Evidence for shared ancestry between Actinobacteria and Firmicutes bacteriophages

Matthew Koert, Courtney Mattson, Steven Caruso, Ivan Erill

https://doi.org/10.1101/842583 version 1.0
Submitted by Steven Caruso 2019-12-10 15:26

Reviews

Reviewed by anonymous reviewer, 2020-01-21 10:23
Download the review (PDF file)
Reviewed by anonymous reviewer, 2020-02-26 18:29

This brief article discusses the similarity between the genome sequences of several Actinobacteria-infecting phages and those of a small group of Firmicutes phages. The authors hypothesize that this similarity suggests that these Actinobacteria phages have shifted from an ancestor infecting Firmicutes towards their current actinobacterial hosts. This host shift may have occurred recently, as supported by the GC content of these phages which is significantly different from that of their hosts, indicating that they are still in the process of ameliorating it to adapt to their new hosts.

The article is well written, concise and clear. The results are interesting and the conclusions seem to be well supported. I only have two comments:

1- Protein sequence phylogenetic analysis: The authors study the phylogeny of terminase protein sequences using the Neighbor-Joining (NJ) method with the JTT substitution model. It is not mentioned if all positions in the multiple sequence alignment were used for phylogenetic inference or if the alignment was trimmed (e.g., to remove gaps and/or ambiguously aligned positions). This needs to be clarified. In addition, NJ with the JTT model is probably not the best approach to analyze fast-evolving sequences as those frequently found in viruses. The authors should instead use maximum likelihood and/or Bayesian inference with a more appropriate substitution model. Programs like IQ-tree offer the possibility to choose the best-fitting model automatically.

Unfortunately, the diversity of phages used as references for anchoring the phylogenetic placement of the phages under study led to amino acid sequence alignments of highly divergent terminase genes, which contained no conserved blocks when examined with Gblocks, preventing the use of Bayesian inference methods. We therefore used the distance-based NJ method to essentially establish coherent clusters of sequences, which clearly define the supler-cluster composed of Actinobacteriophages and *Lactococcus* and *Faecalibacterium* phages as a cohesive clade of related phages. This has been clearly articulated in the text. The terminase clustering of these phages is supported by whole-proteome phylogenetic analyses (SplitTree &

VICTOR), as well as by previously published phage phylogenies that have been referenced and discussed.

2- The GC content alone remains a very rough estimate of adaptation to the host. It would be useful to have data about more precise indicators, such as the Codon Adaptation Index (CAI). See for example https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4032129/

We agree with the reviewer that CUB is likely a better indicator of host adaptation. CAI is known to be heavily affected by mutational bias (i.e. %GC bias). We therefore have used nRCA, an index that explicitly corrects for mutational bias, to perform an analysis of CUB in these phages. The results, overlaid in the new Figure 5, indicate that CUB in Firmicutes phages is well aligned with that of their hosts, whereas this is not the case for Actinobacteria phages. The new results are presented and discussed in the text.

Reviewed by anonymous reviewer, 2020-03-25 16:24

This work reports on the relatedness of a cluster of phages infecting Firmicutes and Actinobacteria. It shows several genomic alignments and trees, and terminates by a GC% analysis. Data are convincing but need to be presented more rigorously. Below a list of changes that are necessary prior acceptance, in my view.

Major points

1. *There is confusion in the mind of authors between genome relatedness and host range. One should keep in mind that a phage ancestor of the described cluster might have infected the last common ancestor of the Actinobacteria and Firmicutes branches, and then diverged along the various branches. There is no need to invoke wide host range when one finds genome relatedness. Or it should be tested experimentally.*
   We agree with the reviewer that a broad switch in host is not the only, or even most plausible, explanation for the observed data. We have reworded the text accordingly.

   The true and convincing cases of large host range concern identical phages (meaning 99 to 100% identity) infecting different genera/families (e. g. https://www.ncbi.nlm.nih.gov/pubmed/29615108). This same reference should be added when FPoengus is mentioned. It also reports the relatedness of FPoengus with the Lactococcus phage 1706.
   We have added the reference to the general discussion and when introducing FPoengus.

The first sentence of the abstract is particularly obscure. Phages are known NOT to display broad host ranges. May be a negation was lost along the line?

We have removed the first sentence of the abstract.

But the discussion comes back with similar confusions: "it is well known that phages will often infect several hosts within the same genus": no reference here. In fact the opposite is known, as most phages infect a few strains per species. Ref 28 on Mu is given to assert that phage systems have been "engineered" to transcend genus boundaries. In the case of Mu, it is a natural trait, not man made. The ref 29 about "virus transfer across phyla" does not show phage infection, but (putative) phage transduction.

We have qualified the sentences and introduced references to back up the point that many phages do actually display broad host range within their primary host genus.

2. Figures are difficult to read. In Figure 1 (where phage genomes are), scripts are too small. Genes have no arrow (a big weakness of phamerator maps). Color codes are not explained. Shaded area have no scale.

We have reformatted Figure 1 to make legends more readable. A scale and explanation for the color shading is now provided in the caption. Phamerator represents genes in the reverse strand below the genomic position ruler, instead of using arrows. This has also been explained in the figure caption.

Figure 5 is really difficult to read, with all these bars crossing each other. It should be possible (with prism for instance) to display all points as clouds, with genera on x axis, and GC% on y axis. A color code would permit to contrast bacterial genomes and phage clusters. One would see how many members are being compared in each case. The difference between B1 and other Streptomyces clusters is not clear.

We have remade Figure 5 to show clouds of points. We experimented with overlaid clouds for each genus using a color scale, but the resulting plots were extremely hard to read. We have therefore retained the cluster-based x axis organization, separated by genera. The figure legend has been updated accordingly.

3. Figure legends should be greatly improved. It is scarcely understandable which data were used to build each of the trees. Fig5 legend is obscure. The term 'cluster' is repeated everywhere in this legend. Which Faecalibacterium species were tested, how was the representative member chosen ? Actually, this species might be a genus (https://www.ncbi.nlm.nih.gov/pubmed/30547746).

For Figure 5, all the information on phages and hosts was listed in Table S3. We explicitly mention this in the figure legend. We also explicitly denote the origin of the cluster assignments, which has been introduced in the Methods section, in the figure

legend. We also list the corresponding supplementary tables used in the phylogenetic analyses in the figure legends.

4. Annotation of the genes common to this phage cluster should go a bit deeper. In particular, gp58 of RavenPuff is annotated as putative RNA polymerase in oengus. The Koonin group recently studied a group of phages with genes distantly related to RNA polymerases (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5736458/). A further search in this direction would provide more strength to this study. Could this gene be considered a marker gene of the cluster?

The reviewer is correct that the RNAP annotation of FP_oengus should be further investigated. We performed BLAST and HHpred searches, as well as detailed analysis of the multiple sequence alignment of homologs for this protein (all mapping exclusively to this heteregenous phage group). We found that HHpred identifies a PF05183.13 (RdRP) domain, and that two RNAP motifs (including the signature catalytic loop DxDxD motif) identified for crAss-like phages and YonO-like RNAPs are present in the multiple sequence alignment. We have introduced these findings in the text.

5. Low et al. published a remarkable and remarked concatenated tree of phage genomes last year (https://www.ncbi.nlm.nih.gov/pubmed/31110365). Where is the phage cluster in this tree ? Does it group well? Please situate your work relative to Low et al. Could you try building up a tree with your dataset following the footsteps of this paper?

The Low *et al.* tree puts together (99% support for root branch) *Arthrobacter* phage Mudcat with *Rhodococcus* phage ReqiPoco6 and Rhodococcus phage ReqiPepy6, and these with the *Lactococcus* phage group. We have cited the paper and mentioned these results in the Results section.

6. Sullivan's group has proposed a network clustering v-CONTACT approach (https://www.ncbi.nlm.nih.gov/pubmed/31061483) which works well at the genus level apparently. How do your genomes cluster in such graphs?

v-CONTACT shows that the Rhodococcus and Lactococcus phages analyzed here form well-defined genera. The paper also provides data on the faction of protein clusters shared by these phages, and this data clearly support the results of our study. We have cited the paper and mentioned these results in the Results section.

7. The two above-mentioned papers are seminal in the field of phage genomics, they should be referred to, and put into dialog with your work.

We refer to and discuss the two studies mentioned by the reviewer in a new Results paragraph. We find that the data in these two studies supports our phylogenetic inference

results, and we also include comparison to another large-scale phage phylogeny (ViPTree).

Minor points

1. First paragraph of results, just before ref 21, 'consistent with that observed for other Siphoviridae' is too vague, you probably mean consistent with Cluster J of Mycophages (this is ref21 title)?

We have made the comment more explicit as suggested and added a reference to the cluster J Mycophages paper.

2. Second Result paragraph: clusters AM AU and AW were not introduced: what are they referring to, how were they built? For a broad readership, insert "Actinobacteria hosts such as" just after "elicited significant hits against…".

The mechanism and reference for the creation of clusters in PhagesDB has now been explicitly mentioned in the Methods section.

3. Same paragraph: Why is ref 5 mentioned for the phage 1706?

This was an error in reference updating. It has been fixed.

4. End of last paragraph: Figure SX is probably S2.

The reviewer is correct. We have amended the reference.

5. Figure 3 and 4: make broader lines, one hardly sees their colors.

We have made the lines broader to facilitate readability.

6. Discussion: the conservation of two large genomic blocks does not suggest to me that they are 'primary functional elements', they are just less often exchanged. They can constitute the backbone of this cluster if you want. But in phage genomes one hardly thinks in terms of essential and accessory genes.

The sentence has been addressed accordingly.

7. I am not sure that the closer relatedness in GC% is a proof of ancestry. Compared to the wealth of Lactococcus phages sequenced, this 1706 is really an outsider. In Faecalibacterium, much less is known, but FP_oengus was an orphan in the Cornuault et al study (https://www.ncbi.nlm.nih.gov/pubmed/29615108).

We have edited the discussion substantially regarding this aspect (see answer to next comment below).

8. I hope authors understood they should revise their mind on the “rapid jump over order and phylum” (sentence just after ref 5 at the end of the discussion) as mentioned as my top criticism on this paper. Or they should prove this jump experimentally. It would be great.

The discussion has been substantially edited. We have incorporated the alternative explanation suggested by the reviewer (that shared ancestry can explain the observed phylogeny without the need for any trans-order jumps) and we have provided an explanation for the different rates of amelioration observed in Firmicutes- and Actinobacteria-infecting phages, based on the notion that Actinobacteria tend to be slow growers, with reduced translational selection, which could potentially explain why the Actinobacteria-infecting phages in this super-group lag behind in terms of %GC content and CUB optimization.