Cell Systems

Transparent Peer Review Record

Degradation of host translational machinery drives tRNA acquisition in viruses

Joy Y. Yang, Wenwen Fang, Fabiola Miranda-Sanchez, Julia M. Brown, Kathryn M. Kauffman, Chantel M. Acevero, David P. Bartel, Martin F. Polz, Libusha Kelly

Summary

Initial Submission: Received March 26, 2020

Scientific editor: Ernesto Andrianantoandro, Ph.D.

First round of review: Number of reviewers: Two
Two confidential, Zero signed
Revision invited May 12, 2020
Major changes anticipated
Revision received January 06, 2021

Second round of review: Number of reviewers: Two
Two original, Zero new
Two confidential, Zero signed
Accepted May 24, 2021

Data freely available: Yes
Code freely available: Yes

Editorial decision letter with reviewers' comments, first round of review

Dear Dr. Kelly,

I hope this email finds you well. The reviews of your manuscript are back and I've appended them below. On balance, the reviewers appreciate the goals of the work presented here; they've provided constructive comments that are aligned with our hopes for the paper. Accordingly, we're happy to invite a revision.

We appreciate that the COVID-19 pandemic challenges and limits what you and your lab can do, so to make sure we're absolutely on the same page about the feasibility of revisions, let's schedule a Zoom call at our earliest mutual convenience.

To help guide this revision, I've highlighted points in the reviews that seem to warrant special attention. Overall, given the current times, I do not advocate more experimental work. That said, if experimental data ***already exist*** that would help address the reviewer concerns (e.g. Reviewer #2's concern 3), we should discuss how best to incorporate that into the revised manuscript.

I hope you find this feedback helpful. If you have any questions or concerns, I'm always happy to talk, either over email or by phone. More technical information and advice about resubmission can be found below my signature. Please read it carefully, as it can save substantial time and effort later.

I look forward to seeing your revised manuscript.

All the best,

Ernesto Andrianantoandro, Ph.D.
Scientific Editor, Cell Systems

Reviewers' comments:

Reviewer #1: Bacteriophages are viruses that invade bacteria and use their hosts' machinery to reproduce. However, some phages have larger genomes that encode tRNAs, leading to the question of why do some phages carry tRNAs at the cost of shuttling larger genomes. Yang et al. show that phages carrying their own tRNAs can reduce their dependence on the host's resources at later stages of infection and allow them to actively degrade the necessary components at initial infection. Using the T4-like Vibriophage 2.275.O that carries 18 tRNA genes, the authors show that although there is a codon bias for the phage tRNAs towards phage genes it cannot explain the need for encoding tRNA genes. They further demonstrate that the need for phage tRNAs, especially at later stages of phage replication, can be explained by the active degradation of the host's contents by 80% in the first 15 minutes and phage production after 30 minutes. Finally, through analyzing tRNA anticodon diversity and tRNA characteristics, Yang et al. show that the phage seeks to acquire tRNAs that can decode a broad range of codons and supplement the diminishing pool of host tRNAs during phage replication. The work presented support the authors' conclusion that the main reason phage 2.275.O carries tRNA genes is to enable translation at later stages of replication when the host cell's genome is degraded.
1. As the authors conclude that the main reason why phages would carry tRNAs is to allow replication to continue at later stages of infection, it would be interesting to see and discuss if deleting the phage

tRNAs would inhibit translation of the late-stage phage genes and increase the replication time required for phage production.
2. The data presented shows that phage 2.275.O has a diverse tRNA population compared to the host's tRNA population. The host organism used for the analyses involving the phage and host tRNAs is not stated clearly. Phage 2.275.O is stated to infect at least two different species, how similar or different are the tRNAs (e.g. codon usage) from those species and does the diversity of the phage tRNAs compared to the host's tRNAs in the statistical analyses apply to all of the phage's known hosts? Also, does the phage have a preferred host and will thus encode tRNAs that would be more beneficial for phage replication in that specific host?
Minor comments:
1. Define tRNA at the beginning
2. In-text citations (XX et al.) sometimes do not have a comma after the author name
3. Figure 2A: the axis lines and labels appear to be missing
4. Supplementary Figure 5: need a legend for the different colours
5. Page 12 first paragraph: should "Figure 4B" be "Figure 4C"?
6. Need to clearly define what the "host" is for the analyses and RNA sequencing experiments

Reviewer #2: This work investigates the question of why some bacterial phages carry tRNA genes. The first phage carrying tRNA genes was found in E. coli T4; previous studies decades ago suggest that the phage tRNA genes are used to optimize codon usage of phage genes in translation. Using a phage with a very large genome with 18 tRNA genes, this work also asks this question first and finds that the codon optimization hypothesis has some merits. However, the authors show that the main utilization of the phage tRNAs is to supply tRNA to phage gene translation at a time of infection when the host nucleic acids have been severely degraded. This is a new and exciting finding about the origin of phage tRNA genes and provides an important new insight into this long-standing question.

Major comments:
1. p.10, 13, balance of phage and host tRNA: To me this would be the most important data to support the author's new thesis. The phage carries tRNA for only 13 amino acids, so at a minimum, sufficient supply of host tRNA for the other 7 amino acids needs to be present late in infection. Although the tRNA-seq measurement of the phage tRNA (supplemental figures) was done using a state-of-the-art method, the host tRNA measurements seem to be done only using qRT-PCR (Fig. 4C). This is unsatisfying in two major ways. First, the data of the host tRNA level should be in the same sequencing data for phage tRNA; it is unclear why this analysis was not described. Second, because of the many modifications, qRT-PCR of tRNA is notoriously not quantitative; rather, quantitative validation of sequencing results is typically done by Northern blots in the field. The authors should present a comprehensive analysis of both host and phage tRNA from sequencing. Furthermore, some of the codon usage analysis during infection should be done using the combined levels of both host and phage tRNA at that time.

2. p.7, tRNA sequencing: how much host tRNAs are left compared to viral tRNA in the same data? How much selectivity is there for host tRNA degradation? The answer to these questions should be present readily in the tRNA sequencing data.

3. p.14, the point of degrading cellular RNA to salvage for nucleotides: the most ribonucleotides in a cell to be recycled are from ribosomes. There is no measurement of ribosome degradation to fully justify this important point. However, degrading ribosomes may be counter-productive, since the

phage needs ribosome for high level translation of their structural proteins. Presumably there is a balance of degrading versus keeping ribosome at a certain level. Some experimental evidence should be presented to gain insights into this.

Minor comments:
1. p.7, tRNA processing: 5' also needs to be processed by RNase P. Presumably the host RNase P is used here. In the RNA-seq data, how much RNase P RNA is still present after the massive degradation of host transcriptome?

2. p.10: Host data all crunched together in Fig. 4C and impossible to see what's going on. A couple of figure citations are incorrect in this section.

Authors' response to the reviewers' first round comments

Attached.

Editorial decision letter with reviewers' comments, second round of review

Dear Dr. Kelly,

I'm very pleased to let you know that the reviews of your revised manuscript are back, the peer-review process is complete, and only a few minor, editorially-guided changes are needed to move forward towards publication.

In addition to the final comments from the reviewers, I've made some suggestions about your manuscript within the "Editorial Notes" section, below. Please consider my editorial suggestions carefully, ask any questions of me that you need, make all warranted changes, and then upload your final files into Editorial Manager. ***We hope to receive your files within 5 business days, but we recognize that the COVID-19 pandemic may challenge and limit what you can do. Please email me directly if this timing is a problem or you're facing extenuating circumstances.***

As you look forward to acceptance, please do consider submitting one of the protocols you've developed in this paper to STAR Protocols, or extending this offer to one of your trainees. STAR Protocols is geared towards trainees and its key purpose is to provide complete and consistent instructions for how to conduct reproducible experiments. If you have any questions, please email starprotocols@cell.com.

I'm looking forward to going through these last steps with you. Although we ask that our editorially-guided changes be your primary focus for the moment, you may wish to consult our FAQ (final formatting checks tab) to make the final steps to publication go more smoothly. More technical information can be found below my signature, and please let me know if you have any questions.

All the best,

Ernesto Andrianantoandro, Ph.D.
Scientific Editor, Cell Systems

Editorial Notes

Title:

Your title is too generic and does not capture the main advance of the paper. I suspect it could be more effective. Please revise to mention some specifics, such as degradation/recycling of host machinery. As you re-consider your title, note that an effective title is easily found on Pubmed and Google. A trick for thinking about titles is this: ask yourself, "How would I structure a Pubmed search to find this paper?" Put that search together and see whether it comes up is good "sister literature" for this work. If it does, feature the search terms in your title. You also may wish to consider that PubMed is sensitive to small differences in search terms. For example, “NF-kappaB” returned ~84k hits as of March, 2018, whereas “NFkappaB” only returned ~8200. Please ensure that your title contains the most effective version of the search terms you feature.

Abstract:

Your abstract reads wonderfully, but it is unfortunately too long. Please condense to 150 words or less.

Manuscript Text:

Please remove the Significance Statement and move any non-redundant text to the Introduction or Discussion, where appropriate.

Also:

- House style disallows editorializing within the text (e.g. strikingly, surprisingly, importantly, etc.), especially the Results section. These terms are a distraction and they aren't needed—your excellent observations are certainly impactful enough to stand on their own. Please remove these words and others like them. “Notably” is suitably neutral to use once or twice if absolutely necessary.

STAR Methods:

Please revise the methods section to according to our STAR Methods format - for details follow the link **here**. In addition to the standard STAR Methods format, Cell Systems papers will need to contain a comprehensive and structured "Data and Code Availability" statement. so please modify the first section of your STAR Methods as follows, noting that the particular examples used might not pertain to your study.

> RESOURCE AVAILABILITY
>
> **Lead Contact:** Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jane Doe (janedoe@qwerty.com).
>
> **Materials Availability:** This study did not generate new materials. -*OR*- Plasmids generated in this study have been deposited at [Addgene, name and catalog number]. -*OR*- etc.
>
> **Data and Code Availability:**
>
> - **Source data statement** (described below)
> - **Code statement** (described below)
> - **Scripts statement** (described below)
> - Any additional information required to reproduce this work is available from the Lead Contact.

Data and Code Availability statements pertain to the source data and original code reported in the study. In this context, **source data** is defined as the collection of individual, unprocessed observations used to generate the figures reported in the paper. Examples include scRNA-seq and proteomic datasets, but also CSV spreadsheets used to generate graphs, and original micrographs in TIFF format. **Code** is defined as any computationally implemented program, algorithm, or pipeline necessary to reproduce the analysis or conclusions reported in a paper. Smaller **scripts** that have been used to visualize data and generate figures should also be included in the statement, as described below.

Data and Code Availability statements are reported in the first section of the STAR Methods**. They have four parts and each part must be present. Each part should be listed as a bullet point, as indicated above. For convenience, a .docx template for Data and Code availability statements can be downloaded here.**

> **Part 1 pertains to source data.** Examples can be used in any number or combination, making sensible modifications as necessary:
>
> - [Data-type] source data have been deposited at [data-type-specific repository] and are publicly available under the accession numbers: [Insert].
> - [Data-type] source data have been deposited at [general repository] and are publicly available at [insert DOI].

- [Data-type] source data are available in the paper's Supplemental Information.
- The [data-type] source data reported in this study have not been deposited in a publicly available repository because [reason why data are not public] . They have been archived locally [insert archiving plan]. To request access [insert instructions].
- This paper analyzes existing, publicly available data. These datasets' accession numbers are provided in the Key Resource Table.
- Source data are not provided in this paper but are available from the Lead Contact on request. *(Note: Cell Systems discourages this practice. If you need to make this statement, please discuss it with your editor first.)*

Part 2 pertains to original code. Examples can be used in any number or combination, making sensible modifications as necessary:

- [Adjective] original code is publicly available at [repository name and DOI].
- [Adjective] original code is available in this paper's Supplemental Information.
- The original code reported in this study is not publicly available repository because [reason why data are not public]. Original code has been archived locally [insert archiving plan]. To request access [insert instructions].
- This paper does not report original code.

Part 3 pertains to scripts used to generate figures. Examples to be used in any number or combination:

- The scripts used to generate the figures reported in this paper are available at [repository name and DOI].
- The scripts used to generate the figures reported in this paper are available in this paper's Supplemental Information.
- The scripts used to generate the figures reported in this paper are available in the [name software package, with version, and provide reference or URL] and their use is described in the STAR Methods.
- Scripts were not used to generate the figures reported in this paper.
- Scripts used to generate the figures presented in this paper are not provided in this paper but are available from the Lead Contact on request. *(Note: Cell Systems discourages this practice. If you need to make this statement, please discuss it with your editor first.)*

Part 4 is a statement: "Any additional information required to reproduce this work is available from the Lead Contact."

Thank you!

Reviewer comments:

Reviewer #2: The authors adequately addressed most of my comments. The remaining ones may be considered in future studies.

We thank the reviewers for their thoughtful comments on our work. In revising our manuscript to address the reviewer comments and concerns we have **1)** clarified that we obtained tRNA-Seq data for both the host and the phage and added an additional supplemental figure to make this point more clear; **2)** described the codon usage profiles for other hosts of our target virus and added an additional supplemental figure for this analysis; **3)** added two supplemental figures and a supplemental table describing host transcription during infection; and **4)** discussed additional caveats and potential future experiments based on the foundation of work described this paper. We thank the reviewers also for catching some errors that we have also fixed. All changes to the original submission are shown in red in the attached revised manuscript and revised supplementary data. We hope that our responses here sufficiently clarify our work and render it suitable for publication.

Reviewers' comments:

Reviewer #1: Bacteriophages are viruses that invade bacteria and use their hosts' machinery to reproduce. However, some phages have larger genomes that encode tRNAs, leading to the question of why do some phages carry tRNAs at the cost of shuttling larger genomes. Yang et al. show that phages carrying their own tRNAs can reduce their dependence on the host's resources at later stages of infection and allow them to actively degrade the necessary components at initial infection. Using the T4-like Vibriophage 2.275.O that carries 18 tRNA genes, the authors show that although there is a codon bias for the phage tRNAs towards phage genes it cannot explain the need for encoding tRNA genes. They further demonstrate that the need for phage tRNAs, especially at later stages of phage replication, can be explained by the active degradation of the host's contents by 80% in the first 15 minutes and phage production after 30 minutes. Finally, through analyzing tRNA anticodon diversity and tRNA characteristics, Yang et al. show that the phage seeks to acquire tRNAs that can decode a broad range of codons and supplement the diminishing pool of host tRNAs during phage replication. The work presented support the authors' conclusion that the main reason phage 2.275.O carries tRNA genes is to enable translation at later stages of replication when the host cell's genome is degraded.
1. As the authors conclude that the main reason why phages would carry tRNAs is to allow replication to continue at later stages of infection, it would be interesting to see and discuss if deleting the phage tRNAs would inhibit translation of the late-stage phage genes and increase the replication time required for phage production.

Phage tRNA deletions are difficult, especially as this *V. cyclitrophicus* host is a non-model organism. However, we agree that this is a useful future experiment and have added this idea to the discussion.

2. The data presented shows that phage 2.275.O has a diverse tRNA population compared to the host's tRNA population. The host organism used for the analyses involving the phage and host tRNAs is not stated clearly.

Apologies for the oversight! The host species and strain number have been added to the Abstract and Introduction.

Phage 2.275.O is stated to infect at least two different species, how similar or different are the tRNAs (e.g. codon usage) from those species

We have added a new Supplementary figure 5 to address this question. We highlight some interesting observations here and we've included a discussion of these points in the revised manuscript:

- As expected, codon usages within species are very similar; phage codon usage is vastly different from codon usage by any *Vibrio* including that of its hosts'.
- *V. cyclitrophicus* hosts are more similar in codon usage to 2.275.O than *V. lentus* hosts. Because *V. cyclitrophicus* also happens to be the host of isolation, it is tempting to infer that *V. cyclitrophicus* must be the preferred host for 2.275.O; however, doing so would require a more precise definition for host preference. This would certainly be a fascinating direction for future exploration.

and does the diversity of the phage tRNAs compared to the host's tRNAs in the statistical analyses apply to all of the phage's known hosts?

This is definitely true, in fact, it is highly unlikely for a phage to acquire 18 different tRNAs even at random from a uniform distribution, as noted in Figure 5.

Also, does the phage have a preferred host and will thus encode tRNAs that would be more beneficial for phage replication in that specific host?

While the phage's host of isolation is a strain of *Vibrio cyclitrophicus*, this does not necessarily mean that *Vibrio cyclitrophicus* is its preferred host. The phage is able to form larger plaques on its *Vibrio lentus* host (using an agar overlay), however, it is more difficult to culture the phage to high titers on *Vibrio lentus* than on *Vibrio cyclitrophicus*. We note that the experiments were performed in laboratory conditions that were not aimed at recreating the natural environment of infection. We therefore do feel that the question of host preference likely has a more complex answer (can be affected by host growth rate, capacity for biofilm formation, temperature, etc.) and so we would hesitate to attribute host preference to just the tRNAs. In response to this comment we discuss host preference briefly in the revised manuscript.

Minor comments:
1. Define tRNA at the beginning

Done.

2. In-text citations (XX et al.) sometimes do not have a comma after the author name

Done.

3. Figure 2A: the axis lines and labels appear to be missing

Figure 2A is a multidimensional scaling plot, we have added this clarification to the figure description.

4. Supplementary Figure 5: need a legend for the different colours

Legend added.

5. Page 12 first paragraph: should "Figure 4B" be "Figure 4C"?

Done, thanks for catching this.

6. Need to clearly define what the "host" is for the analyses and RNA sequencing experiments

Done.

Reviewer #2: This work investigates the question of why some bacterial phages carry tRNA genes. The first phage carrying tRNA genes was found in E. coli T4; previous studies decades ago suggest that the phage tRNA genes are used to optimize codon usage of phage genes in translation. Using a phage with a very large genome with 18 tRNA genes, this work also asks this question first and finds that the codon optimization hypothesis has some merits. However, the authors show that the main utilization of the phage tRNAs is to supply tRNA to phage gene translation at a time of infection when the host nucleic acids have been severely degraded. This is a new and exciting finding about the origin of phage tRNA genes and provides an important new insight into this long-standing question.

Major comments:
1. p.10, 13, balance of phage and host tRNA: To me this would be the most important data to support the author's new thesis. The phage carries tRNA for only 13 amino acids, so at a minimum, sufficient supply of host tRNA for the other 7 amino acids needs to be present late in infection. Although the tRNA-seq measurement of the phage tRNA (supplemental figures) was done using a state-of-the-art method, the host tRNA measurements seem to be done only using qRT-PCR (Fig. 4C). This is unsatisfying in two major ways. First, the data of the host tRNA level should be in the same sequencing data for phage tRNA; it is unclear why this analysis was not described. Second, because of the many modifications, qRT-PCR of tRNA is notoriously not quantitative; rather, quantitative validation of sequencing results is typically done by Northern blots in the field. The authors should present a comprehensive analysis of both host and phage tRNA from sequencing. Furthermore, some of the codon usage analysis during infection should be done using the combined levels of both host and phage tRNA at that time.

We believe there may have been a misunderstanding in this point. While Figure 4B depicts results from qPCR that quantifies host chromosome copies, Figure 4C shows both the phage and host tRNA from RNA-seq. We have made adjustments to the figure caption to clarify this point. To further elaborate on this question here, in figure 2D, the y-axis is defined by the host tRNA expression and the phage tRNA expression, and so the trend in late genes being better encoded for by the expressed population of phage tRNA, which become available mid-infection.

In figure 5B, we see that the tRNAs that the phage do not carry are either (1) typically expressed more highly by the host than tRNA that the phage do carry, implying that perhaps the phage does not need to carry these because they can be supplied by the host, or (2) barely utilized by the phage genes, implying that perhaps the phage does not need to carry these because it does not need them to encode their proteins. In addition, we apologize that we need to revise the “13 amino acids” to “15 amino acids.” This is because we observed that additional tRNA called using another tool, Aragorn, do appear to be expressed.

2. p.7, tRNA sequencing: how much host tRNAs are left compared to viral tRNA in the same data? How much selectivity is there for host tRNA degradation? The answer to these questions should be present readily in the tRNA sequencing data.

We do agree that the comparison of phage vs. host tRNA levels would be better made with the tRNA sequencing results rather than with the RNA-seq results (Figure 4C); however, unfortunately, our tRNA-seq infection timecourse was performed at a lower titer than our RNA-seq infection timecourse. Because of this, the fraction of uninfected host cells in the infection culture was quite substantial and so it was difficult to infer the levels of host tRNA in infected cells. We have added text describing this caveat to the revised manuscript.

However, we do observe phage RNA degradation, as seen in Figure 4 and Supplementary Figures 17 and 18, so it is likely that the phage and host tRNA are likely degraded at similar rates, but because the host genome is degraded, no additional host tRNA can be produced. Additionally, because our time course was taken at 15 minute intervals, we would hesitate to infer a rate of tRNA degradation, as our observed trough of host RNA expression is at 15 minutes. Beyond 15 minutes, the host RNA levels begin to climb again, due to the growth of uninfected host cells. We have added a brief discussion of the background of uninfected hosts to the Figure 4 caption.

3. p.14, the point of degrading cellular RNA to salvage for nucleotides: the most ribonucleotides in a cell to be recycled are from ribosomes. There is no measurement of ribosome degradation to fully justify this important point. However, degrading ribosomes may be counter-productive, since the phage needs ribosome for high level translation of their structural proteins. Presumably there is a balance of degrading versus keeping ribosome at a certain level. Some experimental evidence should be presented to gain insights into this.

Unfortunately, because the ribosomes need to be depleted for RNA-seq, we do not have reliable data to expand upon this point.

Minor comments:
1. p.7, tRNA processing: 5' also needs to be processed by RNase P. Presumably the host RNase P is used here. In the RNA-seq data, how much RNase P RNA is still present after the massive degradation of host transcriptome?

We searched for the protein component of RNase P in our host RNA-seq data and we find that this transcript falls into the main cluster representing the majority of host transcripts (black circles in supplementary figure 19, ‘typical’ expression pattern in supplementary figure 20). In this cluster, transcripts are rapidly degraded, reaching a trough at approximately 15 minutes post-infection. It is possible that despite the mRNA for the protein component of RNase P being degraded, the enzyme itself is still present; we cannot assess this possibility as we did not do proteomics. It is also possible that while transcript levels decline, they are not degraded entirely. We regret that we can’t give a more substantial answer to how the tRNA processing is carried out in the background of host transcriptome degradation.

2. p.10: Host data all crunched together in Fig. 4C and impossible to see what's going on. A couple of figure citations are incorrect in this section.

We have corrected the figure citations. To clarify Figure 4C, we’ve added supplementary figure 7, highlighting each tract by amino acid, for further exploration of the data.