (Figure 3). Inconsistent distribution of these elements has been described for T2, T4 and T6 phages [34]. One notable difference between the T4 and ZZ1 genomes is that ZZ1 appears to lack counterparts in the corresponding regions, except for two mobile genes, a seg-like gene and a mob-like gene. The ZZ1 seg-like mobile endonuclease (CDS142), which is located in the DNA replication module, has 34 to 54% similarity to 6 seg genes in T4 (segA, segB, segE, segC, segD, and segE). However, the ZZ1 seg-like gene has higher similarity with the seg counterparts of the other 4 Acinetobacter phages: 73% amino acid identity with the seg counterpart from Acj9, and 56 to 58% identity with seg counterparts from Acj61, 133 and Ac42. The ZZ1 mob-like homing endonuclease (CDS175), which was found in the cluster of neck genes, lacks sequence similarity with the corresponding T4 genes. Nevertheless it has 52% sequence similarity with the mob counterparts from Acj9 and 56 to 58% identity with seg counterparts from Acj61, 133 and Ac42. The ZZ1 mob-like homing endonuclease (CDS175), which was found in the cluster of neck genes, lacks sequence similarity with the corresponding T4 genes. Nevertheless it has 52% sequence similarity with the mob counterparts from Acj9 and 56 to 58% identity with seg counterparts from Acj61, 133 and Ac42. The ZZ1 mob-like homing endonuclease (CDS175), which was found in the cluster of neck genes, lacks sequence similarity with the corresponding T4 genes.
| Aminoacid | Codon usage (per 10^4) | Anticodon |
|-----------|------------------------|-----------|
| AcaAc2    | ZZ1                    | Ac61133   |
| Arg       | AGG0.49                | 0.63      | 0.41 0.82 0.51 |
|           | AGA5.67                | 6.77      | 2.39 3.10 2.14 |
|           | CGG0.87                | 1.59      | 0.67 2.96 1.40 |
|           | CGA9.48                | 7.82      | 11.74 8.78 11.40 |
|           | CGU20.1                | 19.27     | 22.63 19.70 19.50 |
|           | CGC4.31                | 5.20      | 4.7 9.00 8.95 |
| Leu       | UUG13.70               | 12.44     | 20.1 19.21 19.00 |
|           | UUA32.40               | 42.55     | 19.66 26.64 19.30 |
|           | CUG2.17                | 1.96      | 4.21 8.20 4.07 |
|           | CUA7.99                | 4.77      | 5.93 9.06 7.98 |
|           | CUC21.00               | 16.75     | 19.96 15.17 21.10 |
| Ser       | AGU11.00               | 10.53     | 8.32 9.88 6.66 |
|           | AGC4.89                | 4.03      | 5.14 6.02 6.88 |
|           | UCG5.57                | 4.33      | 5.67 6.00 6.88 |
|           | UCA18.90               | 22.01     | 24.13 19.88 22.26 |
|           | UCU24.50               | 18.35     | 16.50 14.41 18.00 |
|           | UCC2.46                | 2.26      | 1.40 4.92 2.52 |
| Ala       | GCG6.10                | 7.04      | 5.08 11.06 7.60 |
|           | GCA25.00               | 32.25     | 30.23 23.78 34.20 |
|           | GCC26.70               | 22.05     | 30.31 19.98 23.80 |
| Gly       | GGG4.02                | 4.14      | 2.15 8.10 5.71 |
|           | GGA12.00               | 9.78      | 14.25 11.65 12.80 |
|           | GGU29.60               | 32.64     | 30.19 26.33 23.50 |
|           | GGC13.00               | 11.47     | 14.31 14.72 18.50 |
| Pro       | CCG4.08                | 4.47      | 5.52 6.23 4.77 |
|           | CCA17.70               | 15.86     | 17.13 16.16 14.70 |
|           | CCU12.30               | 13.86     | 14.68 12.82 16.40 |
|           | CCC0.66                | 1.94      | 0.67 2.84 1.10 |
| Thr       | ACG4.23                | 4.62      | 4.90 7.49 5.69 |
|           | ACA25.10               | 28.24     | 24.48 19.41 25.39 |
|           | ACU26.68               | 24.18     | 25.57 24.02 22.50 |
|           | ACC4.46                | 6.08      | 5.61 11.27 8.25 |
| Val       | GUG7.67                | 6.37      | 7.37 13.90 9.61 |
|           | GUA21.20               | 21.19     | 17.53 20.23 17.30 |
|           | GUU31.30               | 22.39     | 30.02 25.51 26.50 |
| Gln       | GAG8.99                | 4.66      | 12.37 7.84 13.50 |
|           | GAA5.63                | 5.08      | 7.61 11.37 6.73 |
|           | GAG29.90               | 30.16     | 28.87 27.86 28.00 |
|           | GAC12.90               | 8.49      | 12.59 13.51 11.20 |
|           | GAA55.00               | 56.14     | 54.97 48.23 54.40 |
| Ilc       | UUA5.63                | 8.09      | 3.85 7.29 5.35 |
|           | AUU46.50               | 55.97     | 34.34 45.25 37.50 |
|           | UUC17.70               | 9.34      | 32.64 16.61 26.30 |
|           | UGA1.45                | 1.30      | 1.23 1.78 1.32 |
|           | UAG0.34                | 0.21      | 0.43 0.51 0.21 |
|           | UAA3.02                | 3.39      | 3.10 2.74 1.25 |
| Asn       | AAU36.70               | 48.30     | 31.44 36.13 30.80 |
|           | AAC19.60               | 10.12     | 24.68 19.06 23.20 |
|           | AAG12.42               | 49.70     | 38.63 41.73 39.90 |
|           | GAC19.40               | 11.89     | 21.16 19.35 21.40 |
| Cys       | UGU8.39                | 8.40      | 7.72 8.77 6.77 |
|           | UGC2.66                | 2.44      | 3.26 4.06 4.54 |
|           | UGG13.90               | 8.07      | 25.04 15.45 20.50 |
| His       | CAU13.20               | 15.90     | 12.63 14.49 11.80 |
|           | CAC5.74                | 3.34      | 6.56 7.31 8.49 |
|           | CAG18.70               | 12.60     | 20.41 19.14 18.70 |
| Lys       | AAG54.60               | 60.73     | 49.98 49.49 50.20 |
| Phe       | UUU31.50               | 34.48     | 19.33 25.45 24.10 |
|           | UUC13.90               | 8.07      | 25.04 15.45 20.50 |
| Tyr       | UAU26.70               | 34.25     | 22.84 27.23 21.30 |
|           | UAC18.50               | 6.65      | 16.92 13.49 16.30 |
| Met       | AUG26.40               | 26.56     | 27.29 28.74 27.90 |
| Trp       | UGG11.60               | 13.31     | 12.86 13.22 13.10 |

**Figure 7** (See legend on next page.)
few group I-like introns have been found in other phages, such as T4, the S. aureus phage Twort [35], the Bacillus phages I-BasI and I-Hmuf [36], the Lactobacillus phage LL-H, and the Lactococcus phage r1t [37]. However, none of the three T4 intron nuclease genes were found in phage ZZ1, Acj61, or Acj9, and only one I-TevI-like homing endonuclease from the GIY-YIG family was found in the 133 and Ac42 genome (see Figure 4). These mobile element genes do not commonly exist in the T4-like phages, and the differences in mobile elements between the 5 Acinetobacter phages and T4 phage genomes are consistent with the results for another T4-like phage [7,13]. Apparently, there are ill-defined barriers to intron promiscuity in bacteriophages [38,39].

Conclusions

We determined and revised the complete genome sequence of the phage ZZ1 that infects pathogenic A. baumannii strains. A total of 256 potential proteins and 8 tRNAs were predicted. The BLASTP analysis reveals that only a small portion of its proteins (110 proteins, 43%) are clearly related to coliphage T4 proteins and share up to 73% protein sequence identity with the corresponding T4 proteins. Further analysis revealed that 179, 164, 157, and 143 proteins from ZZ1 share up to 86, 85, 81, and 83% amino acid identity, respectively, with Acinetobacter phage Acj9, Acj61, 133, and Ac42 proteins, respectively. Nine ZZ1 genes lack significant matches to any of the phage sequences in the GenBank database except for the 4 other Acinetobacter phages. In addition, the high degree of conservation of some tRNAs between ZZ1 and the other T4-like phages suggests that they may have an important functional role in addition to overcoming the phage codon usage problem. The number and identity of the homologous tRNAs supports the evolutionary relationship between ZZ1 and the other phages, suggesting some elements of vertical evolution in these tRNAs. Overall, although more than 200 T4-like phages have been examined [4,13], only a very limited number of the T4-like Acinetobacter phage genomes have been explored or exploited. As additional genome sequencing of Acinetobacter phage species are completed, more Acinetobacter phages similar to T4 might be discovered. Host-specific adaptation mechanisms might be revealed by a more comprehensive understanding of the genomic diversity within the Acinetobacter bacteriophage population in the future.

Methods

Bacterial and phage strains

A. baumannii AB09V was isolated from the sputum of one hospitalised patient at the Henan Province People's Hospital in Zhengzhou, China. After obtaining the approval of the Life Science Ethics Committee of Zhengzhou University and written informed consent, sputum samples were collected for the purposes of this study. The automated system BD Phoenix (Becton Dickinson Diagnostic Systems, Sparks, MD, USA) was used on clinical samples for the identification of bacteria and for antibiotic susceptibility tests. Moreover, as the host of phage ZZ1, AB09V has been further confirmed as A. baumannii using sequence information derived from the 16S rRNA gene in our previous work [1]. Phage ZZ1 was propagated in the A. baumannii strain AB09V as previously described [1].

Phage genome resequencing

The nucleic acid from the phage ZZ1 was extracted and purified from phage lysate using a MiniBEST Viral RNA/DNA Extraction Kit Ver. 4.0 (TAKARA BIO Inc., Tokyo, Japan) according to the manufacturer’s protocol. Phage DNA was then sent to Zhejiang California International NanoSystems Institute for sequencing. Phage DNA was fragmented with a Covaris S220. After end-repair and ligation of adaptors, libraries were amplified by polymerase chain reaction (PCR), purified with a QiAquick® PCR extraction kit (Qiagen, Venlo, The Netherlands), and sequenced on an Illumina Solexa Sequencing platform (Illumina, San Diego, USA) with a read length of 2 × 250. The whole genome sequences of ZZ1, with a total length of 166,682 bp, were obtained [1]. However, when we analysed in further detail the entire ZZ1 genome to understand the genetic characteristics of this phage, we found many split genes, which were considered to be due to errors from the sequence read assembly. Thus, in this study, we amended these errors by PCR + sequencing directly from phage genomic DNA. Ultimately, the length of revised single copy ZZ1 genome was 166,687 bp.

Bioinformatics analyses

Open reading frames (ORFs) were identified using two bioinformatics software programs, GenMarkS [40] and fgenesV0 (http://linux1.softberry.com/berry.phtml). The predicted translational regions were corroborated by manual inspection. These CDSs were considered valid if
they possessed at least 30 amino acids (aa), showed a putative ribosome binding site (RBS) at a convenient distance, and began with an AUG, UUG, or GUG codon. The BLASTP program (version 2.2.15) [41] from NCBI was used to search for sequence similarity of the predicted CDSs against the NCBI nonredundant protein sequence database. Protein domain searches were conducted through BLAST and by using Batch CD-Search [42]. Cumulative GC skew was measured with GenSkew at http://genskew.csb.univie.ac.at/. ZZ1 phage codon usage was analysed with the CUSP and CAI programs of the EMBOSS package, version 6.2.0 [43]. tRNAs were identified using the tRNAscan-SE server [44] and confirmed using the ARAGORN program [45] as well as by nucleotide comparison using BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi). TMpred was used to identify membrane-spanning regions in CDSs (http://www.ch.embnet.org/software/TMPRED_form.html).

Comparative genomics
The phage genome sequences downloaded from NCBI and the Tulane T4-like genome (http://phage.bioc.tulane.edu) are listed in Table 1. Basic genome features of all of the phages, such as their host specificity, taxonomy, and genome molecular type, were also collected from the NCBI database and the literature. The dot matrix view from NCBI was used to examine whole-genome similarity using the default parameters. Further genomic comparisons at the nucleotide level were made with Mauve 2.2.0 [46] using a progressive alignment with default settings. Comparisons at the proteomic level were made using CoreGenes (http://binf.gmu.edu:8080/CoreGenes2.0/custdata.html) [47].

Availability of supporting data
The ZZ1 sequence data were deposited at GenBank under accession number NC_018087.3, which replaces the previous accession number. The CDS prefix indicates a predicted protein coding sequence, followed by numbers specifying the locus tag in the GenBank file (i.e., CDS001 is locus ZZ1p0001). The GenBank accession numbers of the other phage genomes analysed in this study are listed in Table 1.

Additional file

Additional file 1: CDSs of the ZZ1 genome and best BLASTP hits.

Abbreviations
CDS: Coding sequence; HPR: Hyper plastic region; LGT: Lateral gene transfer; ORF: Open reading frame; RBS: Ribosome binding site.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JJ conceived the study and drafted the manuscript. JJ, ZJL, SWM, DHH, SWW, SJC, GZ, YHL, JW, and XTW participated in the genome analysis. GQZ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This study was financially supported by funding from a Key Project of the Science and Technology Development of Henan Province Science and Technology Bureau (No. 142102310440), the Education Department of Henan Province (No. 2011C310014), and a Project of the Science and Technology Development of Henan Province (No. 201303199).

Author details
1 Department of Pathogen Biology and Immunology, Henan Medical College, Shuanghu Road #100, Zhengzhou 450001, P. R. China. 2 Department of Pathogen Biology, Basic Medical College of Zhengzhou University, Kexue Road #100, Zhengzhou 450001, P. R. China. 3 Clinical Laboratory, Henan Provincial People’s Hospital, Zhengzhou 450003, P. R. China.

Received: 6 February 2014 Accepted: 10 September 2014
Published: 14 September 2014

References
1. Jin J, Li ZJ, Wang SW, Wang SM, Huang DH, Li YH, Ma YY, Wang J, Liu F, Chen XD, Li GX, Wang XT, Wang HQ, Zhao GQ: Isolation and characterization of ZZ1, a novel lytic phage that infects Acinetobacter baumannii clinical isolates. BMC Microbiol 2012, 12(1):156.
2. Petrov VM, Ratnayaka S, Nolan JM, Miller ES, Karam JD: Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Viral J 2010, 7:292.
3. Miller ES, Kutter E, Mosig G, Ariaifa F, Kuniawa S, Ruge R: Bacteriophage T4 genome. Microbiol Mol Biol Rev 2003, 67(1):156–186.
4. Tetart F, Desplats C, Kutateladze M, Monod C, Ackermann HW, Krisch HM: Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J Bacteriol 2001, 183(1):355–366.
5. Ackermann HW, Krisch HM: A catalogue of T4-type bacteriophages. Arch Virol 1997, 142(12):2329–2345.
6. Arbiol C, Comeau AM, Kutateladze M, Adaramia R, Krisch HM: Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. Genome Biol Evol 2010, 2:140–152.
7. Comeau AM, Bertrand C, Letarov A, Tetart F, Krisch HM: Modular architecture of the T4 phage superfamilies: A conserved core genome and a plastic periphery. Virolology 2007, 362(2):384–396.
8. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldbyrum TW, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM: Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J Bacteriol 2003, 185(17):5220–5233.
9. Petrov VM, Nolan JM, Bertrand C, Levy D, Desplats C, Krisch HM, Karam JD: Plasticity of the gene functions for DNA replication in the T4-like phages. J Mol Biol 2006, 361(1):46–68.
10. Mann NH, Cibele MR, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM: The genome of S-PM2, a “photosynthetic” T4-type bacteriophage that infects marine Synechococcus strains. J Bacteriol 2005, 187(9):3188–3200.
11. Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD: Genetic diversity among five T4-like bacteriophages. Virol J 2006, 3:30.
12. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW: Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol 2005, 3:e144.
13. Liao WC, Ng WY, Lin HL, Syu WJ, Liu TT, Chang CH: T4-like genome organization of the Escherichia coli O157:H7 lytic phage AR1. J Virol 2011, 85(13):6567–6578.
14. Desplats C, Krisch HM: The diversity and evolution of the T4-type bacteriophages. Res Microbiol 2003, 154(6):259–267.
15. Brussow H, Desere F: Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol 2001, 39(2):213–222.
16. Chibani-Chennoufi S, Sidoti J, Bruttn C, Dillmann ML, Kutter E, Cadri F, Sarker SA, Brussow H: Isolation of Escherichia coli bacteriophages from the stools of pediatric diarrhea patients in Bangladesh. J Bacteriol 2004, 186(20):8287–8294.
17. Zuber S, Ngom-Bru C, Barretto C, Bruttin A, Brussow H, Denou E. Genome analysis of phage JS98 defines a fourth major subgroup of t4-like phages in Escherichia coli. J Bacteriol 2007, 189:8206–8214.

18. Fileje J, Baptiste E, Susko E, Krisch HM. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol Biol Evol 2006, 23(9):1688–1696.

19. Krisch HM, Hamlett NV, Berger H. Polynucleotide ligase in bacteriophage T4D recombination. Genetics 1972, 72(2):187–203.

20. Singer BS, Gold L, Gauss P, Doherty DH. Determination of the amount of homology required for recombination in bacteriophage T4. Cell 1982, 31(1):25–33.

21. Repoila F, Tetart F, Eleaume H, Krisch HM. Snapshot of the genome of the pseudo-T4-type bacteriophage RB49. J Bacteriol 2002, 184(10):2789–2804.

22. Desplats C, Dez D, Tetart F, Eleaume H, Krisch HM. Complete genome sequence of multidrug-resistant Acinetobacter baumannii strain 1656–22, which forms sturdy biofilm. J Bacteriol 2011, 193(22):6393–6394.

23. Liu S, Wang Y, Xu J, Li Y, Guo J, Ke Y, Yuan X, Wang L, Du X, Wang Z, Huang L, Zhang N, Chen Z. Genome sequence of an OXA23-producing, carbapenem-resistant Acinetobacter baumannii strain of sequence type ST75. J Bacteriol 2012, 194(21):6001–2000.

24. Gan HM, Lean SS, Suhaili Z, Thong KL, Yeo CC. Identiﬁcation of tmRNA genes in nucleotide sequences. Nucleic Acids Res 2004, 32(5):955–964.

25. Liou ML, Liu CC, Lu CW, Hsieh MF, Chang KC, Kuo HY, Lee CC, Chang CT, Yang CY, Tang CY. Genome sequence of Acinetobacter baumannii TYTH-1. J Bacteriol 2012, 194(24):6974.

26. Liu S, Wang Y, Xu J, Li Y, Guo J, Ke Y, Yuan X, Wang L, Du X, Wang Z, Zhang H, Li G, Zhang N. Complete genome sequence of an OXA23-producing, carbapenem-resistant Acinetobacter baumannii strain of sequence type ST75. J Bacteriol 2012, 194(21):6000–6001.

27. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Panunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan J, Bardarov S, Kovic J, Lawrence JG, Jacobs WR Jr, Hendrix RW, Hafall GF. Origins of highly mosaic mycobacteriophage genomes. Cell 2003, 113(2):171–182.

28. Kwan T, Liu J, Dubrow M, Gros P, Pelletier J. The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. Proc Natl Acad Sci U S A 2009, 106(14):5174–5179.

29. Mackiewicz P, Zakrzewska-Czwerinska J, Zawialik A, Dudek MR, Cebas S. Where does bacterial replication start? Rules for predicting the oriC region. Nucleic Acids Res 2004, 32(13):3781–3791.

30. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages. Genome Res 2007, 17(10):1486–1495.

31. Kunisawa T. Synonymous codon preferences in bacteriophage T4: a distinctive use of transfer RNAs from T4 and from its host Escherichia coli. J Theor Biol 1992, 153(3):267–268.

32. Bell-Pedersen D, Quirk S, Clyman J, Belfort M. Intron mobility in phase T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res 1990, 18(13):3763–3770.

33. Eddy SR, Gold L. The phase T4 mbrd intron: a deletion mutant of a version found in the wild. Genes Dev 1991, 5(6):1032–1041.

34. Quirk SM, Bell-Pedersen D, Tomaschewski J, Ruger W, Belfort M. The inconsistent distribution of introns in the T-even phages indicates recent genetic exchanges. Nucleic Acids Res 1989, 17(1):301–315.

35. Landthaler M, Shub DA. Unexpected abundance of self-splicing introns in the genome of bacteriophage T20r: introns in multiple genes, a single gene with three introns, and exon skipping by group I introns. Proc Natl Acad Sci U S A 1999, 96(12):7005–7010.

36. Landthaler M, Shen BW, Stoddard BL, Shub DA. I-Hasl and I-Hml: two phase intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. J Mol Biol 2006, 358(4):1137–1151.

37. van Sinderen D, Karsens H, Kok J, Terpstra P, Ruiters MH, Venema G, Nauta A. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage rlt. Mol Microbiol 1996, 21(13):343–355.

38. Edgell DR, Belfort M, Shub DA. Barriers to intron promiscuity in bacteria. J Bacteriol 2000, 182(19):5281–5289.

39. Foley S, Bruttin A, Brussow H. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of Streptococcus thermophilus bacteriophages. J Virol 2000, 74(2):611–618.

40. Beesemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 2001, 29(12):2607–2618.

41. Belsham R, Katozaurakis A. BlastAlign: a program that uses blast to align problematic nucleotide sequences. Bioinformatics 2005, 21(1):122–123.

42. Marchler-Bauer A, Lu S, Anderson JB, Chitoua F, Derbyshire MK, deWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Grunewald M, Hurvitz DI, Jackson JD, Ke Z, Lanzucyki CJ, Lu F, Marchler GH, Mulkokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 2011, 39(Database issue):D225–229.

43. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000, 16(6):276–277.

44. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997, 25(1):955–964.

45. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 2004, 32(11):1–16.

46. Darling AC, Mau B, Blattner FR, Perna NT. MAUVE: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004, 14(7):1349–1356.

47. Zafar N, Mazumder R, Seto D. CoreGenes: a computational tool for identifying and cataloging “core” genes in a set of small genomes. BMC Bioinformatics 2002, 3:12.