Evidence for shared ancestry between Actinobacteria and Firmicutes bacteriophages

Matthew Koert, Courtney Mattson, Steven Caruso & Ivan Erill

Department of Biological Sciences, University of Maryland Baltimore County (UMBC), Baltimore, MD (USA)

Abstract

Bacteriophages are known to display a broad range of host spectra, typically infecting a small set of related bacterial species. The transfer of bacteriophages between more distant clades of bacteria has often been postulated, but remains mostly unaddressed. In this work we leverage the sequencing of novel cluster of phages infecting *Streptomyces* bacteria and the availability of large numbers of complete phage genomes in public repositories to address this question. Using phylogenetic and comparative genomics methods, we show that several clusters of Actinobacteria-infecting phages are more closely related between them, and with a small group of Firmicutes phages, than with any other Actinobacteriophage lineage. These data indicate that this heterogeneous group of phages shares a common ancestor with well-defined genome structure. Analysis of genomic %GC content shows that these Actinobacteriophages are poorly adapted to their Actinobacteria hosts, suggesting that this phage lineage originated in an ancestor of the Firmicutes, adapted to high %GC content members of this phylum and later migrated to the Actinobacteria.

Introduction

Frequently referred to as phages, bacteriophages are viruses capable of infecting bacteria. It has been estimated that phages are the most abundant entities in the biosphere [1] and, through their regulation of bacterial populations, bacteriophages play an essential role in many global processes of the biosphere, such as carbon and nitrogen cycling [2]. In the last decade, decreasing sequencing costs have dramatically increased the number and diversity of bacteriophage genome sequences [3]. This influx of phage genomic data has reinforced the notion that phages are not only key players in geobiological processes, but also the largest reservoirs of genetic diversity in the biosphere [4]. The Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) has undertaken a sustained effort to isolate and sequence phages infecting Actinobacteria species [3]. Among these, Mycobacteria-infecting phages have been studied the most, providing a remarkably deep

sample of bacteriophages infecting a given bacterial genus [3]. Studies of genetic diversity in over 600 Mycobacteria-infecting phage genomes have revealed extensive mosaicism, and genetic exchange among relatively distant groups of Mycobacteriophages. Rarefaction analyses suggest that the Mycobacteriophage gene pool is not an isolated environment, and that it is enriched by an influx of genetic material from outside sources [5]. Here we report on the genomic characterization of a new cluster of *Streptomyces* phages (Cluster BI). Gene content and protein sequence phylogenies indicate that members of BI and related Actinobacteriophage clusters share a common ancestor with *Lactococcus* and *Faecalibacterium* phages [6,7]. Analysis of genomic %GC content indicates that these Actinobacteriophages are still undergoing amelioration, suggesting that they may have originated as a result of an interphylum migration event from related Firmicutes phages.

Materials and Methods

Genome data

Genomes for relevant *Streptomyces* phages and for reference Actinobacteria and Firmicutes bacteriophages were retrieved in GenBank format from the NCBI GenBank database [8] using custom Python scripts. These scripts also derived nucleotide and amino acid FASTA-formatted files from the GenBank records, and autonumerically reassigned *locus_tag* and *gene* GenBank identifiers for consistent pham annotation with PhamDB. For phages without a public GenBank record, nucleotide FASTA files were downloaded from PhagesDB [3] and auto-annotated with DNA Master [9] to generate a GenBank-formatted file. %GC content data was obtained from the corresponding NCBI assembly records. Group %GC content was compared using a Mann-Whitney U test with $\alpha=0.05$ using a custom Python script and the scipy.stats module.

Gene content phylogeny

PhamDB was used to compute protein families, or phams, for the bacteriophage genomes under analysis [10]. The PhamDB-generated database was then imported into Phamerator [11] and the resulting pham table was exported as a comma-separated file and processed with spreadsheet software and the Janus program (Lawrence Lab) to obtain a Nexus-format file with presence/absence of each pham in each genome as a binary character. This Nexus file was used as input for SplitTree [12]. Network and tree phylogenies were inferred with the NeighborNet and BioNJ algorithms using a gene content distance [13] and branch support for the resulting phylogeny was estimated from 1,000 bootstrap pseudoreplicates. A genome-based phylogeny

was generated with the VICTOR webservice [14]. Intergenomic protein sequence distances were computed with 100 pseudo-bootstrap replicates using the Genome-BLAST Distance Phylogeny (GBDP) method optimized (distance formula d_6) for prokaryotic viruses [14,15] and a minimum evolution tree was computed with FASTME on the resulting intergenomic distances [16].

Protein sequence phylogeny

A profile Hidden Markov Model (HMM) of terminase protein sequences was built with HMMER (hmmbuild) using a ClustalW multiple sequence alignment of all annotated terminase, TerL or terminase large subunit sequences in the genomes under analysis [17,18]. This profile HMM was used to search (hmmsearch) the protein FASTA file derived from each genome with a cutoff e-value of 10^{-3}. Putative terminase sequences identified by the profile HMM were aligned with ClustalW using default parameters. Tree inference was performed on the resulting multiple sequence alignment using the BioNJ algorithm with a Gamma distribution parameter of 1 and the Jones-Taylor-Thornton substitution model, and branch supports were estimated from 1,000 bootstrap pseudoreplicates [19].

Results

Conserved architecture of BI cluster Streptomyces *phage genomes*

In the last few years, our group has characterized and sequenced several *Siphoviridae* bacteriophages capable of infecting *Streptomyces scabiei* RL-34 [20]. Genomic analysis indicated that these bacteriophages belong to the BI cluster, which also encompasses bacteriophages isolated by other teams on different *Streptomyces* hosts, such as *Streptomyces lividans* JI1326 (*Streptomyces* phage Bing) or *Streptomyces azureus* NRRL B-2655 (*Streptomyces* phage Rima). Cluster BI phages have linear genomes ranging from 43,060 to 57,623 bp, encompassing from 55 to 91 protein coding genes and no predicted tRNA genes. Comparative analysis of these bacteriophage genomes (Figure 1) reveals nucleotide sequence conservation to be predominant only in the virion structure and assembly genes module, which presents a genetic arrangement consistent with that observed in other *Siphoviridae* [21]. Within this module, the terminase gene shows the highest degree of sequence conservation, followed by segments of the portal, capsid maturation and tape measure protein coding genes (Figure 1). Beyond the structure and assembly module, moderate nucleotide sequence conservation is only observed for the genes coding for a predicted hydrolase in the lysis module, and for the DNA

primase/polymerase and an helix-turn-helix (HTH) domain-containing protein in the replication module.

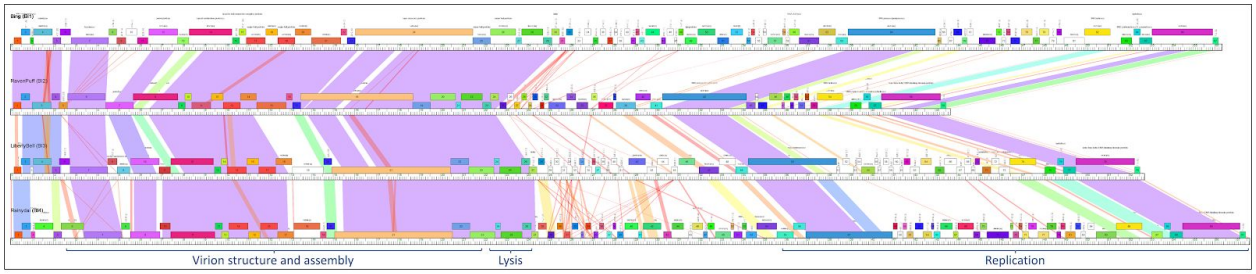


Figure 1 - Phamerator-generated map of four representative BI cluster *Streptomyces* phage genomes (Bing (BI1), RavenPuff (BI2), LibertyBell (BI3) and Rainydai (BI4)). Shaded areas between genomes indicate nucleotide similarity. Groups of orthologous protein sequences are denoted by colored boxes. Orphams (proteins in a pham containing a single member) are shown as white boxes.

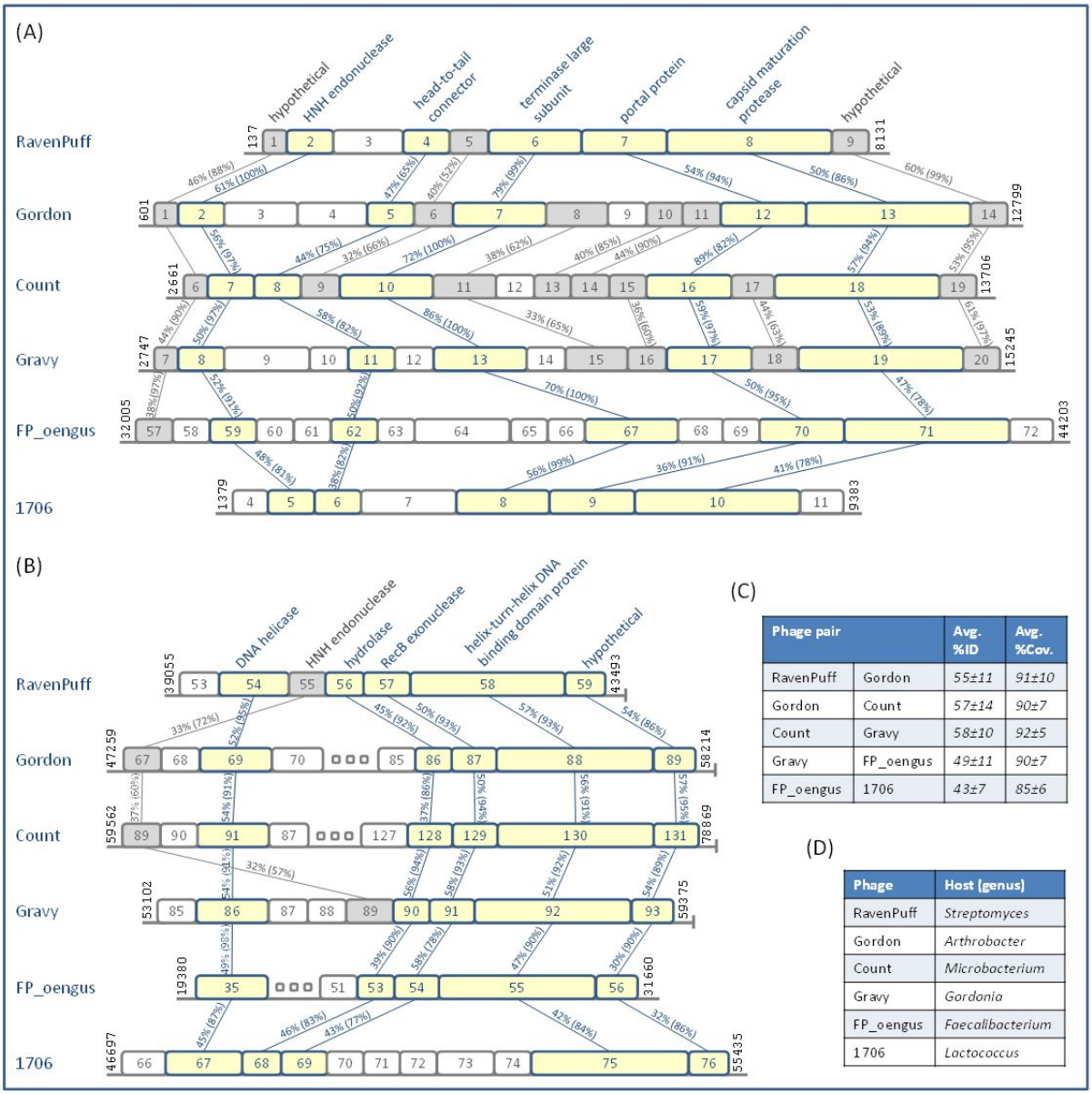
Interphylum conservation of structure and replication proteins

Functional annotation of BI cluster genomes was performed using BLASTP searches against both the NCBI GenBank and the PhamDB databases, as well as the HHpred service [22,3,8,20]. During the annotation process, BI cluster protein sequences frequently elicited significant hits against *Arthrobacter* (clusters AM, AU and AW), *Gordonia* (cluster DJ), *Rhodococcus* (cluster CC) and *Microbacterium* (cluster EL) bacteriophages, rather than against other *Streptomyces* phage clusters. It was noticed also that BLASTP searches against NCBI bacterial genomes often returned significant hits against putative prophages in several Firmicutes genomes. This prompted us to search for potential homologs of BI cluster proteins in the genomes of bacteriophages isolated from Firmicutes hosts, and we identified several *Lactococcus lactis* bacteriophage genomes related to *Lactococcus* phage 1706 [5,6] and a *Faecalibacterium* phage (FP_oengus) harboring multiple homologs of BI cluster proteins.

To contextualize this finding, we compiled complete genome sequences of bacteriophages in all the aforementioned clusters and in the *Lactococcus* and *Faecalibacterium* group, as well as reference members from other *Streptomyces*, *Arthrobacter*, *Gordonia* and *Rhodococcus* clusters, reference Firmicutes phages (e.g. *Staphylococcus* virus Twort, *Bacillus* virus SPO1, *Lactococcus* phage P335, *Leuconostoc* phage 1-A4, *Bacillus* phage Bam35c) and other bacteriophages identified by BLASTP as containing proteins with significant similarity to BI cluster proteins. Using PhamDB and Phamerator, we generated a table orthologous protein

sequence groups (phams) across this heterogeneous set of bacteriophage genome sequences (Table S1). A quick assessment of predicted phams revealed that the phams with the largest number of members within this dataset clearly outlined a supercluster of Actinobacteriophages encompassing *Arthrobacter* (clusters AM, AU and AW), *Gordonia* (cluster DJ), *Rhodococcus* (cluster CC), *Microbacterium* (cluster EL) and *Streptomyces* (cluster BI) phages. Importantly, 10 out of the 11 phams that are present in all these 41 Actinobacteriophages were also found in the *Lactococcus* and *Faecalibacterium* group. Overall, the Actinobacteriophage supercluster shared 27 large phams (27.6 $\pm_{SD}$ 17.8 members) with the *Lactococcus* and *Faecalibacterium* phage group, and 9 of the 15 largest phams were shared between both groups (Table S1). In contrast, *Lactococcus* and *Faecalibacterium* phages did not present any shared phams with the putatively related *Lactococcus* phage P335 [6], and they only shared five small phams (2 members) with reference Firmicutes phages. Likewise, the identified Actinobacteriophage supercluster only shared 35 small phams (6.0 $\pm_{SD}$ 3.1 members) with other Actinobacteriophages.

Graphical analysis of the genomic distribution of orthologs spanning both the Actinobacteriophage supercluster and the *Lactococcus* and *Faecalibacterium* phages (Figure 2) revealed that most of the orthologous genes were contained within two conserved regions at opposite ends of the genome. The first conserved region encompasses a sizable fraction of the virion structure and assembly genes module seen in BI cluster phages, containing a HNH endonuclease, a head-to-tail connector, the terminase large subunit, the portal protein and a capsid maturation protease (Figure 2A). The second conserved region corresponds to the end of the replication module observed in cluster BI phages and contains a DNA helicase, a HNH endonuclease, a RecB exonuclease, the HTH domain-containing protein and a conserved hypothetical protein (Figure 2B). Pairwise amino acid identity and alignment coverage for conserved orthologs among Actinobacteriophages were moderately high (56% $\pm_{SD}$ 12 and 91% $\pm_{SD}$ 7), and remained surprisingly high between *Gordonia* phage Gravy and *Faecalibacterium* phage FP_oengus (49 $\pm_{SD}$ 11 and 90 $\pm_{SD}$ 7), suggesting a relatively close evolutionary relationship.



Phage pair		Avg. %ID	Avg. %Cov.
RavenPuff	Gordon	*55±11*	*91±10*
Gordon	Count	*57±14*	*90±7*
Count	Gravy	*58±10*	*92±5*
Gravy	FP_oengus	*49±11*	*90±7*
FP_oengus	1706	*43±7*	*85±6*

Phage	Host (genus)
RavenPuff	*Streptomyces*
Gordon	*Arthrobacter*
Count	*Microbacterium*
Gravy	*Gordonia*
FP_oengus	*Faecalibacterium*
1706	*Lactococcus*

Figure 2 - Comparative analysis on representative genomes of the main genomic regions (A and B) containing conserved orthologs. Shaded boxes indicate orthologs conserved in at least two (grey) or in all the species shown (yellow), with the numbers across the lines connecting them showing the pairwise amino acid identity and alignment coverage. Gene numbers and genomic positions are provided for reference in each genome. (C) Average pairwise amino acid similarity and alignment coverage for orthologs conserved across all species. (D) List of representative phages and their host genera.

Shared ancestry between Actinobacteria and Firmicutes phages

The presence of two genomic regions showing substantial numbers of orthologous genes across a group of Actinobacteriophages infecting multiple hosts and a small set of Firmicutes phages

strongly pointed to an evolutionary relationship among these phages. To validate and examine this hypothesis, we used SplitsTree to infer the neighbor tree and estimate bootstrap support for the splits. The results (Figure 3, Figure S1, Data S1) show consistent branching (99.9% bootstrap support) of the Actinobacteriophage supercluster with both *Lactococcus* and *Faecalibacterium* phages, clearly establishing that these Firmicutes phages and the Actinobacteriophage supercluster phages share more gene content with each other than with reference Actinobacteriophages and Firmicutes phages. To further validate and support this result, we performed phylogenetic inference on the protein sequence of the large terminase subunit (Figure 4), a very common marker for bacteriophage phylogenetic analysis [23–26]. The inferred tree also shows solid support (100% bootstrap support) for a joint branching of the Actinobacteriophage supercluster phages and *Lactococcus* and *Faecalibacterium* phages, giving further credence to the notion that these phages share a common ancestor. Identical support for the joint branching of the Actinobacteriophage supercluster phages and *Lactococcus* and *Faecalibacterium* phages was obtained through independent phylogenetic inference using a bootstrapped minimal evolution algorithm operating on intergenomic protein sequence distances inferred from pairwise genome-wide reciprocal tBLASTX (Figure SX).

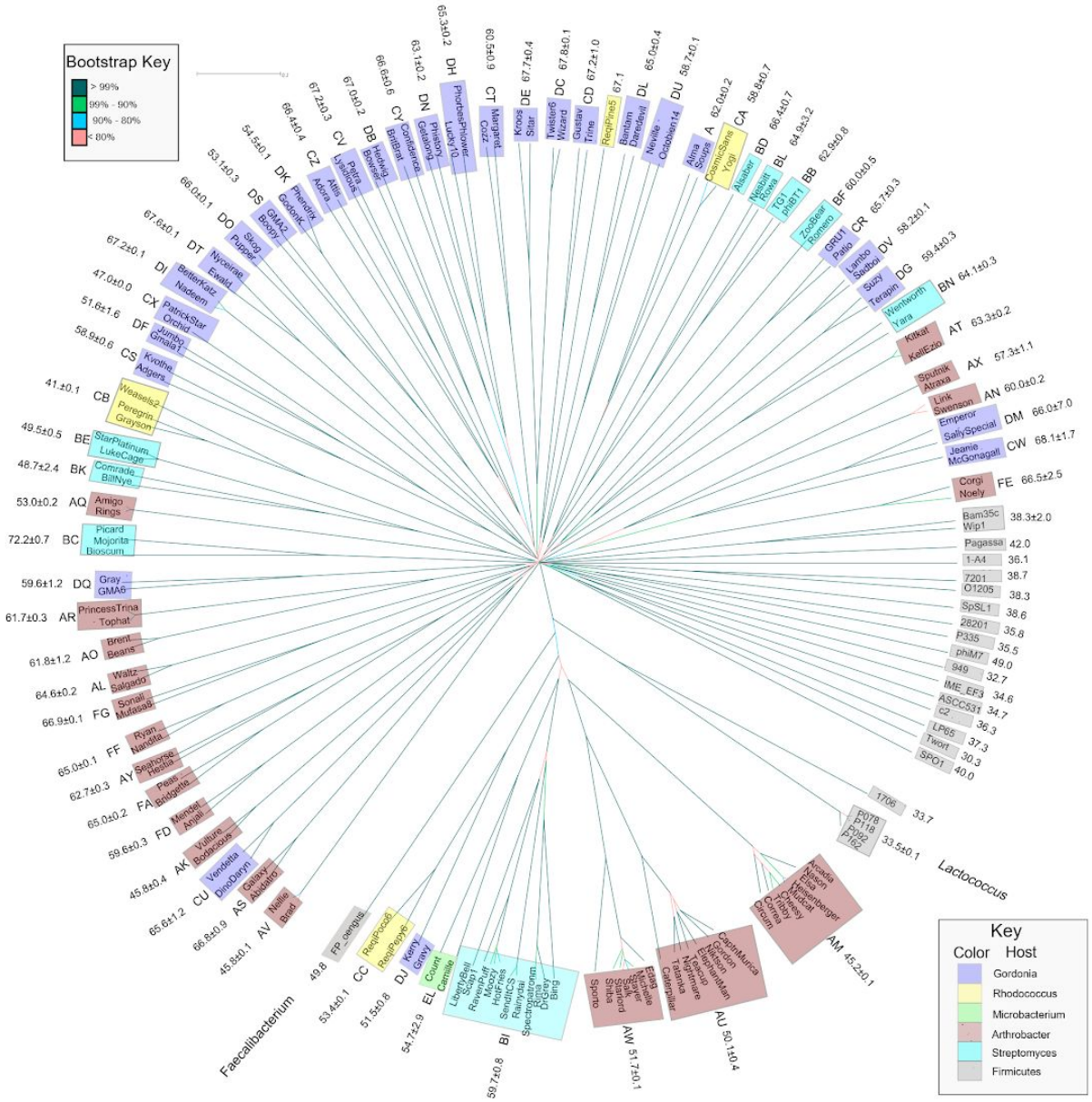


Figure 3 - BioNJ tree for analyzed phages. Bootstrap branch supports for 1,000 pseudoreplicates are shown as percent values on branches. Average genomic %GC content values are shown for different phage groups. Where available, cluster names are also indicated.

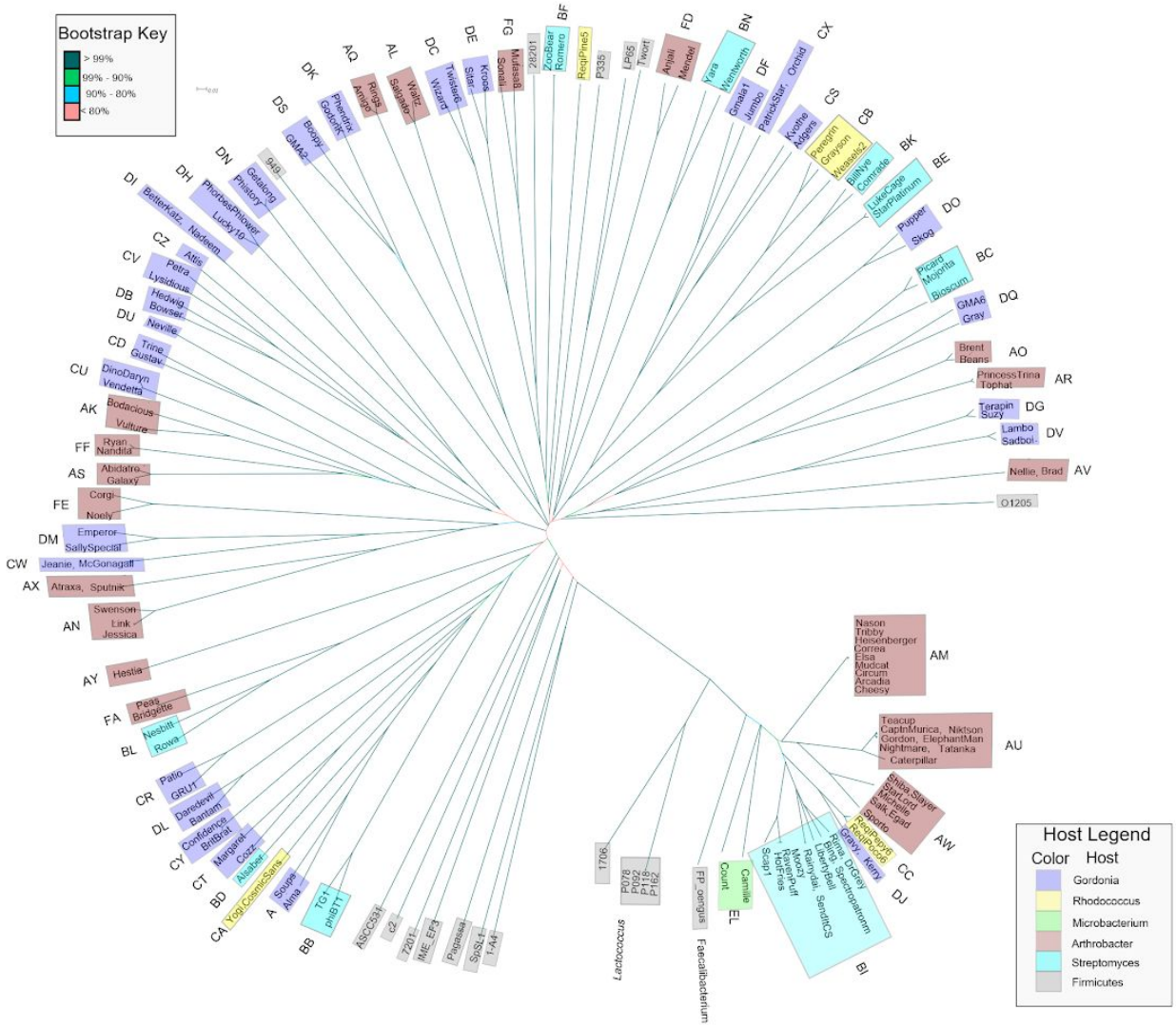


Figure 4 - Neighbor Joining (BioNJ) tree for the large terminase subunit protein sequences. Bootstrap branch supports for 1,000 pseudoreplicates are shown as percent values on branches.

Divergence in %GC content between bacteriophages and their hosts

We analyzed the %GC content of bacteriophage genomes to assess their alignment with the genomic %GC content of their hosts. The results (Figure 5, Table S2) show that the average %GC content of the Actinobacteriophage supercluster is significantly lower (20-30% lower, $p<0.05$, Mann-Whitney U test) than that of their hosts and also significantly lower than the average %GC content of other phage clusters infecting the same genera. This is also true for *Faecalibacterium* phage FP_oengus and *Lactococcus lactis* phages, although the difference in %GC content between phages and hosts is much smaller (~12%). Besides the members of the here identified

supercluster, several other Actinobacteriophage clusters (most notably AV, CX, BK, BE and CB) also present %GC content that is significantly lower than the one observed in their natural hosts and than the average for phage clusters infecting the respective genera.

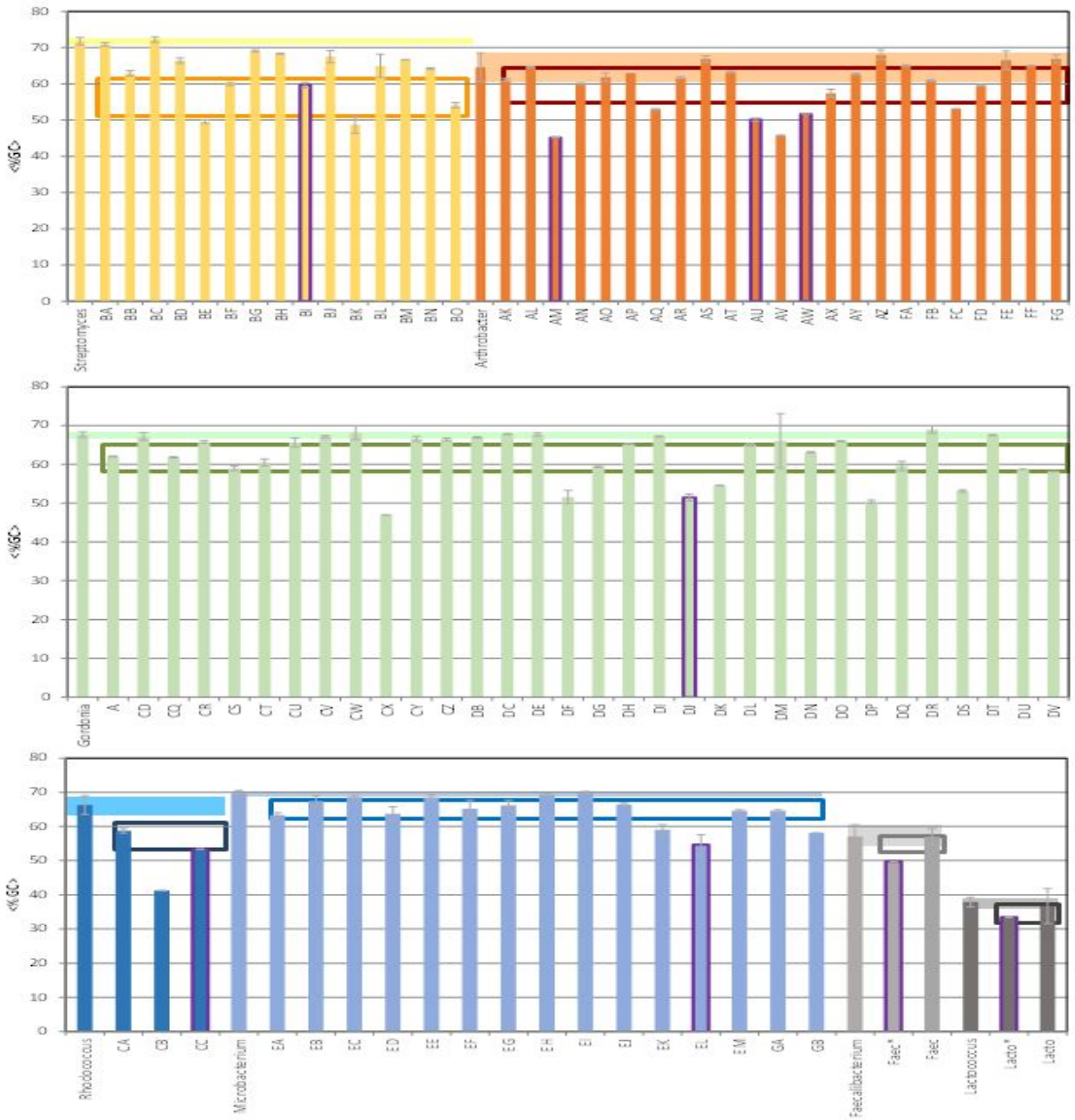


Figure 5 - Average %GC of phage cluster genomes and of complete genomes from each cluster host genus (using one representative per species with available complete genome). Bars for Actinobacteriophage supercluster clusters (BI, AM, AU, AW, DJ, CC and EL) and for *Faecalibacterium* and *Lactococcus* clusters(Faec and Lacto) are highlighted. For comparison, the mean ± standard deviation for %GC content estimated on the host genomes is overlaid with a filled box on the phage cluster bars and the mean ± standard deviation for %GC content among phages infecting each

genus is overlaid as an empty box. The %GC content for *Faecalibacterium* hosts was estimated from one presentative per species with whole genome shotgun assemblies.

Discussion

It is well known that bacteriophages will often infect several different hosts within the same bacterial genus, and that this host range can vary widely among phages within a given genus. As a consequence, it has been postulated that the intragenera host–phage interaction network is nested, with generalist phages infecting multiple hosts and specialist phages infecting particularly susceptible strains [27]. In contrast, relatively little is known about the ability of bacteriophages to infect across genera or broader taxonomic spans. Phage systems have been engineered to transcend genus boundaries [28] and effective transfer of virus-like particles has been documented across phyla [29], but the occurrence in a natural setting of infections across distantly related bacterial groups has not been demonstrated. The recent availability of a significantly large amount of complete bacteriophage genomes infecting a wide variety of bacterial hosts provides an opportunity to explore the genetic relationship among bacteriophages infecting distantly related hosts.

The identification of unexpected sequence similarity between orthologous protein sequences of phages infecting distantly related bacterial hosts within the Actinobacteria and the Firmicutes phyla led us to systematically explore their phylogenetic relationship. Both the gene content and terminase protein sequence phylogenies reported here (Figure 3 and Figure 4) indicate that Actinobacteriophages infecting hosts from five different bacterial families (*Gordoniaceae*, *Mycobacteriaceae*, *Micrococcaceae*, *Microbacteriaceae* and *Nocardiaceae*) in two bacterial orders (Corynebacteriales and Micrococcales) are more closely related to each other than to any other sequenced phage infecting their respective hosts, forming a host-heterogeneous supercluster. Furthermore, phylogenetic analyses also reveal that these Actinobacteriophages are closely related to phages infecting two different Firmicutes orders (Lactobacillales and Clostridiales) that, in turn, are more closely related to the Actinobacteriophage supercluster than to other Firmicutes-infecting phages. This close evolutionary relationship is mostly driven by the conservation of two large genomic blocks involving replication and structural proteins (Figure 2), suggesting that these are primary functional elements for this heterogeneous group of phages. This hypothesis is consistent with the observation of substantial variability in the intervening region between both blocks among the closely related BI cluster phages (Figure 1).

Analysis of genomic %GC content in this group of related Actinobacteria and Firmicutes phages reveals that their %GC content is systematically lower than that of their host genera and than that of similar phages infecting those genera. While the difference in %GC content between phages and their hosts is relatively small for *Lactococcus* and *Faecalibacterium* phages (~13%) it becomes much larger for Actinobacteria phages and hosts (20-30%). This indicates that Actinobacteria phages are still in the process of ameliorating their %GC content. In conjunction with the inferred phylogenies, the %GC analysis results suggest that this heterogenous group of related phages might have originated in a Gram-positive host, possibly related to *Lactococcus*, and spread first to high %GC Firmicutes before jumping to Actinobacteria hosts. Recent analyses of genetic diversity in Mycobacteriophages have put forward the notion that bacteriophages infecting Mycobacteria do not constitute an isolated environment. Instead, rarefaction analyses suggest that the Mycobacteriophage gene pool is constantly enriched by an influx of genetic material from external sources [5]. The identification here of a group of related phages spanning multiple families within the Actinobacteria and encompassing also two Firmicutes orders provides strong support for the notion that bacteriophages are able to cross over order and phylum distances in relatively short evolutionary spans, potentially contributing to the systematic enrichment of the gene pool available within the population of phages infecting any given genus. Lastly, it should be noted that the Actinobacteriophage clusters identified here are not the only outliers in terms of %GC content divergence from their hosts, suggesting that further sequencing may enable the identification of other close evolutionary relationships between bacteriophages infecting distantly-related hosts.

Author Contributions

conceptualization, IE and SMC; methodology, IE and SMC; software, IE, MK.; validation, IE, SMC, MK and CM.; formal analysis, IE, MK; investigation, IE, SMC, CM, MK; resources, SMC, IE; data curation, IE, SMC, MK; writing—original draft preparation, IE; writing—review and editing, IE, SMC, CM, MK; visualization, IE, MK; supervision, IE and SMC; funding acquisition, SMC.

Acknowledgments

This work was supported by the UMBC Department of Biological Sciences and the Howard Hughes Medical Institute SEA-PHAGES program.

Funding

The authors wish to thank Ralph Murphy for his excellent technical support. The authors also wish to thank Graham F. Hatfull, Deborah Jacobs-Sera, Welkin H. Pope, Daniel R. Russell, Steven G. Cresawn and the Howard Hughes Medical Institute SEA-PHAGES program for their support.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage. 2014;4: e28281–e28281. doi:10.4161/bact.28281
2. Casjens SR. Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol. 2005;8: 451–458. doi:https://doi.org/10.1016/j.mib.2005.06.014
3. Russell DA, Hatfull GF. PhagesDB: the actinobacteriophage database. Bioinforma Oxf Engl. 2017;33: 784–786. doi:10.1093/bioinformatics/btw711
4. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, et al. Origins of highly mosaic mycobacteriophage genomes. Cell. 2003;113: 171–182.
5. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, et al. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Kolter R, editor. eLife. 2015;4: e06416. doi:10.7554/eLife.06416
6. Garneau JE, Tremblay DM, Moineau S. Characterization of 1706, a virulent phage from Lactococcus lactis with similarities to prophages from other Firmicutes. Virology. 2008;373: 298–309. doi:10.1016/j.virol.2007.12.002
7. Kot W, Neve H, Vogensen FK, Heller KJ, Sørensen SJ, Hansen LH. Complete Genome Sequences of Four Novel Lactococcus lactis Phages Distantly Related to the Rare 1706 Phage Species. Genome Announc. 2014;2. doi:10.1128/genomeA.00265-14
8. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2017;45: D37–D42. doi:10.1093/nar/gkw1070
9. Pope WH, Jacobs-Sera D. Annotation of Bacteriophage Genome Sequences Using DNA Master: An Overview. Methods Mol Biol Clifton NJ. 2018;1681: 217–229. doi:10.1007/978-1-4939-7343-9_16
10. Lamine JG, DeJong RJ, Nelesen SM. PhamDB: a web-based application for building Phamerator databases. Bioinforma Oxf Engl. 2016;32: 2026–2028. doi:10.1093/bioinformatics/btw106
11. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics. 2011;12: 395. doi:10.1186/1471-2105-12-395
12. Kloepper TH, Huson DH. Drawing explicit phylogenetic networks and their integration into SplitsTree. BMC Evol Biol. 2008;8: 22. doi:10.1186/1471-2148-8-22
13. Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. Nat Genet. 1999;21: 108–110. doi:10.1038/5052

14. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinforma Oxf Engl. 2017;33: 3396–3404. doi:10.1093/bioinformatics/btx440
15. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14: 60. doi:10.1186/1471-2105-14-60
16. Lefort V, Desper R, Gascuel O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol Biol Evol. 2015;32: 2798–2800. doi:10.1093/molbev/msv150
17. Eddy SR. Accelerated Profile HMM Searches. PLOS Comput Biol. 2011;7: e1002195. doi:10.1371/journal.pcbi.1002195
18. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539. doi:10.1038/msb.2011.75
19. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 1997;14: 685–695. doi:10.1093/oxfordjournals.molbev.a025808
20. Blocker D, Koert M, Mattson C, Patel H, Patel P, Patel R, et al. Complete Genome Sequences of Six BI Cluster Streptomyces Bacteriophages, HotFries, Moozy, Rainydai, RavenPuff, Scap1, and SenditCS. Microbiol Resour Announc. 2018;7. doi:10.1128/MRA.00993-18
21. Pope WH, Jacobs-Sera D, Best AA, Broussard GW, Connerly PL, Dedrick RM, et al. Cluster J Mycobacteriophages: Intron Splicing in Capsid and Tail Genes. PLOS ONE. 2013;8: e69273. doi:10.1371/journal.pone.0069273
22. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33: W244-248. doi:10.1093/nar/gki408
23. Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, et al. The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. J Bacteriol. 2005;187: 1091–1104. doi:10.1128/JB.187.3.1091-1104.2005
24. Bardina C, Colom J, Spricigo DA, Otero J, Sánchez-Osuna M, Cortés P, et al. Genomics of Three New Bacteriophages Useful in the Biocontrol of Salmonella. Front Microbiol. 2016;7. doi:10.3389/fmicb.2016.00545
25. Merrill BD, Ward AT, Grose JH, Hope S. Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genomics. 2016;17: 679. doi:10.1186/s12864-016-3018-2
26. Sharaf A, Oborník M, Hammad A, El-Afifi S, Marei E. Characterization and comparative genomic analysis of virulent and temperate Bacillus megaterium bacteriophages. PeerJ. 2018;6: e5687. doi:10.7717/peerj.5687
27. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host–phage interactions. Proc Natl Acad Sci. 2011;108: E288–E297. doi:10.1073/pnas.1101595108
28. Murooka Y, Takizawa N, Harada T. Introduction of bacteriophage Mu into bacteria of various genera and intergeneric gene transfer by RP4::Mu. J Bacteriol. 1981;145: 358–368.
29. Chiura HX. Generalized gene transfer by virus-like particles from marine bacteria. Aquat Microb Ecol. 1997;13: 75–83.

Supporting Information Captions

Figure S1 - Consensus network inferred on SplitTree with the NeighborNet algorithm using a gene content distance. The branches corresponding to the Actinobacteriophage supercluster and the *Lactococcus* and *Faecalibacterium* phage group are highlighted in red.

Figure S2 - Phylogenetic tree from minimal evolution inference on BLAST-derived inter-genomic distances. The branches corresponding to the Actinobacteriophage supercluster and the *Lactococcus* and *Faecalibacterium* phage group are highlighted in red.

Table S1 - List of phage genomes analyzed in this work.

Table S2 - Groups of orthologous proteins (phams) in the set of analyzed phage genomes.

Table S3 - %GC content of phages and their hosts.

Data S1 - Nexus-formatted SplitTree file for the pham-based tree.