nature portfolio

Peer Review File

DNA viruses are constrained to ecological niches and share similar environmental adaptations with hosts

Corresponding Author: Dr Kim Handley

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Hoggard et al., study aims to understand the impact of environmental conditions, mainly salinity and sediment/water column, on virus communities. I have read the revised paper and the response to reviewers. The authors attempted to address the reviewers comments, and the manuscript is improved, but the outcomes of those analyses were not adequately weighted in interpreting the results, leaving me with a similar conclusion to the first round of reviews. Due to overwhelmingly strong relationship with host community, and the more subtle variation beyond that, which is not clearly independent of this, interpretation is challenging in the environmental chemistry, physical conditions (salinity) direction. That is, the very strong relationship between viruses and hosts is not convincingly distentangled from the idea of environmental variation playing in sort of addition role, which is still the angle for which the authors are going for. Indeed in some analyses the environmental correlation 'goes away' after accounting for host community.

As noted, the new analyses reveal that the strongest driver of viral community variation is host community. There are no figures in the main text related to this, but Table S5 shows it clearly and that these correlations are extremely high, and higher than those to the environmental parameters. In general, therefore, the impact if environmental conditions can't really be constrained beyond the host without a more explicit tests. Where there were tests, it shows that there is no effect beyond host especially for the water column: Correlations between combined environmental variables or salinity and vOTU composition were no longer apparent in water samples after taking the effect of MAG composition into account (partial 3-way Mantel test, p-values ≥0.65), reinforcing the importance of host community composition on viral distributions. Although, the effect remained for sediment samples (r=0.30) for combined and 0.65 for salinity, p-values <0.01).

Second, the issue of the small effect sizes of the protein/amino acid trains are now acknowledged, but it doesn't change the fact that they are seemingly miniscule. I appreciate the figures of protein structure, but more details can be provided about how these representatives were chosen. Also, the authors seemed to have tried to consider that GC was a driving factor, but the text on this is contradictory. "Based on Waiwera estuary data, viral gene GC proportions decreased linearly and significantly with increasing salinity in water and sediment." vs "This again suggests that salinity was not a driver of viral GC content trends in the estuary." The answer seems to relate to the location of the proteins, but the effect size are so small, that it is not conclusive to me, especially considering the correlations with hosts.

The title: Adaptations of DNA viruses are influenced by host and environment with proliferations constrained by environmental niche. I am still detecting a lot of effort to connect the changes to environmental niche (suggesting physical and chemical interactions beyond host interactions for which there is negligible data in my view). To me, a better title would need to accurately describe the main results (viral community correlation to host community), not very subtle or inconclusive aspects.

Why are there no Megaviricities at all in this manuscript? At what stage are they removed from the analysis or are there none in this data?

Table S4. Is PhiX, the sequencing standard, appearing here?If so, shouldn't it be removed from mapping/ecological results, and also here?

Table S5. Combined environmental parameters: Shouldn't the analyses iteratively determine if environmental parameters are making an impact, and so one wouldn't want to include many correlated factors, like potentially strongly correlated nutrient values? Also, I think it would be better to try to model which parameters have an additive improvement to the correlations. Seemingly for example, MAGS + salinity alone could be tested instead of all combined environmental parameters together. If the correlation gets worse, it would suggest that it is host and not specifically salinity that plays a role. Notably, MAGS + combined environmental parameters are much worse than just correlations to MAGS.

In my opinion, the IMG/VR analyses don't add that much, and are a bit distracting. I focused most of my review on the more directly applicable results to the main thrust of the manuscript. The most interesting aspect of that data may be the halophiles, but here we also do not see a large (or any effect, via Figure 5b).

Since the MAG and vOTU analyses seem to be completely independent on the bioinformatic side, it seems likely that some vOTUs are also within MAGs which could drive some of the remarkably strong correlations. It would be valuable to characterize how much of the correlation between host and vOTUs could be driven by this.

(Remarks on code availability)
Well organized.

Reviewer #3

(Remarks to the Author)
In their manuscript, "Adaptations of DNA viruses are influenced by host and environment with proliferations constrained by environmental niche," Hoggard et al. investigated viral and microbial diversity and distributions across a salinity gradient in the Waiwera estuary. Viral and microbial diversity and distributions were characterized using shotgun metagenomics and metatranscriptomics of water and sediment, and the genomic features associated with non-saline, brackish, and marine viral community members, including amino acid composition, predicted protein isoelectric points, and GC content were characterized.

Based on their observations, the authors concluded that 1) sampled viral communities were dominated by members of the Caudoviricetes; 2) environmental barriers (e.g. salinity) constrain viral distributions; and 3) genomic features of environmentally constrained viral communities show signatures of environmental adaptation (lower isoelectric points, etc.), including evidence of osmoadaptation independent of similar adaptation by microbial hosts.

General Comments

The authors have done a nice job addressing the comments and concerns of the reviewers. Although the associations between amino acid composition and salinity remain weak, more accurate classification of viral proteins has resulted in larger R2 values in some cases (Figs. 4A, S4), and structural predictions indicate that these changes are at the protein surface as would be expected.

The authors' assertion that increases in acidic amino acid proportions (specifically Asp) in viral proteins are viral adaptations that are host-independent remains less convincing. This conclusion is based on an observed negative correlation between viral aspartic acid proportions (which increase with salinity) and MAG tRNA gene and transcript proportions. The authors show that MAG Asp tRNA gene and transcript proportions decrease with salinity when these data are pooled by site (Fig. 5C). Yet, Asp tRNA transcript proportions show no significant change with salinity (and even display a slightly positive slope) on a per MAG basis (Fig. S9B). It is unclear why the authors used the pooled data set in this analysis rather than the proportions per MAG, but the pooled data set appears to contain potential confounding variables. The authors are encouraged to re-analyze their correlations using the per MAG proportions before drawing conclusions.

Additionally, positive correlations (for example, between Glu amino acid proportions in viral proteins and MAG tRNA transcripts, Fig. 5D) may still represent instances of viral adaptation beyond a reflection of host adaptation. Have the authors looked at the rate of change in viral amino acid proportions with increasing salinity relative to the rate of change in MAG tRNA transcripts? This might indicate, even for positive correlations, changes in viral amino acid proportions above a background level expected from host adaptations to salinity. While not conclusive, this could bolster arguments that changes to amino proportions go beyond host adaptation.

Specific Comments

Figure 2 legend: The vOTUs per estuary zone add up to 857, but this number should be 858?

Lines 286-289: Are these analyses shown anywhere? If not, please indicate data not shown.

Line 435: Changes in amino acid usage in viruses are referenced, but the Figure referenced is Fig. S9a. Did the authors mean Fig. S10?

Lines 412-417 and Fig. S8b,c: The authors suggest that tRNA repertoires do not cluster by phyla or estuarine zone. This

conclusion is based on tRNA gene counts, but gene number does not always correlate with transcript number or codon bias. Please confirm this trend using transcript abundances.

(Remarks on code availability)
The code is accessible, well documented, and usable.

Reviewer #4

(Remarks to the Author)
I found the authors adequately addressed all of reviewer's 1 comments in this revised manuscript. Overall, I found the reviewed manuscript clear, with results presented with the right literature context and caveats. I only have one minor suggestion: if I understand correctly, the authors study virus sequences obtained from cellular fraction metagenomes (l. 534-535: "Filtered water samples included viruses associated with cells or biofilms trapped on 0.22 micron filters."). This is not an issue in itself, and is actually also the case for most of the IMG/VR database. However it tends to yield datasets enriched in prophages (see doi 10.1186/s40168-024-01905-x), most of them being integrated into the host genome, and thus possibly having a stronger adaptation to host in terms of GC content and codon usage as a byproduct of this temperate lifestyle and host-integrated state. This may be worth mentioning, e.g. l. 390-401 or in the final summary l. 507-527.

(Remarks on code availability)
The code repository includes the documentation and scripts necessary to reproduce the analysis.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have significantly revised the manuscript and I think the framing has improved. The Mantel tests show that the viral community and bacterial community is almost perfectly correlated in both water column and sediment, with no added relationship between viruses and salinity or combined environmental parameters in the water column. Nevertheless, there are significant correlations between amino acid/protein traits with the environmental parameters.

I appreciate that the authors have softened a bit their title and 'environmental driver' part, and acknowledge the strong host correlation factor. But, I still think 'constrains' is too strong: DNA viruses have adaptations that correlate with ecological parameters.

I believe the version did not address this point from reviewer 2: "Second, the issue of the small effect sizes of the protein/amino acid trains are now acknowledged, but it doesn't change the fact that they are seemingly miniscule." As a further comment, perhaps some literature references that provide context to the magnitude of the shift would be valuable.

Figure 4a, it would be helpful to plot the individual points for the plots where possible (I think this is Nature Communications best practices), which may aid with interpretation, or may overwhelm the signal.

(Remarks on code availability)
Code is available.

Reviewer #3

(Remarks to the Author)
In their revised manuscript "DNA viruses are constrained by ecological niche and exhibit environmental adaptations in common with prokaryote hosts," the authors have done a thorough job of addressing the concerns of the reviewers. The manuscript is clear and the scope of the claims have been adjusted to coincide with the presented results. I have no further concerns or suggestions.

(Remarks on code availability)
The code is accessible, well documented, and usable.

Reviewer #4

(Remarks to the Author)
I thank the authors for addressing the remaining reviewer comments, and have no further suggestion at this stage.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

Hoggard et al., study aims to understand the impact of environmental conditions, mainly salinity and sediment/water column, on virus communities. I have read the revised paper and the response to reviewers. The authors attempted to address the reviewers comments, and the manuscript is improved, but the outcomes of those analyses were not adequately weighted in interpreting the results, leaving me with a similar conclusion to the first round of reviews. Due to overwhelmingly strong relationship with host community, and the more subtle variation beyond that, which is not clearly independent of this, interpretation is challenging in the environmental chemistry, physical conditions (salinity) direction. That is, the very strong relationship between viruses and hosts is not convincingly distentangled from the idea of environmental variation playing in sort of addition role, which is still the angle for which the authors are going for. Indeed in some analyses the environmental correlation 'goes away' after accounting for host community.

>> RESPONSE: We agree with the reviewer that the host-effect is overwhelmingly the biggest effect in our dataset, and also that it is not possible from our data to conclusively determine whether osmoadaptive-like traits are the result of direct environmental effects or indirect effects obtained via coevolution with the host, although biases associated with protein location suggest the former. To address this issue, we have further modified the manuscript: (1) to emphasise that results indicate host is the primary driver of viral distributions; (2) with a reduced emphasis on causation when evaluating relationships between genomic signatures and habitat; and (3) to acknowledge that further studies are needed to determine whether the amino acid biases observed are simply present due to co-evolution with the host, or due to direct adaptation to environmental factors. Examples are given below.

The abstract has been modified as follows:

> "… Results suggest that viruses feature some of the same osmoadaptive traits as their hosts. These included slightly elevated ratios of acidic to basic amino acids and decreased protein isoelectric points at higher salinities, particularly in virus major tail and capsid proteins, which did not solely reflect biases inherited due to reliance on host machinery (tRNA pools). Further studies are needed to determine the primary driver of these modifications in viruses (e.g. environment or host) and whether these traits further restrict virus distributions beyond the effects of host-range limitation. Overall, our findings indicate that successful proliferations of viruses into distinct biomes (e.g. freshwater, saline, terrestrial) are rare, with viruses constrained to specific ecological niches." (lines 23-30)

The following was added to the introduction:

> "However, the extent to which proliferation across habitat boundaries is additionally influenced by adaptation of viruses to the physicochemistry of

> distinct niches is less well understood – whether in direct response to environmental pressures or indirectly via coevolution with the host." (lines 56-59)

The results section of amino acid bias now ends with the following statement:

> "Overall, our observations indicate that adaptations to salinity, common to prokaryotes, are also present in viruses. Results comparing between the different gene categories suggest that these adaptations could be due to the influence of the external environment. Whether these traits represent a biologically meaningful additional restriction on virus proliferation beyond host-range limitation requires further study, and experimental evidence is needed to determine the extent that such adaptations are due to the direct effects of environment versus co-evolution with the host." (lines 431-437)

The summary has been modified to emphasis the significance of the host as follows:

> "… Strong correlations were observed between estuarine virus and prokaryote community distributions suggesting host-range limitation, and that virus distributions are primarily constrained by the distributions of their hosts. … The modification of protein amino acid repertoire in response to environmental salinity appears to be a common adaptive strategy spanning both prokaryotes and viruses. Further studies are needed to determine the primary driver of these modifications in viruses (e.g. direct environmental adaptation, or indirectly acquired via adaptation to the host), the effect on virus fitness, and the influence of virus-host interaction strategy as employed by lytic versus temperate phages. … Our findings indicate that successful proliferation of viruses transferred to distinct biomes (e.g. freshwater, saline, terrestrial) is limited, and that the distribution of viruses and hosts are similarly constrained to specific ecological niches." (lines 559-574)

As noted, the new analyses reveal that the strongest driver of viral community variation is host community. There are no figures in the main text related to this, but Table S5 shows it clearly and that these correlations are extremely high, and higher than those to the environmental parameters. In general, therefore, the impact if environmental conditions can't really be constrained beyond the host without a more explicit tests. Where there were tests, it shows that there is no effect beyond host especially for the water column: Correlations between combined environmental variables or salinity and vOTU composition were no longer apparent in water samples after taking the effect of MAG composition into account (partial 3-way Mantel test, p-values ≥0.65), reinforcing the importance of host community composition on viral distributions. Although, the effect remained for sediment samples (r=0.30) for combined and 0.65 for salinity, p-values <0.01).

>> RESPONSE: Please see our response above relating to the strong correlations with host. **Figure 2f** shows the strong relationship between viruses and hosts. We have also added Mantel test statistic plots (based on data in **Table S5**) to **Figure 2** (please see **Figure 2e**).

Second, the issue of the small effect sizes of the protein/amino acid trains are now acknowledged, but it doesn't change the fact that they are seemingly miniscule. I appreciate the figures of protein structure, but more details can be provided about how these representatives were chosen. Also, the authors seemed to have tried to consider that GC was a driving factor, but the text on this is contradictory. "Based on Waiwera estuary data, viral gene GC proportions decreased linearly and significantly with increasing salinity in water and sediment." vs "This again suggests that salinity was not a driver of viral GC content trends in the estuary." The answer seems to relate to the location of the proteins, but the effect size are so small, that it is not conclusive to me, especially considering the correlations with hosts.

>> RESPONSE: We have added additional details to the methods on how representative proteins were chosen for the protein structure figures, as follows:

> "Viruses were selected based on the following criteria: Target proteins were first identified via word searches of DRAM-v annotations, including filtering to remove gene hits with conflicting results across different annotation databases, and followed by further manual curation based on structural prediction results. Viruses containing putative annotations for all four target proteins were selected from freshwater (Waiwera vOTUs), marine (Waiwera vOTUs), and hypersaline (IMG/VR) environments for further analysis. After this initial filtering process, six replicate viruses from each of the three habitat and four protein types remained (72 proteins from 18 virus genomes). Three-dimensional structures of protein complexes (major tail and major capsid) and individual proteins (large terminase and DNA polymerase) were predicted from protein sequences via Alphafold 3 server[128] with the default settings. Proteins were visualised in UCSF ChimeraX[129], with electrostatic potential calculations via the *coulombic* function with a standardised range (coulombic range -30,30). Two putative large terminase proteins were excluded during analyses due to poor structural predictions, resulting in 70 proteins in total." (lines 745-757).

We have also revised the section on GC to clarify that we found no consistent relationship between viral GC content and salinity, as is also lacking in prokaryotes:

> "Elevated GC content is also often (although not always) associated with prokaryotes from hypersaline environments, and may be a factor of the environments that many halophiles commonly inhabit, independent of salinity[20,59]. In halophiles, key traits of salinity adaptation, including amino acid biases, are present irrespective of overall GC content, which may co-vary in response to other environmental factors such as nitrogen limitation[60] or UV-induced damage[20,59]. In this study, the lack of consistent association between viral GC content and salinity (i.e., GC was lowest in marine relative to both non-saline and hypersaline environments), suggests that salinity is also not a driver of viral GC trends." (lines 310-317)

We have also clarified that our overall interpretation of viral GC contents is that host is likely to be the key driver of differences:

> "Further analysis of IMG/VR viruses across a variety of source ecosystems (e.g., aquatic non-saline, aquatic saline, aquatic hypersaline, terrestrial, host, or air) showed proportions of viral gene GC content differed strongly by environment, indicating that factors influencing viral genomic GC content are broad (**Figure 4b; Figure S5a**). GC content distributions for IMG/VR viruses from wetlands, groundwater, sediment, deep subsurface, plant litter, and the air were multimodal and tended to span the ranges of GC proportions characteristic of aquatic and terrestrial viruses (**Figure S5a**), and may reflect distinct microenvironments within these ecosystems. A recent study examining viral genomic content in relation to broad host categories for ~10,000 viral genomes also found multi-modal viral GC contents, largely driven by a trimodal pattern for dsDNA viruses dominating the dataset[31]. It is unclear whether any differences in GC content among or within ecosystems are due to direct environmental adaptation and/or selective pressures acting directly on host GC content that are acquired by viral genomes via host-specific adaptation. However, host characteristics may underlie different distributions due to host-related selective pressures on viral genome content[31], which can include host translational biases[61] and viral-host gene exchange[62]. Intrinsic characteristics of viral lineages may also contribute to differences. For example, the link between virus and host GC content and codon usage bias is also stronger for bacteriophage than for eukaryotic viruses[35]. Here, when subsetting IMG/VR viruses by realm, large differences in GC content were evident among realms (i.e., typically lower in *Varidnaviria* versus *Riboviria*, *Duplodnaviria*, and *Monodnaviria*) (**Figure S5b**). Differences among viral GC contents from aquatic non-saline, aquatic marine, hypersaline, and terrestrial environments were also more strongly associated with *Duplodnaviria* and *Monodnaviria*, further suggesting that underlying differences in viral biology or host domain between the realms influence the association between virus GC content and environmental source." (lines 318-339)

The title: Adaptations of DNA viruses are influenced by host and environment with proliferations constrained by environmental niche. I am still detecting a lot of effort to connect the changes to environmental niche (suggesting physical and chemical interactions beyond host interactions for which there is negligible data in my view). To me, a better title would need to accurately describe the main results (viral community correlation to host community), not very subtle or inconclusive aspects.

>> RESPONSE: We believe the analyses on the extent to which viruses are potentially constrained to environmental niche due to direct adaptation is an important addition to this study. However, we acknowledge the reviewer's concern that the current title and elements of the manuscript text still under emphasise host influence. To address this

concern, we have updated the title and the manuscript text to reframe the analyses on direct environmental adaptation of viruses as an additional element of the study, in contrast to a central finding. Please see comment above for more detail and examples regarding updates to the text. We have updated the title as follows:

> "DNA viruses are constrained by ecological niche and exhibit environmental adaptations in common with prokaryote hosts"

Why are there no Megaviricities at all in this manuscript? At what stage are they removed from the analysis or are there none in this data?

>> RESPONSE: No viruses from the class *Megaviricetes* were identified based on the methods used to predict the taxonomy of vOTUs in this study. However, 23 out of 858 vOTUs were identified (n=16) or estimated (n=7) to have genomes >200 kbp, and are defined as jumbo phage in **Table S2**. This includes one virus with a complete genome of 532 Kbp long that may instead be considered a mega virus. Of these, 21/23 (including the mega virus) were predicted to be from the class *Caudoviricetes*, while taxonomy was not predicted for the remaining two viruses. This information has been added as follows: "… including 22 predicted jumbo viruses (>200 Kbp) and one mega virus (>500 Kbp)[36,37] (**Table S2**)" and "*Caudoviricetes* (tailed dsDNA phage; phylum Uroviricota) were the dominant class overall (*n*=616, including *n*=21/23 of the jumbo viruses and mega virus)" (lines 81-82 and 101-103)

Table S4. Is PhiX, the sequencing standard, appearing here?If so, shouldn't it be removed from mapping/ecological results, and also here?

>> RESPONSE: PhiX from Illumina was used as a standard at low concentration (1% v/v) due to the high natural diversity of our samples. About 0.9% of reads aligned to PhiX in the initial sequence data. These were in the unindexed/unbarcoded fraction of reads and were retained by the sequencing facility following demultiplexing (i.e. they were not part of the data we received or analysed). However, as PhiX contamination of isolate genomes has been reported in the literature (although possibly where samples were not multiplexed and therefore unbarcoded), we undertook further analyses to ensure that the PhiX sequencing standard is not present in our data. PhiX reference genome sequences downloaded from NCBI (NC_001422.1) and Illumina (Illumina real-time analysis reference: https://support.illumina.com/sequencing/sequencing_software/igenome.html [accessed December 12, 2025]) were searched against all Waiwera viral contigs from this study via blastn (-perc_identity 80 –task blastn). Searches were conducted using the full PhiX reference genome as well as the reference genome spliced into fragments 100 bp in length. In both cases, the only hits to Waiwera viral contigs were over an alignable fraction of ≤45 bp at an average nucleotide identity ≥80% (or, alternatively, over an alignable fraction of ≤30 bp at ≥90% identity). Therefore, we are confident that all sequences (both representative and clustered sequences) contributing to the 12

vOTUs classified as *Monodnaviria; Sangervirae; Phixviricota; Malgrandaviricetes; Petitvirales; Microviridae* do not represent the PhiX Illumima sequencing spike-in standard, but are instead genuinely derived from the environment.

We have updated the manuscript to clearly state this, as follows:

> "PhiX was used as a standard at low concentration (1% v/v) in Illumina DNA sequencing libraries, and the resulting reads were removed in the unindexed/unbarcoded fraction of reads. As the Illumina PhiX sequencing standard is derived from a bacteriophage, we undertook further analyses to ensure it was not present in our data. PhiX reference genome sequences (5386 bp long) were downloaded from NCBI (NC_001422.1) and Illumina (Illumina real-time analysis reference: https://support.illumina.com/sequencing/sequencing_software/igenome.html [accessed December 12, 2025]) and searched against all viral contigs from this study via blastn (-perc_identity 80 –task blastn, BLAST v2.16.0)[104]. Searches were conducted using the full PhiX reference genome as well as the reference genome spliced into fragments 100 bp in length. In both cases, no hits were obtained that shared ≥80% identity to PhiX and >45 bp alignment length (or ≥90% identity and >30 bp alignment length), indicating no Phix sequencing standard contamination of our viral contigs." (lines 620-630)

Table S5. Combined environmental parameters: Shouldn't the analyses iteratively determine if environmental parameters are making an impact, and so one wouldn't want to include many correlated factors, like potentially strongly correlated nutrient values? Also, I think it would be better to try to model which parameters have an additive improvement to the correlations. Seemingly for example, MAGS + salinity alone could be tested instead of all combined environmental parameters together. If the correlation gets worse, it would suggest that it is host and not specifically salinity that plays a role. Notably, MAGS + combined environmental parameters are much worse than just correlations to MAGS.

>> RESPONSE: Simple Mantel tests previously included separate testing against each individual environmental variable and the 'combined environmental variables'. Partial Mantel testing was similarly conducted testing vOTU community dissimilarities against each individual environmental variable with separate testing of the 'combined environmental variables' (while controlling for MAG community dissimilarities, provided as the third parameter to *partial.mantel()*) (results presented in **Table S5**).

We thank the reviewer for their recommendation regarding testing each variable iteratively to select variables to be include in the 'combined environmental' data. We have incorporated this now, and have updated the text relating to Mantel test results as follows:

> "There was a strong correlation between prokaryotic (MAG) and viral (vOTU) Bray-Curtis dissimilarities across samples (Mantel test $r \geq 0.95$, p-values=0.001 for water or sediment) (**Figure 2e,f; Table S5**). … There was also a strong

correlation between vOTU community dissimilarities and changes in key combined environmental parameters (Euclidean distance matrices including DNPOC+salinity in water samples and salinity+sulphate in sediment samples; see Methods for further detail on selected variables) (Mantel test $r \geq 0.89$, p-values=0.001), as well as when testing salinity, sulphate, and DNPOC concentrations individually against viruses ($r \geq 0.84$, p-values=0.001) or prokaryotes ($r \geq 0.86$, p-values=0.001), with the exception of DNPOC in sediment samples. Iterative testing of environmental variables identified that DNPOC (water samples) and sulphate (sediment samples) each added explanatory value to the correlations with vOTU community changes beyond the effect of salinity alone. Correlations between combined environmental variables or salinity and vOTU composition were no longer apparent in water samples after taking the effect of MAG composition into account (partial 3-way Mantel test, p-values ≥ 0.15) (**Figure 2e; Table S5**), reinforcing the importance of host community composition on viral distributions. Although, the effect remained for DNPOC and nitrite in water samples ($r \geq 0.46$, p-values ≤ 0.05), and salinity, sulphate, nitrate, DNPOC, and the combined environmental variables in sediment samples (r=0.55 for combined and 0.65 for salinity, p-values <0.01; remaining variables $r \geq 0.22$, p-values ≤ 0.05). These results highlight the potential role of additional environmental parameters in structuring virus communities beyond host community structure or salinity alone. Increased mixing across the water column may mask some of the effect of environmental parameters on virus communities compared to the results seen here for sediment samples, however, this requires further study." (lines 146-173)

We have also updated the text relating to Mantel test methods, as follows:

"To assess key environmental parameters in combination, the highest statistically significant Mantel correlation statistic was first identified for each of the four datasets (vOTUs or MAGs, in water or sediment) based on results of simple Mantel tests against each individual environmental variable. The data for this variable was then used to initiate a 'combined environmental' variable. Next, iterative partial Mantel tests (p*artial.mantel(x, y, z)*, incorporating distance matrices of vOTU or MAG communities, the environmental variable being tested, and controlling for the 'combined environmental' variables data) were then conducted against each remaining significant environmental variable from the previous iteration. For each iteration, the significant variable with the highest correlation statistic was added to the 'combined environmental' data until no additional significant variables remained. The final 'combined environmental' data for each dataset (vOTUs or MAGs, in water or sediment) included the following: DNPOC+salinity (vOTUs; water); salinity+DNPOC+sulphate (MAGs; water); salinity+sulphate (vOTUs; sediment); and salinity+Mud.content+sulphate+DRP+pH (MAGs; sediment). Finally, we conducted simple Mantel tests comparing the distributions of vOTUs and MAGs (using Bray-Curtis dissimilarity matrices) with each other or with environmental factors (Euclidean distance matrices), and partial Mantel tests comparing

associations between changes in vOTUs distributions and environmental factors after taking into account MAG distributions." (lines 695-712)

In my opinion, the IMG/VR analyses don't add that much, and are a bit distracting. I focused most of my review on the more directly applicable results to the main thrust of the manuscript. The most interesting aspect of that data may be the halophiles, but here we also do not see a large (or any effect, via Figure 5b).

>> RESPONSE: We consider the IMG/VR data to be an important supportive addition to this study by further identifying that the trends observed in the Waiwera estuary data (including ecological niche partitioning of viruses, and the comparative roles of host-adaptation vs direct environmental adaptation) are not solely a characteristic of this estuary alone (or estuaries in general), but are also evident across diverse ecosystems globally. We have rephrased the end of the introduction to clarify that the purpose of the IMG/VR analyses is to confirm our findings are not unique to the estuary as follows:

"We then expanded our analysis to include all high-quality viral genomes in the IMG/VR database, to assess whether patterns observed were not unique to this estuarine dataset but were also reflected in a large global database of viruses from diverse ecosystems." (lines 68-71)

We have also revised figures to ensure it is immediately obvious which panels refer to the estuarine dataset and which of a smaller subset of panels refer to the IMG/VR dataset.

Since the MAG and vOTU analyses seem to be completely independent on the bioinformatic side, it seems likely that some vOTUs are also within MAGs which could drive some of the remarkably strong correlations. It would be valuable to characterize how much of the correlation between host and vOTUs could be driven by this.

>> RESPONSE: We have added an additional analysis of prophage cobinned within prokaryote MAGs that shows that cobinning of prophage within prokaryote MAGs is not a strong driver of the close associations identified between vOTU and MAG communities in this study. The scope of this analysis was limited to putative prophage, as integrated viruses replicating as a part of the host genome would directly correlate with host abundances (notwithstanding instances of active viral replication of induced prophage within a sub-population of host cells). In contrast, copy numbers of actively replicating lytic viruses would not be expected to correlate closely with host MAG abundances due to the complexities of predator-prey population dynamics and virus replication burst size resulting in many viruses produced per host cell.

This analysis has now been included in the manuscript, as follows:

"There was a strong correlation between prokaryotic (MAG) and viral (vOTU) Bray-Curtis dissimilarities across samples (Mantel test $r \geq 0.95$, p-values=0.001 for water or sediment) (**Figure 2e,f; Table S5**). When investigating the rates of

prophage cobinned within prokaryote MAGs, only four viral contigs (each clustered into distinct vOTUs) were identified as putative prophage based on excision via VIBRANT, VirSorter2, or CheckV and also cobinned into a prokaryote MAG, and that corresponded to the filtered data set of vOTUs ≥50% complete. This suggests that cobinning of prophage within prokaryote MAGs is unlikely to be a strong driver of the close associations identified between vOTU and MAG communities in this study. There was also a strong correlation between vOTU community dissimilarities and changes in key combined environmental parameters…" (lines 146-156)

Reviewer #2 (Remarks on code availability):

Well organized.

Reviewer #3 (Remarks to the Author):

In their manuscript, "Adaptations of DNA viruses are influenced by host and environment with proliferations constrained by environmental niche," Hoggard et al. investigated viral and microbial diversity and distributions across a salinity gradient in the Waiwera estuary. Viral and microbial diversity and distributions were characterized using shotgun metagenomics and metatranscriptomics of water and sediment, and the genomic features associated with non-saline, brackish, and marine viral community members, including amino acid composition, predicted protein isoelectric points, and GC content were characterized.

Based on their observations, the authors concluded that 1) sampled viral communities were dominated by members of the Caudoviricetes; 2) environmental barriers (e.g. salinity) constrain viral distributions; and 3) genomic features of environmentally constrained viral communities show signatures of environmental adaptation (lower isoelectric points, etc.), including evidence of osmoadaptation independent of similar adaptation by microbial hosts.

General Comments

The authors have done a nice job addressing the comments and concerns of the reviewers. Although the associations between amino acid composition and salinity remain weak, more accurate classification of viral proteins has resulted in larger R2 values in some cases (Figs. 4A, S4), and structural predictions indicate that these changes are at the protein surface as would be expected.

The authors' assertion that increases in acidic amino acid proportions (specifically Asp) in viral proteins are viral adaptations that are host-independent remains less convincing. This conclusion is based on an observed negative correlation between viral aspartic acid proportions (which increase with salinity) and MAG tRNA gene and

transcript proportions. The authors show that MAG Asp tRNA gene and transcript proportions decrease with salinity when these data are pooled by site (Fig. 5C). Yet, Asp tRNA transcript proportions show no significant change with salinity (and even display a slightly positive slope) on a per MAG basis (Fig. S9B). It is unclear why the authors used the pooled data set in this analysis rather than the proportions per MAG, but the pooled data set appears to contain potential confounding variables. The authors are encouraged to re-analyze their correlations using the per MAG proportions before drawing conclusions.

>> RESPONSE: We have updated our analyses and discussion on tRNA as follows:

(1) The plots correlating salinity with tRNA gene transcription in individual MAGs were included to present the variability in tRNA transcription patterns (previously **Figure S9b**, now **Figure S11b,d**). However, these analyses excluded zero values (i.e. tRNA genes that are present but untranscribed), which affected the trend of the linear models. We have revised the figures to include the zero values for tRNAs where MAGs were identified as present in a given sample. These are shown across two panels, one showing the linear models, and the other showing individual data points per tRNA gene. This shows a negative correlation between salinity and prokaryote MAG aspartate tRNAs transcription, and trends overall are consistent with the spearman's correlations shown in **Figures 5e** and **S12a** that compare the relationship between virus amino acid proportions and MAG tRNAs.

(2) We have expanded our analysis by adding corresponding correlations between prokaryote MAG amino acid proportions and MAG tRNA genes and gene transcript proportions (**Figures 5e** and **S12a**). This shows similar trends and adaptations between viruses and prokaryotes with respect to amino acid preference and prokaryote tRNA gene and transcript abundances (e.g. positive correlations for glutamate and negative for aspartate). However, tRNA pools cannot fully explain amino acid preferences in either viruses or their putative hosts.

We have added the following to the results and discussion based on this new analysis:

> "A similar association was observed between cognate tRNA gene transcripts and glutamic acid proportions in prokaryotes (significant for ribosomal proteins, and positive but non-significant for secreted/translocated proteins exposed to higher salinities and the whole protein fraction). … Additionally, the non-significant relationship observed for prokaryotic secreted/translocated proteins suggests that tRNA preference also does not fully explain protein osmoadaptations in prokaryotes." (lines 459-476)

> "Likewise, when considering prokaryote charged amino acid proportions and tRNA gene transcript frequencies, strong negative correlations were observed for aspartate (**Figure 5e**)." (lines 487-488)

> "… our results highlight a disconnect between prokaryote tRNA gene and transcript proportions and virus amino acid encoding within the Waiwera estuary, although a similar disconnect was also observed for prokaryotes." (lines 523-525)

Additionally, positive correlations (for example, between Glu amino acid proportions in viral proteins and MAG tRNA transcripts, Fig. 5D) may still represent instances of viral adaptation beyond a reflection of host adaptation. Have the authors looked at the rate of change in viral amino acid proportions with increasing salinity relative to the rate of change in MAG tRNA transcripts? This might indicate, even for positive correlations, changes in viral amino acid proportions above a background level expected from host adaptations to salinity. While not conclusive, this could bolster arguments that changes to amino proportions go beyond host adaptation.

>> RESPONSE: We have added a new supplementary figure presenting direct comparisons of linear modelling results for mean predicted protein amino acid proportions per genome in viruses and prokaryotes and the relative abundance of prokaryote transcription of tRNAs for each amino acid (**Figure S12b**), and have updated the manuscript, as follows:

> "Based on linear modelling results, for each increase of 1 ppt of salinity across the estuary, the relative abundance of MAG transcription of tRNAs for glutamic acid increased by 1.21×10^{-6}, while glutamic acid proportions in predicted proteins increased in viruses by 1.74×10^{-4} and in MAGs by 1.18×10^{-4} (**Figure S12b**). Together, these data indicate that host-adaptation could be a primary driver of the predicted changes in glutamic acid proportions in virus proteins in this study, and may have contributed to the small shift in overall acidity observed with salinity in the viral proteins as a whole (**Figure 4a**). However, these results also highlight that with increasing salinity in the estuary the relative proportion of glutamic acid in viruses and prokaryotes increased at approximately 140 (viruses) and 100 (prokaryotes) times the rate of increase in transcription of tRNAs for this amino acid (relative to the overall pool of prokaryote tRNA transcription). The reduced association observed for major tail proteins (**Figure 5e**) further suggests that factors beyond host tRNA repertoires or host-adaptation may also influence the acidity of these particular proteins. Additionally, the non-significant relationship observed for prokaryotic secreted/translocated proteins suggests that tRNA preference also does not fully explain protein osmoadaptations in prokaryotes." (lines 462-476)

Specific Comments

Figure 2 legend: The vOTUs per estuary zone add up to 857, but this number should be 858?

>> RESPONSE: One vOTU (vOTU_6681) was unassigned in the process of categorising vOTUs into estuary zones and was excluded from subsequent estuary zone-based analyses. This vOTU only had reads mapped at sediment site 3 (and no other sites). As noted in the methods, read coverage values for sites 3 and 7 were excluded from this categorisation process to limit edge effects between the freshwater-to-brackish and brackish-to-marine transition zones. The absence of detected genome coverage across all remaining sites resulted in no zone being assigned for this vOTU. We have now clarified this in the legends for Figures 2 and 3 as follows:

> "Numbers of vOTUs per estuary zone (non-saline, brackish, and marine) were 184, 109, and 494 for water, and 10, 24, and 36 for sediment, respectively. One vOTU (vOTU_6681) was not assigned to an estuary zone due to insufficient genome coverage and was excluded from subsequent analyses."

Lines 286-289: Are these analyses shown anywhere? If not, please indicate data not shown.

>> RESPONSE: We note that *Nature Communications* does not allow 'data not shown' statements. We have therefore included the referenced analyses of viral proteins from saline samples only (i.e. excluding freshwater sample sites) as an additional supplementary figure, **Figure S6**, and have added this reference in text, as follows:

> "In linear models comparing only viral proteins from saline samples we found that the effect sizes were reduced for major capsid and endolysin proteins (albeit significant, R^2 <0.01, *p*-values <0.02), and the trend was lost for major tail proteins (R^2 <0.01, *p*-values = 0.372) (**Figure S6**)." (lines 367-370)

Line 435: Changes in amino acid usage in viruses are referenced, but the Figure referenced is Fig. S9a. Did the authors mean Fig. S10?

>> RESPONSE: The reviewer is correct, and we have updated this reference to **Figure S10**.

Lines 412-417 and Fig. S8b,c: The authors suggest that tRNA repertoires do not cluster by phyla or estuarine zone. This conclusion is based on tRNA gene counts, but gene number does not always correlate with transcript number or codon bias. Please confirm this trend using transcript abundances.

>> RESPONSE: We have now included figure panels with additional analyses based on tRNA transcript coverage (now **Figure S9**). The maximum tRNA transcript coverage at any site was presented here to assess the extent of clustering by taxonomy or assigned estuary zone based on tRNA transcript abundances where each MAG was most actively transcribing. We have also updated the text to now reference both gene count and tRNA transcript analyses, as follows:

> "However, prokaryote tRNA repertoires (based on Bray-Curtis dissimilarities of gene counts and maximum transcription of tRNAs per amino acid) did not cluster according to phyla or estuary habitat type (**Figure S9b**). Likewise, when restricting analyses to a phylogenetically-related group of prokaryotes that were present across the estuary (actinobacterial Luna-1 subcluster genera *Rhodoluna* in non-saline and *Aquiluna* in saline water samples), gene counts and maximum transcription of tRNAs for each amino acid again did not cluster by lineage or estuary habitat type (non-saline, brackish, or marine) (**Figure S9c**)." (lines 446-453)

Reviewer #3 (Remarks on code availability):

The code is accessible, well documented, and usable.

Reviewer #4 (Remarks to the Author):

I found the authors adequately addressed all of reviewer's 1 comments in this revised manuscript. Overall, I found the reviewed manuscript clear, with results presented with the right literature context and caveats. I only have one minor suggestion: if I understand correctly, the authors study virus sequences obtained from cellular fraction metagenomes (l. 534-535: "Filtered water samples included viruses associated with cells or biofilms trapped on 0.22 micron filters."). This is not an issue in itself, and is actually also the case for most of the IMG/VR database. However it tends to yield datasets enriched in prophages (see doi 10.1186/s40168-024-01905-x), most of them being integrated into the host genome, and thus possibly having a stronger adaptation to host in terms of GC content and codon usage as a byproduct of this temperate lifestyle and host-integrated state. This may be worth mentioning, e.g. l. 390-401 or in the final summary l. 507-527.

>> RESPONSE: Thank you for the suggestion. It does seem plausible that samples such as ours that principally capture the cellular fraction could be enriched in prophage relative to viral particle only fractions, depending on whether conditions favoured lysogeny at the time of sampling. The results reported by Kosmopoulos et al. (2024) varied somewhat by habitat with virus-specific versus non-specific samples in some cases containing more predicted temperate viruses. Additionally, our own analyses do not include a robust categorisation of viruses as lytic and temperate phage and prophage. So, it is unclear whether there is a prophage bias. However, we have added a recommendation to the summary on the potential differences between lytic and temperate phages:

> "The modification of protein amino acid repertoire in response to environmental salinity appears to be a common adaptive strategy spanning both prokaryotes and viruses. Further studies are needed to determine the primary driver of these modifications in viruses (e.g. direct environmental adaptation, or indirectly

> acquired via adaptation to the host), the effect on virus fitness, and the influence of virus-host interaction strategy as employed by lytic versus temperate phages." (lines 565-569)

Reviewer #4 (Remarks on code availability):

The code repository includes the documentation and scripts necessary to reproduce the analysis.

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

The authors have significantly revised the manuscript and I think the framing has improved. The Mantel tests show that the viral community and bacterial community is almost perfectly correlated in both water column and sediment, with no added relationship between viruses and salinity or combined environmental parameters in the water column. Nevertheless, there are significant correlations between amino acid/protein traits with the environmental parameters.

I appreciate that the authors have softened a bit their title and 'environmental driver' part, and acknowledge the strong host correlation factor. But, I still think 'constrains' is too strong: DNA viruses have adaptations that correlate with ecological parameters.

>> RESPONSE: In the title and elsewhere in the manuscript 'constrained' refers to ecological niche partitioning of viruses and not to adaptation. We have therefore left this in the title, although we have rephrased slightly so that the title does not refer to "constrained by ecological niche" but rather "constrained to ecological niches". As noted in the manuscript and by the reviewer previously, it is well known that viruses are constrained by their prokaryote hosts. It is also well established that prokaryotes are constrained by ecological niche, and logical that viruses are therefore constrained to niches as shown in this study.

I believe the version did not address this point from reviewer 2: "Second, the issue of the small effect sizes of the protein/amino acid trains are now acknowledged, but it doesn't change the fact that they are seemingly miniscule." As a further comment, perhaps some literature references that provide context to the magnitude of the shift would be valuable.

>> RESPONSE: Small effects sizes were previously addressed. This section of text now also includes reference to the small magnitude of changes observed in prokaryotes:

> "The observed increase in ratios of acidic-to-basic amino acids with increased salinity, across this threshold, is small. However, charged amino acids are predominantly found at the exposed surfaces of proteins rather than on unexposed surfaces or buried within protein interiors[63], therefore it is not unexpected that effect sizes are small when considering the protein as a whole. Moreover, similarly small differences exist in the amino acid osmoadaptations among prokaryotic taxa[19,69]. Assuming the same trends in viruses as in prokaryotes, the variability observed within samples is, in part, also likely to be due to the influence of lineage-specific biases on osmoadaptive trends." (lines 373-380)

Figure 4a, it would be helpful to plot the individual points for the plots where possible (I think this is Nature Communications best practices), which may aid with interpretation, or may overwhelm the signal.

>> RESPONSE: The data underlying several of the plots in Figure 4a are very large (> 7,000 data points for some individual genes/proteins, and >700,000 data points for the 'all genes/proteins' category, as indicated in the figure). In this case, violin plots were chosen as they provide a better visual representation of the distribution at each salinity level, whereas the large number of individual points obscured these distributions and resulted in figures that were more difficult to interpret.

Reviewer #2 (Remarks on code availability):

Code is available.

Reviewer #3 (Remarks to the Author):

In their revised manuscript "DNA viruses are constrained by ecological niche and exhibit environmental adaptations in common with prokaryote hosts," the authors have done a thorough job of addressing the concerns of the reviewers. The manuscript is clear and the scope of the claims have been adjusted to coincide with the presented results. I have no further concerns or suggestions.

Reviewer #3 (Remarks on code availability):

The code is accessible, well documented, and usable.

Reviewer #4 (Remarks to the Author):

I thank the authors for addressing the remaining reviewer comments, and have no further suggestion at this stage.