nature portfolio

Peer Review File

Giant viruses as reservoirs of antibiotic resistance genes

Reviewer #1 (Remarks to the Author):

This work presents an overview of the ARGs encoded in the giant viruses' genomes (NCLDV) compared to phages. The authors searched for various ARGs in isolated viruses as well as metagenomes, including a newly published dataset originating from mine tailings. The analyzed metrics include, the number and percentage of ARGs, their type, the taxonomy and particularities of the carriers. To conclude this work, they show that ARGs from giant viruses have the potential to be functional by studying the resistance to trimethoprim procured by two different viral dfr genes in E. coli.

The authors did a great job by 1- identifying a missing subject from the literature that needed to be investigated, 2- giving themselves the means and data to answer to their question (positively).
Therefore I think this is absolutely suited for publication, after some clarifications.

Analysis

- Could you present the trees made to classify the giant viruses from mine tailings ?

- l. 137 : Have you given coverage information to METABAT2 ?

- Section 3. 2 : An important point is that you included large/jumbo phages in the study according to the genome sizes shown in table S2. This make your NCLDV/phage comparison more reasonable. It would be interesting to know if the acquisition of ARGs depends on genome size with a plot showing, for giant viruses and phages, the % of ARGs, compared to genome size of isolated genomes primarily. One would expect, that ARGS are auxiliary genes that are added with genome size-increase but we could be surprised (see further) !

-l. 373-378 : It would be interesting to get more details into the mechanisms of the ARGs with some concrete examples. Table S6 deserves to be in the main text. You could instead include a heatmap of the main ARG genes in different genomes according to their types (for example one gene per row, some of which are named clearly, and viral taxa by columns).

- Have you encountered any differences in ARGs types from the vertebrate and insect-infecting Poxviridae ? It is not clear to me why vertebrate-dependent Poxviridae would encounter many bacterias or fungi (l. 491 – 500) while infecting their host.

- Section 3.6 : What is the distribution of genomic distance in-between the IS and the ARGs ?

- l. 448 : And what is their sequence identity ?

- l. 562 : The idea that the ARGs could serve to interact with sympatric bacterias is facinating ! This last paragraph is very well written. Just to add on to that thought I would like to offer some observations that you are of course free to consider or not. Just to mention that we have found other ARGs in ancient viral genomes from the study https://doi.org/10.1038/s41467-022-33633-x. This includes dihydrofolate reductase and beta-lactamase related proteins in the 8 large genomes (> 500kb contigs). The data is in the NCBI as the 42,000 years old Hydrivirus

(OW988864), and other not complete genomes of various ages (bioproject : PRJEB47746). The two dihydrofolate reductase from those large genomes as well as 25 others in smaller contigs, were annotated as “Bifunctional dihydrofolate reductase/thymidylate synthase” (not in the main text but in the shared data). Thymidylate synthase can be considered a core gene. This could be a putative explanation to the large amounts of ARGs in giant viruses and their low Ka/Ks. Another example perhaps is the biochemical study of a ribonuclease, beta-lactamase in tupanvirus doi: 10.1038/s41598-020-78658-8. It might be that those genes have important functions in giant viruses but are targets of antibiotics, and thus, giant viruses maintain the mutations that infer the resistance as you implied.

Editing
- There are too many abbreviations. While they are supposed to make it faster to read, they actually often slow down the process. IGs could be replaced by “isolated genomes” (or isolates, although not scientifically exact), MLS by “macrolides, lincosamides, streptogramines”, MGE by “mobile genetic elements”…

- In a similar way, DHFR deserves to be explained once (simply as dihyrofolate reductase).

- If you wish to gain space, the section 3.2 (in the results), could eventually be simplified.

- l. 467 – 468 : excellent ! But then, l. 468-469 you go very much into detail. I had to read three times then look at the figure to understand the difference with the first sentence. Either make it clearer, or simplify with for example : “ The overall, and average possibility of giant virus IGs to carry ARGs was shown to be higher than that of giant virus MAGs (Fig. 1B-C).”

- l. 469-472 : I don’t understand what is the added value of that sentence. It makes it a bit confusing.

- l. 129 : Maybe make it clearer that you are describing the already published work with, for example : “Details of sampling strategy […] bioinformatic analysis leading to the contigs used it the current study are described in the associated paper [21].”
Because the first time I read that paragraph I thought you had analyzed new samples using the same protocol.

Reviewer #2 (Remarks to the Author):

This study examines the genomes and constructed MAGs of giant viruses infecting eukaryotes and finds that they carry antibiotic resistance genes (ARGs). Further, they carry ARGs at a higher rate even than bacteriophages. They found that ARGs tended to be associated with insertion sequences, and revealed evidence of horizontal gene transfer with bacteria as well as eukaryotes. They also carried out a validation study to demonstrate that one of the main ARGs carried was functional in E. coli. Overall, it is an interesting study given the unexpected findings, but it also suffers from raising more questions than answers and generally giving the impression of an exploratory study. The manuscript would benefit from a major overhaul to better clarify the

research questions guiding the study and to provide more robust logic to the motivation for the study and interpretation of the results.

Specific Comments:
Overall- the Abstract needs a tighter logical framework motivating this study and needs to lay out research questions up front-- In its present form, it conveys a very exploratory study in which the authors did "this and that" and observed "this and that"

line 22-23-- The opening sentence of the abstract "in the microbiota" is grammatically off.

line 24-25 abstract "ARG of those viruses"-- needs a bit of explaining in the Abstract itself- as it is not expected that viruses themselves carry ARGs, only that they accidentally pick them up during transduction--- though this study contributes to a body of knowledge indicating that this is not quite the case and viruses themselves can carry ARGs in their genomes.

lines 26-27 Abstract-- since these large viruses mainly infect eukaryotes--- it is not clear what the relevance would be of them carrying ARGs, since they encode resistance to antibiotics in bacteria?

lines 43-45--- Agreed that it is probably hard to get into this kind of detail in the Abstract, but the paper itself would benefit more from a "so what"? Knowing that giant viruses carry ARGs and there is some evidence that HGT occurred or is occurring (we don't have any sense of evolutionary history- or HGT rates)--what would it mean to "take this into account"?

Lines 56-57- This sentence is a bit convoluted because it starts off talking about HGT, but then ends by again mentioning HGT as if it is an additional thought.

line 60-61-- "transduction....is likely to be the most influential on the whole" <<mechanism of HGT>> is a subjective statement.... This is but one example. Work to make the Introduction more objective and logical.

Lines 79-80-- If NCLDVs infect eukaryotes- it's not clear how they participate in HGT among prokaryotes----

Lines 79-83-- This is the only coverage in the Introduction of giant viruses, the main topic of the paper--- thus, there is not a very compelling argument for why giant viruses should be studied in the context of ARGs. I suggest sacrificing most of the opening paragraph about "antibiotic resistance is a problem"-- which is very general (everyone in this field knows it's a problem)- and instead add information here to the Introduction to expand on why to study giant viruses.

line 84-- More clear research questions are needed than "To address such an important knowledge gap"--- again, this gives the impression that the study is highly exploratory.

line 103-105-- work on refining the significance statement "within the whole microbiota" is vague.

line 297-- "ability" is not the right word --- I think simply the phages carried fewer ARGs on average--- this study doesn't tell us about ability.

line 373-- this section on the resistance mechanisms carried by the giant viruses is an important one and could provide insight into whether the ARGs on the whole behave as ARGs in their day to day function, or as something else. The authors should consider whether there's a tendency to encode intrinsic ARGs that are derived from housekeeping genes, versus mobile ARGs enncoding clinically-important resistance.

line 388- not sure what you mean by "occasions"- which implies time-- maybe "instances" is better word choice?

Line 409- what function do you think the dfr genes could be playing in the NCLDV?

line 480-- clarify what is meant by "well-known" versus "novel" ARG. be more specific- some refer to "proto-ARG" as a gene that is an evolutionary precursor to ARGs....

line 484-- I think you mean "predicted" not "predicated"?

line 500-- I am not following the difference between how the authors define "possibility" and "ability".

line 509-512-- the second half of this sentence does not logically follow the first-- how does the use since 1962 connect to detection of corresponding ARGs in sewage bacteria and soil phages.

Reviewer #3 (Remarks to the Author):

The authors report a comprehensive survey of anti-microbial resistance genes in Nucelocytoviricota using their own Nucelocytoviricota MAGs and a large prior collection of Nucelocytoviricota genomes using DeepARG and Resfams. The demonstrate through expression of dfhr genes from a Pithoviridae and an Asfarviridae in Escherichia coli strains can confer an trimethoprim resistance phenotype.

Trimethoprim resistance genes have been previously reported in the Nucleocytoviricota.

Lausannevirus Encodes a Functional Dihydrofolate Reductase Susceptible to Proguanil
L Mueller 1 , P M Hauser 1 , F Gauye 1 , G Greub 2
Antimicrob Agents Chemother. 2017 Mar 24;61(4):e02573-16.
doi: 10.1128/AAC.02573-16. Print 2017 Apr

However, the authors overlooked this study and stated in their abstract “Minimum inhibitory concentration experiments with two Escherichia coli strains carrying trimethoprim resistance genes of NCLDVs provided the first evidence that NCLDV-encoded ARGs can confer an antibiotic resistance phenotype”

The title is also misleading. This paper includes available genomes in the phyum Nucleocytoviricota. It does not at any point distinguish between members of this phylum and the different subclades containing giant viruses (nor does it provide a working definition of a giant virus). The term antibiotic resistance spread commonly refers to the spread of resistance due to human use of antibiotics. The authors do not show that giant viruses contribute to the spread of resistance.

Nucleocytoplasmic large DNA viruses (NCLDVs) is an outdated dated term. It has been replaced by the Nucleocytoviricota which is the phyum name for this group of viruses.

The 3rd and 4th paragraph of the introduction are basically the abstract restated.

Because of these inaccuracies I did not review the article in further detail

Reviewer #4 (Remarks to the Author):

The study of Yi et al. evaluates the genomes of different giant viruses and MAGS and the presence of ARGs in these sequences. They compare the results obtained with giant viruses with those obtained in phage genomes and phage MAGs, observing a higher presence of ARGs in giant viruses (genomes and MAGs). They perform a descriptive study indicating that some families of giant viruses are more prone than others to contain ARGs, and that some ARGs are more frequently detected than others. They devoted some efforts to validate the functionality of two dfr genes (trimethoprim resistance), that are encoded in one family of giant viruses. By cloning these drf genes in E. coli, they observed an increase in the MIC resistance, confirming that, despite the genomic differences with the original bacterial genes, these genes are functional and may confer resistance.

The manuscript is well written an easy to follow. The work presented it is methodologically sound and opens interesting lines of study about the spread of ARGs in different viral populations. It also opens quite interesting insights about the biological role of ARGs in these viruses and the factors that have contributed to a positive selection of the maintenance of ARGs in the virus genome. In this sense some more scientific discussion about the implications of these findings is required.

I have some questions and comments:

Ln 66-70: The first reference showing the presence and abundance of ARGs in phages is not cited here and was published far before the ones cited. Please refer to Colomer-Lluch M, Jofre J, Muniesa M. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. PLoS One. 2011 Mar 3;6(3):e17549. doi: 10.1371/journal.pone.0017549.

Ln 69-70, Moreover, the above-mentioned publication (2011) is really the first study where the

functionality of the resistance conferred by ARGs (blaTEM) encoded in phages was validated. Correct the sentence ad reference accordingly.

Ln 146. Check-quality?

Ln 174. Correct orthologs

Ln 196. To construct the gene trees

Ln 168. If I am not wrong, ResFams database is not actively updated since 2015. Some other databases, some with a more restrictive information, should be used in parallel (optimally selecting others that are regularly updated, e.g ARGminer, CARD, MEGARes, NDARO, ResFinder or SARG) and the findings obtained with ResFams should be compared and confirmed. See for details the information in Papp M, Solymosi N. Review and Comparison of Antimicrobial Resistance Gene Databases. Antibiotics (Basel). 2022 Mar 4;11(3):339. doi: 10.3390/antibiotics11030339. Erratum in: Antibiotics (Basel). 2022 Aug 30;11(9): PMID: 35326803; PMCID: PMC8944830.

Ln 249. Unfortunately, this classification is outdated. In a recent classification of the international committee of viruses done in 2023, the order Caudovirales and the families Microviridae, Siphoviridae and Podoviridae , etc. have been abolished. (Turner et al., Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee Arch Virol. 2023; 168(2): 74. The equivalent for Caudovirales (Table S2) should be Caudoviricetes class, and there are many other families of phages not dependent on they morphology. This new taxonomical classification should be mentioned and updated.

Ln 260. In addition to the degree of identity, what was the coverage of each ARG detected in the different viruses? Indicate what criteria was used to select the minimal length to assign a ORF to an ARG and what was the rational of this criteria (or reference). This is very relevant since it is not the same to have a complete gene with 70 % of identity than a fragment with 100 % of identity but only covering a 20 % of an ARG. There is no data in the manuscript that allows to evaluate the coverage of the ARGs.

Ln 258, I do not see a list with all the ARGs found in the viromes. Just a table S6 where some of the most detected are indicated. This data is important and should be presented as supplementary data. For example, is important to know if efflux pumps have been included in this study and are considered as an ARGs or not. In fact, an efflux pump with unspecific removal of intracellular antibiotics and other substances should not be considered as an ARG, since it serves for antibiotic elimination as a collateral activity, since some efflux pumps release many other substances, and their function is not specific for antibiotic resistance and, consequently, they are not subjected to antibiotic selective pressure (do not increase in the presence of antibiotics). Another question are those efflux pumps specific for a given antibiotic. In figure 3 it seems that general efflux pumps have been included and this should be revised. In this sense, the use of different databases for ARG recognition mentioned above is useful since some more restrictive are not going to consider efflux pumps as a real ARG. The combined used of different databases could lead to a different set of results. Also, in Fig. 3 the item “multiple mechanisms” should be better described and indicated if these are specific for antibiotics or general

mechanisms with other purposes that in parallel confer antibiotic resistance. In this late situation, these should be removed.

In table S4 and S5 and in the text, please clarify the calculations to determine the possibility and ability of encoding ARGs by the different types of viruses. It is not clear in its present form.

Ln 314. As mentioned above, are MFS transporters, an efflux pump specific for tetracycline resistance or are used to export other antibiotics too? Just to clarify if these genes can be considered an specific resistance mechanisms or as general efflux pumps with unspecific activity

Ln 412. Why there is a functional conservation and stability in the evolution of the dfr gene within NCLDVs, that might explain its widespread presence in NCLDVs? What is the advantage this gene might confer to the virus, that its incorporation in the viral genome has been positively selected?.

Ln 420. The presence of IS is indicative that ARGs could have been mobilized by them, however there is no indication of the position of the ISs found. To be responsible of ARG mobilization, they must be located upstream and downstream of the ARGs. Have they evaluated the proximity of IS to ARGs? Flanking regions of the drf gene have been evaluated but IS are not indicated in Fig. 6. Are IS not present or the map is not presented in enough detail to indicate IS?

Ln 497, It may be added that this is particularly true considering that resistances to betalactams is one of the most prevalent in Poxviridae and betalactams in particular are widely used in human medicine.

Considering this, are there other relationships between the type of ARGs most abundant in certain families and the antibiotic pressure they may endure?

Figure 1, 3, fig S4, fig S7, the nomenclature of phage families should be revised according to the new taxonomy (or at least, it should be indicated that these families are not taxonomically considered anymore).

General questions:
• What is the biological meaning of giant viruses carrying ARGs? This might have an explanation in phages, since incorporation of the ARG provide bacteria with a resistance, hence improving its survival. At the same this is beneficial for the phage, since bacteria somehow “tolerate” the entrance of the phage carrying this beneficial gene. But what is the advantage for an eukaryotic cell to be infected by a giant virus and to acquire an ARG? This should be discussed more in depth quoting studies (if any) on the survival of eukaryotic cells in the presence of antibiotics, or ability of eukaryotic cells to express genes conferring resistance to antibiotics, etc.

• In the final part of the discussion, they quote an study indicating that infection of NCLDVs can be inhibited by intracellular microbes of their host. Although this is interesting, I do not envisage how the ARGs in the virus can cause a trouble to the intracellular microbes, inhibiting them. On the contrary, the intracellular microbes can acquire this ARGs and survive even better. The

rational for this reference here should be clarified.

• The study focused on ARGs in giant viruses. However, have the authors found other bacterial sequences in the giant viruses others than ARGs? Toxins, virulent genes, metabolic genes? This could be even more interesting than ARGs. For example, few studies described the ability of eukaryotic cells to express toxins encoded by phages. Could be possible that the same happens in this case?

Responses to the Reviewers' comments

Responses to Reviewer #1's comments

Reviewer #1 (Remarks to the Author):

This work presents an overview of the ARGs encoded in the giant viruses' genomes (NCLDV) compared to phages. The authors searched for various ARGs in isolated viruses as well as metagenomes, including a newly published dataset originating from mine tailings. The analyzed metrics include, the number and percentage of ARGs, their type, the taxonomy and particularities of the carriers. To conclude this work, they show that ARGs from giant viruses have the potential to be functional by studying the resistance to trimethoprim procured by two different viral dfr genes in E. coli.

The authors did a great job by 1- identifying a missing subject from the literature that needed to be investigated, 2- giving themselves the means and data to answer to their question (positively).

Therefore I think this is absolutely suited for publication, after some clarifications.
Response: We thank this reviewer for acknowledging the merits of our manuscript.

Analysis
- Could you present the trees made to classify the giant viruses from mine tailings?
Response: Yes, of course. In the revised manuscript (RM), the phylogenetic tree has been added to Supplementary Fig. 10.

- l. 137: Have you given coverage information to METABAT2?
Response: Yes. Before we run MetaBAT2, we calculated the coverage of all the contigs using Bowtie, Samtools, and the jgi_summarize_bam_contig_depths script. MetaBAT2

was run with the coverage file as one of the inputs. The method of binning has been revised accordingly (RM: Lines 553-558).

- Section 3. 2: An important point is that you included large/jumbo phages in the study according to the genome sizes shown in table S2. This make your NCLDV/phage comparison more reasonable. It would be interesting to know if the acquisition of ARGs depends on genome size with a plot showing, for giant viruses and phages, the % of ARGs, compared to genome size of isolated genomes primarily. One would expect, that ARGS are auxiliary genes that are added with genome size-increase but we could be surprised (see further)!

Response: Thank you. This is a good idea. In the RM, we have examined the correlations between the genomic potential of ARG carriage (i.e., the % of ARGs) of isolated viruses and their genome size. While a weak positive correlation between the genomic potential of ARG carriage and genome size was observed for overall isolated phages, no significant relationship was recorded for overall isolated NCLDVs. Somewhat surprisingly, a significant negative correlation between the genomic potential of ARG carriage and genome size was observed for several NCLDV families. These results indicated that the mechanisms by which NCLDVs acquired ARGs were likely not the same as those of phages. The relevant contents have been added to the RM (Lines 413-424).

-l. 373-378: It would be interesting to get more details into the mechanisms of the ARGs with some concrete examples. Table S6 deserves to be in the main text. You could instead include a heatmap of the main ARG genes in different genomes according to their types (for example one gene per row, some of which are named clearly, and viral taxa by columns).

Response: Thanks for the constructive comment. In the RM, the following three major revisions have been made. First, the Table S6 of the original manuscript has been moved to the main text as Table 1. Second, we have prepared a heat map showing the distribution of antimicrobial resistance gene families in different viral taxa (presented as Supplementary Fig. 3 of the RM). We deviated from the reviewer's suggestion to construct the heat map at the gene level, where each gene was presented in one row. We did so

because substantial dissimilarity between putative ARGs in NCLDVs and those in bacterial genomes was observed. As indicated in Supplementary Table 3 of the RM, only a minor fraction of putative NCLDV ARG sequences exhibited sequence similarity exceeding 80% with known ARG references. In established systems for ARG nomenclature, genes showing limited sequence similarity to known references are generally assigned novel gene names. For instance, Roberts' guidelines for naming macrolide resistance genes specify that genes with ≥ 80% amino acid identity are grouped under the same class and letter designation, while those with ≤ 79% identity receive different designations (Roberts et al. 1999). In this context, we think that it would be inaccurate to present ARGs at the gene level by simply adopting the gene names of the best alignment hits in ARG databases. Third, we have provided a detailed description of the resistance mechanisms of the most frequently detected antimicrobial resistance gene families right after providing a succinct overview on the broad resistance mechanism classification (e.g., "antibiotic inactivation" or "target alteration") (RM: Lines 202-222).

- *Have you encountered any differences in ARGs types from the vertebrate and insect-infecting Poxviridae? It is not clear to me why vertebrate-dependent Poxviridae would encounter many bacterias or fungi (l. 491 – 500) while infecting their host.*

Response: According to a suggestion of Reviewer #4, we have re-annotated ARGs of viral genomes with multiple methods and databases rather than mere DeepARG. Our newly obtained results showed that all the ARGs carried by *Poxviridae* were rifampin resistance genes. Therefore, we deleted the discussion differentiating vertebrate- and insect-infecting *Poxviridae*. Alternatively, given that some *Poxviridae* members can cause human smallpox and cowpox and that rifampin had been utilized for treatment, we have proposed that the occurrence of rifampin resistance genes in *Poxviridae* (whose genomes under investigation were obtained from host-associated environments) might be a direct result from the selective pressure caused by the usage of this antiviral agent (RM: Lines 425-433).

- *Section 3.6: What is the distribution of genomic distance in-between the IS and the ARGs?*

Response: In the RM, we have annotated not only ISs and endonucleases but also several other types of mobile genetic elements (MGEs), including transposases, integrases,

recombinases, resolvases, and relaxases. We have revealed a close spatial association between these MGEs and ARGs, with 37.1% of the MGEs co-occurring with ARGs being situating within 10 kb (upstream or downstream) of their corresponding ARGs (72.9% within 30 kb). More specifically, 24.3% of the ISs/transposases co-occurring with ARGs located within 10 kb of their corresponding ARGs. Note that the active range of ISs/transposases was generally recognized to be 10 kb (Jiang et al. 2019). Despite this, in the Discussion section of the RM, we have made a brief discussion on endonucleases (recently reported with an active range of 10 kb; Barth et al. 2023) rather than ISs/transposases, given that endonucleases were identified as the most dominant MGE type of giant viruses (Lines 497-506).

- l. 448: And what is their sequence identity?

Response: The two NCLDV *dfr* genes exhibited relatively low amino acid identity with their bacterial homologs (33.5% and 36.3%, respectively). Despite this, we speculated that these genes could function in bacteria, because another NCLDV-encoded dihydrofolate reductase that exhibited only 22.2% amino acid sequence identity with its homolog of *Saccharomyces cerevisiae* was reported previously to confer trimethoprim resistance when expressed in *S. cerevisiae* (Mueller et al. 2017). The relevant contents have been added to the RM (Lines 301-306).

- l. 562: The idea that the ARGs could serve to interact with sympatric bacterias is facinating! This last paragraph is very well written. Just to add on to that thought I would like to offer some observations that you are of course free to consider or not. Just to mention that we have found other ARGs in ancient viral genomes from the study https://doi.org/10.1038/s41467-022-33633-x. This includes dihydrofolate reductase and beta-lactamase related proteins in the 8 large genomes (> 500kb contigs). The data is in the NCBI as the 42,000 years old Hydrivirus (OW988864), and other not complete genomes of various ages (bioproject: PRJEB47746). The two dihydrofolate reductase from those large genomes as well as 25 others in smaller contigs, were annotated as "Bifunctional dihydrofolate reductase/thymidylate synthase" (not in the main text but in the shared data). Thymidylate synthase can be considered a core gene. This could be a

putative explanation to the large amounts of ARGs in giant viruses and their low Ka/Ks. Another example perhaps is the biochemical study of a ribonuclease, beta-lactamase in tupanvirus doi:10.1038/s41598-020-78658-8. It might be that those genes have important functions in giant viruses but are targets of antibiotics, and thus, giant viruses maintain the mutations that infer the resistance as you implied.

Response: Thanks for your insightful comment. Based on the two important papers shared by you and another important paper shared by Reviewer #3, the following two major revisions have been made in the RM.

First, a new paragraph has been added in the Introduction section of the RM to summarize the NCLDV-encoded ARGs reported by the three papers and to raise our scientific questions in the context of the findings of these papers. In brief, although a total of 35 NCLDV genomes were found to encode ARGs, the incidence of ARGs across the phylum *Nucleocytoviricota*, their evolutionary characteristics, their dissemination potential, and their association with virulence factors have not yet been explored. For more details, please refer to Lines 66-79 in the RM.

Second, in the Discussion section, two explanations for the widespread presence of ARGs in NCLDVs have been developed based on the three papers. In brief, one explanation is that certain ARG-encoded proteins could exert crucial functions in the reproduction of giant viruses while they were also antibiotic targets, and the other explanation is that some ARG-encoded proteins could have evolved to be pleiotropic rather than mere as an agent to resist antibiotics. For more details, please refer to Lines 434-484 in the RM.

Editing

- There are too many abbreviations. While they are supposed to make it faster to read, they actually often slow down the process. IGs could be replaced by "isolated genomes" (or isolates, although not scientifically exact), MLS by "macrolides, lincosamides, streptogramines", MGE by "mobile genetic elements" …

- In a similar way, DHFR deserves to be explained once (simply as dihyrofolate reductase).

Response: Thanks for your suggestions. To enhance the readability of our manuscript, we have made the following three adjustments regarding abbreviations:

(1) We have removed some uncommon or infrequently used abbreviations and replaced them with their full forms throughout the texts. For example, we have replaced "IGs" with "isolate genomes" or "isolates", "DHFR" with "dihydrofolate reductase", and "MLS" with "macrolide-lincosamide-streptogramin". However, in figures, these abbreviations have been retained to save space, with their full forms being provided in the corresponding figure captions.

(2) For commonly used abbreviations that might not hinder readers' understanding, we have defined them upon their first appearance in the main text. Such abbreviations included: "ARG" for antibiotic resistance gene, "NCLDV" for nucleocytoplasmic large DNA virus, "HGT" for horizontal gene transfer, "MAG" for metagenome-assembled genome, "HMM" for Hidden Markov Model, and "ORF" for open reading frame.

(3) For abbreviations that may not be familiar to all readers, we have provided definitions in subsections where they are heavily used. This ensures that readers can quickly find the meanings of abbreviations. In other subsections where these abbreviations are used less frequently, we have opted for their full forms. Such abbreviations include "MGE" for mobile genetic elements, "VF" for virulence factors, and "ABC-F" for F-subtype ATP-binding cassette proteins.

- If you wish to gain space, the section 3.2 (in the results), could eventually be simplified.
Response: Thanks for the suggestion. In the RM, we have tried to keep the description of the relevant results concise (Lines 121-161), although we recognized that the journal is flexible with regard to manuscript length.

- l. 467 – 468: excellent! But then, l. 468-469 you go very much into detail. I had to read three times then look at the figure to understand the difference with the first sentence. Either make it clearer, or simplify with for example : " The overall, and average possibility of giant virus IGs to carry ARGs was shown to be higher than that of giant virus MAGs (Fig. 1B-C)."
Response: Sorry for the confusion. Inspired by a related comment raised by both Reviewer #2 and Reviewer # 4, we believed that such a confusion could be likely attributed to the lack of clear definitions of the parameters used in Fig. 1B and 1C. Therefore, in the RM,

the parameters used in Fig. 1B and 1C have been defined as “Possibility of ARG carriage” and “Genomic potential of ARG carriage” respectively, with the detailed calculation methods for them also being provided (Lines 606-615). In addition, some changes have been made the relevant sentences (e.g., Lines 130-144 and 146-161). We hope these revisions have made our description of the results shown in Fig. 1B and 1C easy to follow.

- l. 469-472: I don’t understand what is the added value of that sentence. It makes it a bit confusing.

Response: Please refer to our response to your last comment on lines 467-468.

- l. 129: Maybe make it clearer that you are describing the already published work with, for example: “Details of sampling strategy […] bioinformatic analysis leading to the contigs used it the current study are described in the associated paper [21].” Because the first time I read that paragraph I thought you had analyzed new samples using the same protocol.

Response: The relevant sentence has been revised as suggested (RM: Lines 544-546).

Responses to Reviewer #2’s comments

Reviewer #2 (Remarks to the Author):

This study examines the genomes and constructed MAGs of giant viruses infecting eukaryotes and finds that they carry antibiotic resistance genes (ARGs). Further, they carry ARGs at a higher rate even than bacteriophages. They found that ARGs tended to be associated with insertion sequences, and revealed evidence of horizontal gene transfer with bacteria as well as eukaryotes. They also carried out a validation study to demonstrate that one of the main ARGs carried was functional in E. coli. Overall, it is an interesting study given the unexpected findings, but it also suffers from raising more questions than answers and generally giving the impression of an exploratory study. The manuscript would benefit from a major overhaul to better clarify the research questions guiding the

study and to provide more robust logic to the motivation for the study and interpretation of the results.

Response: We thank this reviewer for acknowledging the merits of our manuscript and for providing us constructive suggestions for improving the quality of our manuscript. In the revised manuscript (RM), the Abstract, Introduction, and Discussion sections have been largely rewritten to better clarify the research questions guiding the study and to provide more robust logic to the motivation for the study and interpretation of the results. Below, we would like to list two examples of the major revisions that we have made in the RM.

First, a new paragraph has been added in the Introduction section of the RM to summarize the NCLDV-encoded ARGs reported previously in the literature (i.e., three important papers recommended by Reviewer #1 and Reviewer #3) and to raise our scientific questions in the context of the findings of the previous papers. In brief, although a total of 35 NCLDV genomes were found previously to encode ARGs, the incidence of ARGs across the phylum *Nucleocytoviricota*, their evolutionary characteristics, their dissemination potential, and their association with virulence factors have not yet been explored. For more details, please refer to Lines 66-79 in the RM.

Second, in the Discussion section, two explanations for the widespread presence of ARGs in NCLDVs have been developed based on the three papers recommended by Reviewer #1 and Reviewer #3. In brief, one explanation is that certain ARG-encoded proteins could exert crucial functions in the reproduction of giant viruses while they were also antibiotic targets, and the other explanation is that some ARG-encoded proteins could have evolved to be pleiotropic rather than mere as an agent to resist antibiotics. For more details, please refer to Lines 434-484 in the RM.

Specific Comments:

Overall- the Abstract needs a tighter logical framework motivating this study and needs to lay out research questions up front-- In its present form, it conveys a very exploratory study in which the authors did "this and that" and observed "this and that"

Response: According to the comment, we have written the first several sentences of the Abstract section to better explain the context and research questions of our study. The relevant sentences in the RM read as follows:

"Nucleocytoplasmic large DNA viruses (NCLDVs; also called giant viruses), constituting the phylum *Nucleocytoviricota*, can infect a wide range of eukaryotes and exchange genetic material with not only their hosts but also prokaryotes and phages. A few NCLDVs were report to encode genes conferring resistance to beta-lactams, trimethoprim, or pyrimethamine, suggesting that they are potential vehicles for the transmission of antibiotic resistance genes (ARGs) in the biome. However, the incidence of ARGs across the phylum *Nucleocytoviricota*, their evolutionary characteristics, their dissemination potential, and their association with virulence factors remain unexplored." (Lines 22-29).

line 22-23-- The opening sentence of the abstract "in the microbiota" is grammatically off.
Response: Thanks for the reminder. The wrong expression has been deleted from the RM.

line 24-25 abstract "ARG of those viruses"-- needs a bit of explaining in the Abstract itself- as it is not expected that viruses themselves carry ARGs, only that they accidentally pick them up during transduction--- though this study contributes to a body of knowledge indicating that this is not quite the case and viruses themselves can carry ARGs in their genomes.
Response: Thanks for the comment. We have revised the relevant sentences as suggested. For more details, please refer to our response to your first specific comment.

lines 26-27 Abstract-- since these large viruses mainly infect eukaryotes--- it is not clear what the relevance would be of them carrying ARGs, since they encode resistance to antibiotics in bacteria?
Response: Thanks for the comment. In the Introduction section of the RM, we have specified that the presence of ARGs in NCLDVs indicated a potential role of NCLDVs in mediating the transmission of ARGs in the biome (Lines 66-77). Meanwhile, in the Discussion section of the RM, we have elaborated on the potential biological explanations and implications of the widespread presence of ARGs in NCLDVs (Lines 434-514).

lines 43-45--- Agreed that it is probably hard to get into this kind of detail in the Abstract, but the paper itself would benefit more from a "so what"? Knowing that giant viruses carry

ARGs and there is some evidence that HGT occurred or is occurring (we don't have any sense of evolutionary history- or HGT rates)--what would it mean to "take this into account"?

Response: Thanks for the comment. In the RM, the relevant sentence has been deleted from the Abstract section. Alternatively, in the Discussion section of the RM, we have elaborated on the potential biological explanations and implications of the widespread presence of ARGs in NCLDVs (Lines 434-514).

Lines 56-57- This sentence is a bit convoluted because it starts off talking about HGT, but then ends by again mentioning HGT as if it is an additional thought.

Response: Thanks for the comment. The relevant sentence has been rephrased as follows: "Antibiotic resistance can arise through point mutation or horizontal gene transfer (HGT), with the latter often being cited as a key driver of its rapid spread." (RM: Lines 45-47).

line 60-61-- "transduction....is likely to be the most influential on the whole" <<mechanism of HGT>> is a subjective statement.... This is but one example. Work to make the Introduction more objective and logical.

Response: Thanks for the comment. In the RM, the relevant sentence has been revised to "Of these, transduction, the phage-mediated transfer of genetic material, has been considered as a route for ARG exchange among prokaryotes, because genes conferring resistance to beta-lactam antibiotics, glycopeptides, macrolides, peptide antibiotics, and tetracyclines, were detected in phages from a variety of environments [3-9]." (Lines 48-52). Meanwhile, many other revisions have been made to improve the objectivity and logic of the Introduction section. For example, as mentioned in our response to your general comment, a new paragraph has been added in the Introduction section of the RM to summarize the NCLDV-encoded ARGs reported previously in the literature and to raise our scientific questions in the context of the previous findings (Lines 66-79).

Lines 79-80-- If NCLDVs infect eukaryotes- it's not clear how they participate in HGT among prokaryotes----

Response: Sorry for the confusion. This sentence was not to express that NCLDVs can participate in HGT among prokaryotes directly. Instead, we aimed to say that some NCLDVs are potential effective vehicles of DNA transfer between eukaryotes and prokaryotes. To avoid confusion, revisions have been made to the relevant sentence (RM: Lines 63-65).

Lines 79-83-- *This is the only coverage in the Introduction of giant viruses, the main topic of the paper--- thus, there is not a very compelling argument for why giant viruses should be studied in the context of ARGs. I suggest sacrificing most of the opening paragraph about "antibiotic resistance is a problem"-- which is very general (everyone in this field knows it's a problem)- and instead add information here to the Introduction to expand on why to study giant viruses.*

Response: According to this suggestion, two major revisions have been made in the Introduction section of the RM. First, the first paragraph of about “antibiotic resistance is a problem” has been shorten to a single sentence (Lines 44 and 45). Second, as mentioned in our response to your general comment, a new paragraph has been added to summarize the NCLDV-encoded ARGs reported previously in the literature and to raise our scientific questions in the context of the previous findings (Lines 66-79).

line 84-- More clear research questions are needed than "To address such an important knowledge gap"--- again, this gives the impression that the study is highly exploratory.

Response: Thanks for the suggestion. In the RM, we have tried to specify that the purpose of our study was to explore the incidence of ARGs across the phylum *Nucleocytoviricota*, their evolutionary characteristics, their dissemination potential, and their association with virulence factors (Lines 75-79).

line 103-105-- work on refining the significance statement "within the whole microbiota" is vague.

Response: Thanks for the comment. In the RM, the relevant sentence has been deleted from the Introduction section. Instead, in the Discussion section of the RM, we have

elaborated on the potential biological explanations and implications of the widespread presence of ARGs in NCLDVs (Lines 434-514).

line 297-- "ability" is not the right word --- I think simply the phages carried fewer ARGs on average--- this study doesn't tell us about ability.

Response: Thanks for the comment. In the RM, we have opted to replace "ability to carry ARGs" with "Genomic potential of ARG carriage" that was defined according to a similar term proposed by Eichorst et al. (2018). Please refer to our response to your specific comment on line 500 for more details on the method that was used to calculate "Genomic potential of ARG carriage".

line 373-- this section on the resistance mechanisms carried by the giant viruses is an important one and could provide insight into whether the ARGs on the whole behave as ARGs in their day to day function, or as something else. The authors should consider whether there's a tendency to encode intrinsic ARGs that are derived from housekeeping genes, versus mobile ARGs enncoding clinically-important resistance.

Response: Inspired by this insightful comment and those relevant comments from the other three reviewers, we have made the following two major revisions in the RM.

(1) In the Results section, we have provided a more detailed description of resistance mechanisms of the most frequently detected AMR gene families (especially the *dfr* genes, the F-subtype ATP-binding cassette protein genes, and the *ileS* genes). Please refer to Lines 202-222 for more details.

(2) In the Discussion section, two possible explanations for the widespread presence of ARGs in NCLDVs have been proposed. The first explanation is that certain ARG-encoded proteins could exert crucial functions in the reproduction of giant viruses while they were also antibiotic targets. Such an explanation is likely applicable to dihydrofolate reductase encoded by *dfr* gene and isoleucyl-tRNA synthetase encoded by *ileS* gene. The second explanation is that some ARG-encoded proteins could have evolved to be pleiotropic rather than mere as an agent to resist antibiotics. This explanation may be applicable to beta-lactamase genes and streptogramin vat acetyltransferase genes. For more details, please refer to Lines 434-484.

line 388- not sure what you mean by "occasions"- which implies time-- maybe "instances" is better word choice?

Response: Thanks for the suggestion. The word “occasions” has been replaced with “instances” (RM: Line 238).

Line 409- what function do you think the dfr genes could be playing in the NCLDV?

Response: Thanks for the question. In the RM, we have proposed that dihydrofolate reductase, encoded by *dfr*, could exert crucial functions in the reproduction of NCLDVs while it was also an antibiotic target (Lines 434-442).

line 480-- clarify what is meant by "well-known" versus "novel" ARG. be more specific- some refer to "proto-ARG" as a gene that is an evolutionary precursor to ARGs....

Response: Thanks for the suggestion. Given to the challenge of defining “well-known” and “novel” ARGs derived from the existence of various competing systems for naming a new gene (Roberts et al. 1999; Doi et al. 2008; Hall and Schwarz 2016), we have preferred not to use the words “well-known” and “novel” ARGs in the RM. As such, the relevant sentence has been revised as “Given that some viral ARGs were shown to exhibit low sequence identities with cellular ARGs (as low as 20.4%) [7,17,18], an exploratory threshold of sequence identity (i.e., 25%) was employed in our study.” (Lines 398-400).

line 484-- I think you mean "predicted" not "predicated"?

Response: Yes, you are right. The typo has been corrected (RM: Line 410).

line 500-- I am not following the difference between how the authors define "possibility" and "ability".

Response: Sorry for the confusion. In the RM, we have replaced the term “ability” with “genomic potential”. Moreover, we have provided more details on the definitions of “Possibility of ARG carriage” and “Genomic potential of ARG carriage” and on the methods that were used to calculate the “Possibility of ARG carriage” and “Genomic

potential of ARG carriage" of a given viral taxonomy group. The relevant contents read as follows (RM: Lines 606-615):

"Two quantitative parameters were used to describe the ARG carriage of a given viral taxonomic group. The first one was "Possibility of ARG carriage", which refers to the proportion of genomes within a given group that carry ARGs, expressed as a percentage of the total number of genomes in that group (Equation 1). The second one was "Genomic potential of ARG carriage", which is defined as the percentage of ARG-like ORFs relative to the total number of ORFs in a given genome (Equation 2) [86]. The genomic potential of ARG carriage of a given group can be then calculated as the average value of the genomic potential of all genomes within that group.

$$\text{Possibility of ARG carriage (\%)} = \frac{Number\ of\ ARG-carrying\ genomes}{Total\ number\ of\ genomes} \quad (1)$$

$$\text{Genomic potential of ARG carriage (\%)} = \frac{Number\ of\ ARG-like\ ORFs\ in\ a\ genome}{Total\ number\ of\ ORFs\ in\ a\ genome} \quad (2)"$$

line 509-512-- the second half of this sentence does not logically follow the first-- how does the use since 1962 connect to detection of corresponding ARGs in sewage bacteria and soil phages.

Response: Thanks for the comment. The relevant sentence has been deleted from the RM.

Responses to Reviewer #3's comments

Reviewer #3 (Remarks to the Author):

The authors report a comprehensive survey of anti-microbial resistance genes in Nucelocytoviricota using their own Nucelocytoviricota MAGs and a large prior collection of Nucelocytoviricota genomes using DeepARG and Resfams. The demonstrate through expression of dfhr genes from a Pithoviridae and an Asfarviridae in Escherichia coli strains can confer an trimethoprim resistance phenotype.

Response: We thank this reviewer for acknowledging the merits of our manuscript.

Trimethoprim resistance genes have been previously reported in the Nucleocytoviricota.

Lausannevirus Encodes a Functional Dihydrofolate Reductase Susceptible to Proguanil
L Mueller 1 , P M Hauser 1 , F Gauye 1 , G Greub 2
Antimicrob Agents Chemother. 2017 Mar 24;61(4):e02573-16.
doi: 10.1128/AAC.02573-16. Print 2017 Apr

However, the authors overlooked this study and stated in their abstract "Minimum inhibitory concentration experiments with two Escherichia coli strains carrying trimethoprim resistance genes of NCLDVs provided the first evidence that NCLDV-encoded ARGs can confer an antibiotic resistance phenotype"

Response: Sorry for missing the important reference. Based on this reference and two additional important references recommended by Reviewer #1, the following four major revisions have been made in the revised manuscript (RM):

(1) The incorrect statement "Minimum inhibitory concentration experiments with two *Escherichia coli* strains carrying trimethoprim resistance genes of NCLDVs provided the first evidence that NCLDV-encoded ARGs can confer an antibiotic resistance phenotype." has been deleted from the RM.

(2) The first several sentences of the Abstract section have been completely written to better explain the context and research questions of our study (Lines 22-29).

(3) A new paragraph has been added in the Introduction section of the RM to summarize the NCLDV-encoded ARGs reported by the three important references and to raise our scientific questions in the context of the findings of these papers. In brief, although a total of 35 NCLDV genomes were found to encode ARGs, the incidence of ARGs across the phylum *Nucleocytoviricota*, their evolutionary characteristics, their dissemination potential, and their association with virulence factors have not yet been explored. For more details, please refer to Lines 66-79 in the RM.

(4) In the Discussion section, two possible explanations for the widespread presence of ARGs in NCLDVs have been developed based on the three important references. In brief, one explanation is that certain ARG-encoded proteins could exert crucial functions

in the reproduction of giant viruses while they were also antibiotic targets, and the other explanation is that some ARG-encoded proteins could have evolved to be pleiotropic rather than mere as an agent to resist antibiotics. For more details, please refer to Lines 434- 484 in the RM.

The title is also misleading. This paper includes available genomes in the phyum Nucleocytoviricota. It does not at any point distinguish between members of this phylum and the different subclades containing giant viruses (nor does it provide a working definition of a giant virus). The term antibiotic resistance spread commonly refers to the spread of resistance due to human use of antibiotics. The authors do not show that giant viruses contribute to the spread of resistance.

Response: Thank you for raising these points. In the RM, the following two major revisions have been made to address the points:

(1) The title of our manuscript has been revised to "Giant viruses are unexpectedly large reservoirs of antibiotic resistance genes" (Line 2).

(2) In the Abstract and Methods sections of the RM, we have defined "giant viruses" as the members of the phylum *Nucleocytoviricota* according to the definition used by Aylward et al. (2021). Please refer to Lines 22, 23, 523 and 524 for more details.

Nucleocytoplasmic large DNA viruses (NCLDVs) is an outdated term. It has been replaced by the Nucleocytoviricota which is the phylum name for this group of viruses.

Response: Thanks for the comment. However, we cannot agree with your viewpoint that "Nucleocytoplasmic large DNA viruses (NCLDVs) is an outdated term.", given that we found nearly 400 papers published in the last three years (i.e., since 2022) when we did a search (on June 10, 2024) in Google scholar using "nucleocytoplasmic large DNA viruses (NCLDVs)" as the search keyword. Moreover, a considerable proportion of these papers were published in top journals. An excellent example is the article entitled "Giant virus biology and diversity in the era of genome-resolved metagenomics", which was published in Nature Reviews Microbiology in 2022 (Schulz et al. 2022). Within this important review article, "nucleocytoplasmic large DNA viruses (NCLDVs)" was used interchangeably with "*Nucleocytoviricota*" and "giant viruses". Therefore, we have preferred to keep

"Nucleocytoplasmic large DNA viruses (NCLDVs)" in the RM, wherein it was also used interchangeably with "*Nucleocytoviricota*" and "giant viruses" (e.g., Lines 22 and 23).

The 3rd and 4th paragraph of the introduction are basically the abstract restated.

Response: Thanks for your comment. In the RM, a new paragraph has been added in the Introduction section to summarize the NCLDV-encoded ARGs reported previously and to raise our scientific questions in the context of the previous findings (Lines 66-79). Meanwhile, in the last paragraph of the Introduction section (Lines 80-92), we have shifted to outline the analyses and experiments conducted, rather than to present a summary of the results obtained, to avoid redundant contents with the Abstract section. Despite these revisions, due to the nature of the Abstract section encompassing the research background and knowledge gaps, we recognized that some overlap between the Abstract section and the Introduction section seemed inevitable.

Because of these inaccuracies I did not review the article in further detail

Response: Thanks again for your comments. We have found that the comments provided by you and the other three reviewers have helped us a lot to create a stronger and more accurate manuscript.

Responses to Reviewer #4's comments

Reviewer #4 (Remarks to the Author):

The study of Yi et al. evaluates the genomes of different giant viruses and MAGS and the presence of ARGs in these sequences. They compare the results obtained with giant viruses with those obtained in phage genomes and phage MAGs, observing a higher presence of ARGs in giant viruses (genomes and MAGs). They perform a descriptive study indicating that some families of giant viruses are more prone than others to contain ARGs, and that some ARGs are more frequently detected than others. They devoted some efforts to validate the functionality of two dfr genes (trimethoprim resistance), that are encoded in one family

of giant viruses. By cloning these drf genes in E. coli, they observed an increase in the MIC resistance, confirming that, despite the genomic differences with the original bacterial genes, these genes are functional and may confer resistance.

The manuscript is well written an easy to follow. The work presented it is methodologically sound and opens interesting lines of study about the spread of ARGs in different viral populations. It also opens quite interesting insights about the biological role of ARGs in these viruses and the factors that have contributed to a positive selection of the maintenance of ARGs in the virus genome. In this sense some more scientific discussion about the implications of these findings is required.

Response: We thank this reviewer for acknowledging the merits of our manuscript. In the Discussion section of the revised manuscript (RM), we have elaborated on the potential biological explanations and implications of the widespread presence of ARGs in NCLDVs. For more details, please refer to our responses to your general questions listed below.

I have some questions and comments:

Ln 66-70: The first reference showing the presence and abundance of ARGs in phages is not cited here and was published far before the ones cited. Please refer to Colomer-Lluch M, Jofre J, Muniesa M. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. PLoS One. 2011 Mar 3;6(3):e17549. doi: 10.1371/journal.pone.0017549.

Response: Sorry for missing the important reference. In the RM, the reference has been cited (Lines 52-54, 732, and 733).

Ln 69-70, Moreover, the above-mentioned publication (2011) is really the first study where the functionality of the resistance conferred by ARGs (blaTEM) encoded in phages was validated. Correct the sentence ad reference accordingly.

Response: Sorry again for missing the important publication by Colomer-Lluch et al. (2011), which is truly the first study to validate the functionality of phage-encoded ARGs.

We have revised the incorrect sentence and updated the reference accordingly (RM: Lines 52-54).

Ln 146. Check-quality?

Response: Sorry for the confusion. In the RM, "checkv-quality" has been revised to "CheckV quality tiers" to avoid confusion (Line 572). In our study, we used the CheckV software to determine the completeness of each viral contig under investigation. The software generated an "quality_summary.tsv" output file with an attribute named "checkv-quality" for each viral contig, classifying the viral contigs into "Complete", "High-quality" (> 90% complete), or "Medium-quality" (50–90% complete), etc. We only retained those viral contigs with the quality classification of "Complete" or "High-quality" for further analysis (RM: Lines 572-574).

Ln 174. Correct orthologs

Response: Done as suggested (RM: Line 640).

Ln 196. To construct the gene trees

Response: Done as suggested (RM: Line 666).

Ln 168. If I am not wrong, ResFams database is not actively updated since 2015. Some other databases, some with a more restrictive information, should be used in parallel (optimally selecting others that are regularly updated, e.g ARGminer, CARD, MEGARes, NDARO, ResFinder or SARG) and the findings obtained with ResFams should be compared and confirmed. See for details the information in Papp M, Solymosi N. Review and Comparison of Antimicrobial Resistance Gene Databases. Antibiotics (Basel). 2022 Mar 4;11(3):339. doi: 10.3390/antibiotics11030339. Erratum in: Antibiotics (Basel). 2022 Aug 30;11(9): PMID: 35326803; PMCID: PMC8944830.

Response: Thanks for the constructive suggestion. You are right that ResFams database is not actively updated since 2015. In the RM, we have re-annotated the ARGs of viral genomes using multiple methods and databases rather than mere DeepARG. Briefly, we have employed DeepARG and four regularly updated databases (i.e., CARD, NDARO,

SARG, and NCBI NDARO). Sequences annotated by DeepARG or those by at least two of the four abovementioned databases were finally retained as ARGs. For more details, please refer to Lines 579-605 in the RM.

Ln 249. Unfortunately, this classification is outdated. In a recent classification of the international committee of viruses done in 2023, the order Caudovirales and the families Microviridae, Siphoviridae and Podoviridae, etc. have been abolished. (Turner et al., Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee Arch Virol. 2023; 168(2): 74. The equivalent for Caudovirales (Table S2) should be Caudoviricetes class, and there are many other families of phages not dependent on they morphology. This new taxonomical classification should be mentioned and updated.

Response: Thanks for the constructive suggestion. In the RM, we have updated the taxonomic classification for phages as suggested. Specifically, for the phage sequences from the CheckV database, we have adopted the taxonomic information provided in CheckV 1.5, which aligns with the latest ICTV classification system. We retained the sequences that can be classified at any taxonomic level of phages or have habitat information for our further analysis (Lines 531-537). For the phage sequences obtained from our own mine tailings metagenomes, we have re-classified them using the geNomad software with the "end-to-end" command and default settings, along with the genomad_db_v1.7 database that aligns with the latest ICTV classification system (Lines 574-576). Remarkably, upon careful checking the latest ICTV taxonomy table, we have noted that the family *Microviridae* is retained in the latest classification system, while the order *Caudovirales* and the families *Siphoviridae* and *Podoviridae* have been abolished (source: https://ictv.global/taxonomy; accessed on 2024.03.24).

Ln 260. In addition to the degree of identity, what was the coverage of each ARG detected in the different viruses? Indicate what criteria was used to select the minimal length to assign a ORF to an ARG and what was the rational of this criteria (or reference). This is very relevant since it is not the same to have a complete gene with 70% of identity than a

fragment with 100% of identity but only covering a 20% of an ARG. There is no data in the manuscript that allows to evaluate the coverage of the ARGs.

Response: Thank you for raising this important question. In the original manuscript, we utilized the option "--arg-alignment-overlap 0.6" from DeepARG to filter sequences with a target coverage greater than 60%. However, in the RM, we have updated our ARG annotation approaches as suggested and refined our criteria based on a systematic literature search. Specifically, we have employed an 80% alignment coverage threshold for both query and target sequences (RM: Lines 584 and 592). The selection of 80% alignment coverage is supported by its common usage in recent studies (Wang et al. 2023; Liu et al. 2024), alongside the frequently cited threshold of 70% (Che et al. 2022; Zhang et al. 2022; Liu et al. 2023). While many previous studies adopted only the query coverage or target sequence coverage, we have preferred to increase the annotation precision by applying this coverage threshold to both target and query sequences. We appreciate your valuable input, which has prompted us to enhance the accuracy and clarity of our methodology.

Ln 258, I do not see a list with all the ARGs found in the viromes. Just a table S6 where some of the most detected are indicated. This data is important and should be presented as supplementary data. For example, is important to know if efflux pumps have been included in this study and are considered as an ARGs or not. In fact, an efflux pump with unspecific removal of intracellular antibiotics and other substances should not be considered as an ARG, since it serves for antibiotic elimination as a collateral activity, since some efflux pumps release many other substances, and their function is not specific for antibiotic resistance and, consequently, they are not subjected to antibiotic selective pressure (do not increase in the presence of antibiotics). Another question are those efflux pumps specific for a given antibiotic. In figure 3 it seems that general efflux pumps have been included and this should be revised. In this sense, the use of different databases for ARG recognition mentioned above is useful since some more restrictive are not going to consider efflux pumps as a real ARG. The combined used of different databases could lead to a different set of results. Also, in Fig. 3 the item "multiple mechanisms" should be better described and indicated if these are specific for antibiotics or general mechanisms with other purposes that in parallel confer antibiotic resistance. In this late situation, these should be

removed.

Response: Thank you for your insightful comment. In the RM, the following three major revisions have been made to address your concerns.

(1) Complete ARG list: We have now included a comprehensive list of all ARGs found in the studied viral genomes as supplementary data. This list, along with the identification method, representative alignment, alignment parameters, AMR gene family, ARG type, resistance mechanism, and other relevant information, can be found in Supplementary Table 3 (for NCLDVs) and Supplementary Table 4 (for phages).

(2) Efflux pump genes: In the RM, we identified 181 and 67 pump gene sequences from the studied NCLDVs and phages, respectively. According to the reviewer's suggestion, these efflux pump genes have not been considered as ARGs in our manuscript. That is, they have been excluded from total ARG count in the viral genomes and subsequent analysis. However, to provide more reference information, we have prepared two separate supplementary tables to showcase the specific annotation information of these efflux pump sequences (Supplementary Table 5 for NCLDVs and Supplementary Table 6 for phages).

(3) Multiple mechanisms: After the adoption of the updated ARG annotation approaches, there were no viral sequences labeled as "multiple mechanisms" in the RM (Fig. 2E and 2F).

In table S4 and S5 and in the text, please clarify the calculations to determine the possibility and ability of encoding ARGs by the different types of viruses. It is not clear in its present form.

Response: Thanks for the suggestion. In the RM, the term "ability" has been replaced with "genomic potential" to avoid confusion (Supplementary Table 7, Supplementary Table 8, and in the text). As per your suggestion, we have clarified the methods used to calculate the possibility and genomic potential of ARG carriage of different taxonomic groups of viruses. In the RM, the relevant contents read as follows (RM: Lines 606-615):

"Two quantitative parameters were used to describe the ARG carriage of a given viral taxonomic group. The first one was "Possibility of ARG carriage", which refers to the proportion of genomes within a given group that carry ARGs, expressed as a percentage of the total number of genomes in that group (Equation 1). The second one was "Genomic

potential of ARG carriage", which is defined as the percentage of ARG-like ORFs relative to the total number of ORFs in a given genome (Equation 2) [86]. The genomic potential of ARG carriage of a given group can be then calculated as the average value of the genomic potential of all genomes within that group.

$$\text{Possibility of ARG carriage (\%)} = \frac{Number\ of\ ARG-carrying\ genomes}{Total\ number\ of\ genomes} \quad (1)$$

$$\text{Genomic potential of ARG carriage (\%)} = \frac{Number\ of\ ARG-like\ ORFs\ in\ a\ genome}{Total\ number\ of\ ORFs\ in\ a\ genome} \quad (2)"$$

Ln 314. As mentioned above, are MFS transporters, an efflux pump specific for tetracycline resistance or are used to export other antibiotics too? Just to clarify if these genes can be considered an specific resistance mechanisms or as general efflux pumps with unspecific activity

Response: Thanks for the comment. In the RM, after the adoption of the updated ARG annotation approaches, which require stricter criteria, the two MFS transporter genes shown in Figure 2 of the original manuscript were no longer annotated as a MFS transporter gene or ARG. Therefore, the Figure 2 in the original manuscript and its relevant contents have been deleted from the RM.

Ln 412. Why there is a functional conservation and stability in the evolution of the dfr gene within NCLDVs, that might explain its widespread presence in NCLDVs? What is the advantage this gene might confer to the virus, that its incorporation in the viral genome has been positively selected?

Response: Thanks for raising the two interesting questions. By integrating additional information from the three important references recommended by Reviewer #1 and Reviewer #3, we have proposed that a possible reason for the functional conservation and stability in the evolution of the *dfr* gene within NCLDVs lied in that the protein encoded by it could exert crucial functions in the reproduction of giant viruses while it was also an antibiotic target. Such a scenario could explain its widespread presence in NCLDVs as well. For more details, please refer to Lines 434-456 in the RM.

Ln 420. The presence of IS is indicative that ARGs could have been mobilized by them, however there is no indication of the position of the ISs found. To be responsible of ARG mobilization, they must be located upstream and downstream of the ARGs. Have they evaluated the proximity of IS to ARGs? Flanking regions of the drf gene have been evaluated but IS are not indicated in Fig. 6. Are IS not present or the map is not presented in enough detail to indicate IS?

Response: Thanks for the insightful comment. In the RM, we have expanded our annotations of mobile genetic elements (MGEs) to include not only endonucleases and ISs that have been found previously in giant virus genomes but also transposases, integrases, recombinases, resolvases, and relaxases that have been typically identified as mobility-associated proteins in bacteria (Lines 617-627).

In order to address the reviewer's comment, we have conducted a detailed examination of the genomic proximity between ARGs and MGEs. We revealed a close association between ARGs and MGEs, which was evidenced by our finding that 37.1% of the MGEs co-occurring with ARGs located within 10 kb (upstream or downstream) of their corresponding ARGs (with 72.9% within 30 kb). More specifically, 24.3% of the ISs/transposases co-occurring with ARGs located within 10 kb of their corresponding ARGs. Note that the active range of ISs/transposases was generally recognized to be 10 kb (Jiang et al. 2019). Despite this, in the Discussion section of the RM, we have made a brief discussion on endonucleases (recently reported with an active range of 10 kb; Barth et al. 2023) rather than ISs/transposases, given that endonucleases were identified as the most dominant MGE type of giant viruses. Please refer to Lines 496-506 in the RM for more details.

As to the two giant virus genomes shown in Fig. 4A of the original manuscript, they both carried an endonuclease gene but not any ISs. Among the two endonuclease genes, one was within 10 kb of the ARG co-occurring with it. Please refer to Lines 290-298 in the RM for more details.

Ln 497, It may be added that this is particularly true considering that resistances to betalactams is one of the most prevalent in Poxviridae and betalactams in particular are widely used in human medicine.

Considering this, are there other relationships between the type of ARGs most abundant in certain families and the antibiotic pressure they may endure?

Response: Thanks for the comment. In the RM, after the adoption of the updated ARG annotation approaches, we have found that all the ARGs carried by genomes the family *Poxviridae* were rifampin resistance gene (*rif*). Certain members of the family *Poxviridae* can cause human smallpox and cowpox, with rifampin being utilized for treatment. Given that the *Poxviridae* genomes analyzed in this study were all obtained from host-associated environments, we have proposed that the presence of the *rif* genes within *Poxiviridae* likely stemmed from a direct selection under the pressure of the antiviral agent, rifampin. The revised discussion can be found in Lines 425-433 in the RM.

At this stage, we have not found other relationships between the type of ARGs most abundant in certain families and the antibiotic pressure they may endure.

Figure 1, 3, fig S4, fig S7, the nomenclature of phage families should be revised according to the new taxonomy (or at least, it should be indicated that these families are not taxonomically considered anymore).

Response: Thanks for the constructive suggestion. In the RM, the nomenclature of phage families has been updated in accordance with the latest ICTV classification system (e.g., Figs. 1 and 2). For more details, please refer to our response to your specific comment on Line 249 of the original manuscript.

General questions:

• *What is the biological meaning of giant viruses carrying ARGs? This might have an explanation in phages, since incorporation of the ARG provide bacteria with a resistance, hence improving its survival. At the same this is beneficial for the phage, since bacteria somehow "tolerate" the entrance of the phage carrying this beneficial gene. But what is the advantage for an eukaryotic cell to be infected by a giant virus and to acquire an ARG? This should be discussed more in depth quoting studies (if any) on the survival of eukaryotic cells in the presence of antibiotics, or ability of eukaryotic cells to express genes conferring resistance to antibiotics, etc.*

Response: Thanks for the insightful comment. In the RM, we have proposed three possible reasons for the widespread presence of ARGs in giant viruses. They can be summarized briefly as follows:

The first reason was related to the selection pressure stemmed from a given antiviral agent. Such a reason was likely applicable to the rifampin resistance gene (*rif*). The second reason was that certain ARG-encoded proteins could exert crucial functions in the reproduction of giant viruses while they were also antibiotic targets. Such proteins likely included dihydrofolate reductase encoded by *dfr* gene and isoleucyl-tRNA synthetase encoded by *ileS* gene. The third reason lied in that some ARG-encoded proteins could have evolved to be pleiotropic rather than mere as an agent to resist antibiotics. This explanation might be applicable to beta-lactamase genes and streptogramin vat acetyltransferase genes.

Remarkably, a prior study has shown that one dihydrofolate reductase gene from the giant virus family *Marseilleviridae*, when expressed in *Saccharomyces cerevisiae* (one of the best studied eukaryotes), conferred resistance to trimethoprim in the fungus (Mueller et al. 2017). This important previous study was recommended by Reviewer #3 and has been cited in the RM (including the discussion on the second possible reason).

For more details on the abovementioned contents, please refer to Lines 425-484 in the RM.

• *In the final part of the discussion, they quote an study indicating that infection of NCLDVs can be inhibited by intracellular microbes of their host. Although this is interesting, I do not envisage how the ARGs in the virus can cause a trouble to the intracellular microbes, inhibiting them. On the contrary, the intracellular microbes can acquire this ARGs and survive even better. The rational for this reference here should be clarified.*

Response: Thanks for the comment. In an important reference recommended by Reviewer #1, the authors revealed that after expression in *E. coli*, the beta-lactamase encoded by the giant virus (*Tupanvirus deep ocean*) was able to not only hydrolyze beta-lactam but also degrade RNA from its amoebal host and a variety of bacteria (Colson et al. 2020). We believed that this previous finding provided us a cue to envisage how the ARGs in the giant virus can cause a trouble to the intracellular microbes of its host. Therefore, in the RM, this reference has been cited to discuss the possible interactions between NCLDV-encoded

ARGs and intracellular microbes of their hosts, along with the citation of the reference showing that the infection of NCLDVs can be inhibited by intracellular microbes of their hosts (Lines 472-478). Meanwhile, the possibility of intracellular microbes to acquire NCLDV-encoded ARGs and thereby to survive even better under antibiotic stress has been also discussed in the RM (Lines 496-506).

• *The study focused on ARGs in giant viruses. However, have the authors found other bacterial sequences in the giant viruses others than ARGs? Toxins, virulent genes, metabolic genes? This could be even more interesting than ARGs. For example, few studies described the ability of eukaryotic cells to express toxins encoded by phages. Could be possible that the same happens in this case?*

Response: Thanks for the comment. Upon careful reading the relevant references, we found that: (1) the incidence of metabolic genes across the phylum *Nucleocytoviricota* and their evolutionary characteristics have been systematically examined (Moniruzzaman et al. 2020; Schulz et al. 2020); and (2) several toxin genes encoded by giant viruses have been reported (Deeg et al. 2018), although it remains unknown whether they can be expressed by eukaryotic cells. Therefore, in the RM, we have preferred to further explore virulence factors (VFs, including toxin genes) encoded by viruses and their association with ARGs. In brief, VFs were annotated by aligning viral protein sequences against the VFDB (Virulence Factors Database, http://www.mgc.ac.cn/VFs) using diamond blastp (RM: Lines 627-629), which showed that 68% of the studied NCLDV genomes and 4.88% of the studied phage genomes carried VFs (RM: Fig. 6A and 6E). Additionally, VF-positive viral genomes were found to exhibit a higher possibility of ARG carriage compared to VF-negative genomes (RM: Fig. 6B and 6F). For more details on the relevant contents, please refer to Lines 356-374 in the RM.

References

Aylward, F. O., M. Moniruzzaman, A. D. Ha and E. V. Koonin (2021). A phylogenomic framework for charting the diversity and evolution of giant viruses. PLOS Biology **19**(10): e3001430.

Barth, Z. K., D. T. Dunham and K. D. Seed (2023). Nuclease genes occupy boundaries of genetic exchange between bacteriophages. NAR Genomics and Bioinformatics **5**(3): lqad076.

Che, Y., X. Xu, Y. Yang, K. Břinda, W. Hanage, C. Yang and T. Zhang (2022). High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP- and human/animal-associated bacteria. Microbiome **10**: 16.

Colson, P., L. Pinault, S. Azza, N. Armstrong, E. Chabriere, B. La Scola, P. Pontarotti and D. Raoult (2020). A protein of the metallo-hydrolase/oxidoreductase superfamily with both beta-lactamase and ribonuclease activity is linked with translation in giant viruses. Scientific Reports **10**(1): 21685.

Deeg, C. M., C.-E. T. Chow and C. A. Suttle (2018). The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. eLife **7**: e33014.

Doi, Y., J.-i. Wachino and Y. Arakawa (2008). Nomenclature of plasmid-mediated 16S rRNA methylases responsible for panaminoglycoside resistance. Antimicrobial Agents and Chemotherapy **52**(6): 2287-2288.

Eichorst, S. A., D. Trojan, S. Roux, C. Herbold, T. Rattei and D. Woebken (2018). Genomic insights into the *Acidobacteria* reveal strategies for their success in terrestrial environments. Environmental Microbiology **20**(3): 1041-1063.

Hall, R. M. and S. Schwarz (2016). Resistance gene naming and numbering: is it a new gene or not? Journal of Antimicrobial Chemotherapy **71**(3): 569-571.

Jiang, X., A. B. Hall, R. J. Xavier and E. J. Alm (2019). Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. PLOS ONE **14**(12): e0223680.

Liu, W., C. Xu, T. Li, Z. Ren, S. Hao, Z. Chen, X. Huang and X. Wen (2024). Temporal dynamics and contribution of phage community to the prevalence of antibiotic resistance genes in a full-scale sludge anaerobic digestion plant. Environmental Science & Technology **58**(14): 6296-6304.

Liu, Z., X. Wan, C. Zhang, M. Cai, Y. Pan and M. Li (2023). Deep sequencing reveals comprehensive insight into the prevalence, mobility, and hosts of antibiotic resistance genes in mangrove ecosystems. Journal of Environmental Management **335**: 117580.

Moniruzzaman, M., C. A. Martinez-Gutierrez, A. R. Weinheimer and F. O. Aylward (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nature Communications **11**: 1710.

Mueller, L., P. M. Hauser, F. Gauye and G. Greub (2017). *Lausannevirus* encodes a functional dihydrofolate reductase susceptible to proguanil. Antimicrobial Agents and Chemotherapy **61**(4): 10.1128/aac.02573-02516.

Roberts, M. C., J. Sutcliffe, P. Courvalin, L. B. Jensen, J. Rood and H. Seppala (1999). Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. Antimicrobial Agents and Chemotherapy **43**(12): 2823-2830.

Schulz, F., C. Abergel and T. Woyke (2022). Giant virus biology and diversity in the era of genome-resolved metagenomics. Nature Reviews Microbiology **20**: 721-736.

Schulz, F., S. Roux, D. Paez-Espino, S. Jungbluth, D. A. Walsh, V. J. Denef, K. D. McMahon, K. T. Konstantinidis, E. A. Eloe-Fadrosh, N. C. Kyrpides and T. Woyke (2020). Giant virus diversity and host interactions through global metagenomics. Nature **578**: 432-436.

Wang, B., J. Xu, Y. Wang, E. Stirling, K. Zhao, C. Lu, X. Tan, D. Kong, Q. Yan, Z. He, Y. Ruan and B. Ma (2023). Tackling soil ARG-carrying pathogens with global-scale metagenomics. Advanced Science **10**(26): 2301980.

Zhang, J., D. Yu, L. Dian, Y. Hai, Y. Xin and Y. Wei (2022). Metagenomics insights into the profiles of antibiotic resistome in combined sewage overflows from reads to metagenome assembly genomes. Journal of Hazardous Materials **429**: 128277.

Reviewer #1 (Remarks to the Author):

The authors have thoroughly answered all my remarks.
I have no further comments and hope to see it published.

Reviewer #2 (Remarks to the Author):

The authors have carefully considered the comments from the previous round of review and have prepared a thoughtful revision, including more robust Discussion laying out hypotheses for the observations. I just have a few minor editorial suggestions:

line 38- Abstract "closely correlated" is casual- use precise statistical terminology.

line 50 "because" is mis-leading. I would break this into two sentences and delete because. First point is that transduction is considered a mode of ARG exchange. Second point is that there is evidence for this based on observing ARGs in phages from a variety of environments.

line 67 "pioneering"

line 262 "resistance"

Reviewer #3 (Remarks to the Author):

The authors have satisfactorily addressed my comments.

Reviewer #3 (Remarks on code availability):

The code for using R to make the figures is available at https://github.com/anotherXinzhu/NCLDV_ARGs. However the data imported to run the code is not in the github repository. Perhaps it is buried in the supplementary tables? It would be much easier to produce the results in the data was available in the format used by the R scripts.

Reviewer #4 (Remarks to the Author):

The authors have addressed and considered all my previous comments and have added new interesting references. I endorse the manuscript for publication

Responses to the Reviewers' comments

Responses to Reviewer #1's comments

Reviewer #1 (Remarks to the Author):

The authors have thoroughly answered all my remarks.

I have no further comments and hope to see it published.

Response: Thanks for the comment.

Responses to Reviewer #2's comments

Reviewer #2 (Remarks to the Author):

The authors have carefully considered the comments from the previous round of review and have prepared a thoughtful revision, including more robust Discussion laying out hypotheses for the observations. I just have a few minor editorial suggestions:

Response: We thank this reviewer for acknowledging our efforts to improve the quality of our manuscript and for providing us further suggestions listed below.

line 38- Abstract "closely correlated" is casual- use precise statistical terminology.

Response: The term has been changed to "significantly correlated" (revised manuscript: Line 38).

line 50 "because" is mis-leading. I would break this into two sentences and delete because. First point is that transduction is considered a mode of ARG exchange. Second point is that there is evidence for this based on observing ARGs in phages from a variety of environments.

Response: Done as suggested (revised manuscript: Lines 48-52).

line 67 "pioneering"

Response: Done as suggested (revised manuscript: Line 67).

line 262 "resistance"

Response: Done as suggested (revised manuscript: Line 264).

Responses to Reviewer #3's comments

Reviewer #3 (Remarks to the Author):

The authors have satisfactorily addressed my comments.

Response: Thanks for the comment.

Reviewer #3 (Remarks on code availability):

The code for using R to make the figures is available at https://github.com/anotherXinzhu/NCLDV_ARGs. However the data imported to run the code is not in the github repository. Perhaps it is buried in the supplementary tables? It would be much easier to produce the results in the data was available in the format used by the R scripts.

Response: Yes, the data imported to run the code are buried in the supplementary tables. In the revised manuscript, such data have been uploaded to the github repository as suggested.

Responses to Reviewer #4's comments

Reviewer #4 (Remarks to the Author):

The authors have addressed and considered all my previous comments and have added new interesting references. I endorse the manuscript for publication.

Response: Thanks for the comment.