nature portfolio

Peer Review File

Stabilization mechanism accommodating genome length variation in evolutionarily related viral capsids

Corresponding Author: Dr Simon White

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The results described in this paper are interesting and important. The authors present convincing arguments for a bacteriophage capsid organization which involves split hexamers stabilized by an accessory protein, allowing for a novel bacteriophage capsid organization where the phage can package larger genomes without having to change the T-number. They formulated their discovery in the form of a "theorem" on the conditions when the capsid can have a hexamer centered on a true 3-fold axis which describes the empirical situation well. I believe the paper is well written and the conclusions well supported and clearly described. As such it should be accepted for publication. My two suggestion for an optional revision are: 1. that the authors include a reference to the review Physics Reports 847 (2020) 1–102 in the section "Icosahedral capsids that can be built with split or skew hexamers" which I believe is a fitting overview of the symmetry principles of icosahedral capsids, and 2. that they include a reference to J. Phys. Chem. B 2016, 120, 26, 6051–6060 when first mentioning the internal pressure since this reference describes in detail the mechanism and characteristics of the DNA repulsive pressure in bacteriophage capsids..

Reviewer #2

(Remarks to the Author)
The manuscript by Podgorski et al describe the structures of the capsids of two related phages. The capsid has an interesting feature in that the hexameric capsomers contain a cleft down the center that is filled with a separate protein, gp4. The authors suggest that this allows this class of phage to package a larger genome.

The question of how to evolve different sized capsids is interesting. But the quality of this manuscript prevents publication as is. There are major issues with scientific rigor, presentation of the data, and interpretation that should prevent publication. It's a shame because I think that this could be a very interesting paper.

1) There are major issues with scientific rigor. Further explanation of reconstruction methodology, results, model-building, model-validation are absolutely necessary. The reporting is not in line with the standard of the field.
There are no micrographs nor 2D classes shown.
The Section "Cryo-Electron Microscopy Data Analysis" has an insufficient description of the reconstruction methodology.
Map assessment is missing as well; there are no FSC curves, nor local resolution maps shown.
Description of Model building should be more complete, as well. How many amino acids of each of the proteins were built?
The Materials and Methods section makes it sound as if the structural models were only refined against the map using Isolde. Is this correct? If so, is there a reason why this non-standard method was employed? Because using just Isolde (at least by my understanding) would not get the B-factors right and using this local refinement could mean some regions are better/worse refined than others. This is one of the reasons most atomic structures are refined using global refinement programs such as Phenix or Refmac. This raises real concerns as to the quality of the structures. Is there a reason the author's employed this building method?
Compounding this issue, there is no assessment of the quality of the atomic models. It is standard to have a table describing the refinement of the atomic model and the stereochemistry of the model. This should be reported. Have these structures been deposited at the PDB and EMDB? Are there PDB validation reports? These are standard to provide to reviewers.

2) The description of structures is poor. In general, the figures do not show what the authors describe in the text.

Figure 1 & 2 it is hard to see gp4. The cartoon is too thin to be seen.
From the results section (specifically, the section entitled “The E-loop is over-extended in the major capsid proteins that bridge the 2-fold chasm”), it sounds like the E-loop is longer in some subunits. If so, please make this more clear both in text and in figure 6. Supp fig 6 is extremely hard to parse, as are most structural figures herein. But this is potentially an important finding and should be moved into main text, because this would be v relevant. The authors will need to show a close-up of some density as well, particularly in the metal binding region.
In Figure 3, why are there two different types of hexamers? In a T=7 capsid, the seven subunits in the asymmetric unit are typically (always?) all six subunits within the hexamer itself and one of subunits in the pentamer (see next section). Did these different hexamers come out of 3D classification? How is this possible?

More minor point: Whenever showing an Alphafold model, it is most informative to use color each residue by its associated Alphafold confidence value (which should already be in the B-factor column of the pdb file that alphafold spits out.) This would help to contextualize where Alphafold predictions are correct and where they are less correct.

3) Insufficient analysis, interpretation and discussion of the data in the context of the field.
There are several claims that are incorrect.
Intro: “However, there has not been an empirical observation showing the adaptation of a specific stabilization mechanism to accommodate the evolution of a larger genome”. This is simply not true; the authors need not oversell to make the point that understanding the mechanisms of capsid size control is interesting.
Intro and elsewhere: “Split or skewed hexamers are common in procapsids but were not previously observed in mature capsids, which usually contain 6-fold symmetry hexamers.” Again, this is not true. In a T=7 capsid, the hexamer is NOT 6-fold symmetric. The asymmetric unit of a T=7 capsid includes all six subunits constituting the hexamer, and one pentameric subunit (the pentons are actually 5-fold symmetric).
The author’s mathematical analysis of capsid geometry has some nice ideas but these ideas are not new. I recommend the authors read the work of Mannige and Brooks, particularly their 2009 PNAS and 2010 PLOS One papers. The authors will see that their ideas that have are in the germination stage in this manuscript are well developed in the Mannige papers.
Capsid geometry and internal capacity have been studied for decades, but the authors don’t put their results into context against this large body of work. For example, many studies (too numerous to list here) have shown that capsid size can be altered by modifying the T number or by transitioning between isometric and prolate assemblies. This type of transition can be easily accomplished through evolution, as single point mutations can drive these transitions. This work is not discussed, which is a shame because I think the author’s point might be a good one if supported with solid data and robust comparisons.
Moreover, recent work from the Antson and Kelch groups (Bayfield et al PNAS; Stone et al Nat Comm 2019) have uncovered a mechanism of altering capsid size without changing capsid T number or prolate/isometric states. The potential mechanism for capsid size control described here appears to be conceptually similar (make a bigger capsomer) but in a unique and interesting way. There are other important parallels that should be discussed (E-loop length, packing of A-domains, etc) that are not discussed here.
A discussion of the mechanism identified by the authors (which is a cool mechanism!) should be placed in context against this rich and robust backdrop of prior knowledge.

Reviewer #3

(Remarks to the Author)
Podgorski et al submit a manuscript on their findings on bacteriophage that infect the Actinobacteria phylum. This is a continuation of their previous work and includes a refined structure of the Patience capsid to higher resolution than previously reported. The groups propose a novel capsid reinforcement mechanism that modifies the length of a stabilization protein to accommodate a larger genome while maintaining the same capsid size. Below are a few points to be considered and/or corrected prior to publication.

Page 5, Line 113: “Gp4 (Supplementary Fig. 1) is a small 102 amino acid…” Gp4 is not labeled in the figure. Is this the correct figure, if so, please label it.

Page 5, Line 113; Fig 1: “forms dimer within the hexamer”, it is hard to distinguish the monomers that compose the dimeric oligomers in Fig. 1 (image 4 of 5). Maybe it is a result of the lower resolution rendering in the provided PDF. Since the dimeric nature of gp4 is a major focus of the manuscript, could this be better emphasized either with alternate colors for members of a single dimer or blowup etc

Page 6, Line 132: “The minor capsid protein lacks the peanut-like protein domain found in Patience.” The peanut-like protein domain is not described in the manuscript, could this be clarified and/or referenced?

Page 6, Line 133: “However, the rest of the minor capsid protein of Adjutor is structurally very similar to Patience's (Supplementary Fig. 1)” Is this the correct figure callout? If so, similarity is not clear in this figure.

Page 7, Line 169: “Patience and Adjutor display the same internal capsid diameter (63 nm), but Patience packs a larger genome (70.5 kb) than Adjutor (64.5 kb). This constitutes a 10% increase in genome density, which in tailed phages implies an increase of internal pressure. The concomitant increase of the gp4 length and number of interactions in Patience with respect to Adjutor, was attributed to this increase in internal pressure. We note that the genome of Adjutor is approximately 30% larger than other well studied T=7 phages, for example lambda, HK97 and L5. This suggests that gradual changes in

gp4 length would help accommodate slight genome length increases in the evolution (or engineering) of tailed phages." Are the volumes of HK97 and L5 the same as Patience and Adjutor? Regarding volume, if the same or not the same, how does this affect the hypothesis?

Page 8, Line 185; Fig. 3: "The two chains in Patience and Adjutor (A and D) have E-loops that bridge the 2-fold chasm (Fig. 3). This is achieved by straightening and elongating the E-loop starting at Val73 (Supplementary Fig. 6)." The E-loop is not labeled explicitly. While it is reasonably intuitive, a label would be helpful for the non-structural biologist. The E-loop is not defined in the text with respect to the protein and perhaps a pointer to Supplementary Fig. 7 would be appropriate. A note on the E-loop in the context of other structures such as HK97 and other viruses would also be appropriate. For example, in HK97 the E-loop forms interactions with the N-terminus of other subunits.

Supplementary Fig. 6: Blue side chains on blue cartoon and green on green hard are to see. Hydrogens are not necessary, and the hydrogens make it cluttered and difficult to determine the residue in the figure. Residues should be labeled throughout, Ex. V58 and "Thus, the first salt bridge between Glu85 (E- loop) and Arg148 of the spine helix and the final salt bridge between Arg98 (E-loop) and Glu26 of the…": Are these residues relevant to Supplementary Fig. 6? Why is the E-loop cropped in the images?

Page 8, Line 211: "It has a seven-strand beta-sheet…" Define "it".

Page 12, Line 309: "This is consistent with the observation made computationally where about 2/3 of the capsids explored (12 out of the 20) could accommodate all skew/split hexamers while only about 1/3 (8 out of 20) displayed a 3-fold centered hexamer." What does this say about the larger/genome capacity? what about volume in split/non-split capsids? Presumably they are the same.

Can the authors provide more detail on the EM reconstructions including the reconstruction symmetry. Supplementary Table 1 lists I1. Were any additional symmetry considerations made for the split hexamers?

The document needs to be scanned for typographical errors: skey/split, etc

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I believe that the authors have taken all my comments into account and also successfully answered all the queries by other referees, therefore the revised form of this MS is ready for publication.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have made major improvements to the paper. However there are still several factors that should prevent publication as is.

1) The FSC shown for Adjutor in Fig S1 has red flags. The curve shown for the tight mask is extremely sharp, but this oddity is not explained in the text. Moreover, the FSC curve is not displayed in sufficient detail to ascertain data quality. The FSC curve should be shown to higher resolution than shown here so that the high-resolution correlation can be judged. Moreover, the image is so blurry that it is impossible to read the legend. The x-axes should be identical units for Adjutor and Patience.
2) In the discussion, the authors mention the thermophilic phages P74-26 and P23-45, as previously requested. But I don't think that the authors fully understand the findings from the thermophilic phage structures. These pair of papers on nearly identical phages show that these phages have vastly larger internal capacity than a typical T=7 phage because the E-loop is greatly extended. The longer E-loop allows for much larger hexamers and pentamers than a typical HK97 fold. Instead, the authors chose to compare P23-45 and P74-26 against each other. P23-45 and P74-26 are essentially identical phages and thus the comparison between them is not interesting nor meaningful towards the discussion. The relevant comparison is between the thermophilic phages and other t=7 phages, such as lambda. Thus, the increase in genome capacity rises from ~48kb (lambda) vs ~83kb (P23-45), equivalent to nearly doubling the internal capacity and not 1% as the authors claim. This needs to be fixed.

Additionally, the authors should remove the claim in the Discussion that: “there was no direct evidence on how the molecular mechanism associated with the trimers facilitates evolution of a larger capsid”, as the molecular mechanism described above is already apparent.
Because the authors are also predicating their proposed mechanism of capsid enlargening on an extended E-loop, the authors should put their observation the context that the thermophilic capsid structures display a longer E-loop.
3) The authors have cleaned up most of their errors in describing capsid symmetry. However a few remain. For example, in Figure 2A and in the text, the authors refer to chasm running down the centerline of the hexamer as the “2-fold chasm”. However, this should not be a true 2-fold axis, but rather a pseudo-2fold axis. Afterall, all six subunits of the hexamer are part of the asymmetric unit, indicating that the two sides of the chasm are not truly identical.

Minor points
1) I like the addition of Supplemental Figure 13. To make the author’s point more clear though, I think that they should point out the pseudo-2 fold axis for
2) Supplemental Fig 8 is a nice addition, wherein the authors make the claim that the E-loop is extended in some subunits versus others. The figure shows one E-loop that looks longer than the other, but it is hard to be sure because this could be simply a difference in perspective. The authors should quantify this putative extension, both to unambiguously establish that the E-loop does indeed extend, and also to provide the reader with a sense of how much the E-loop extends. This could be done through a simple calculation such as a measurement of the distance between the Ca at the tip of the E-loop to the Center of Mass of the rest of the HK97 fold (not counting the E-loop).

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Podgorski et al have submitted a revision to their manuscript.

The authors have adequately addressed the prior concerns of this reviewer.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have updated their manuscript and adequately addressed my concerns.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Table of Contents

RESPONSE TO REVIEWER 1

REVIEWER 1.0: The results described in this paper are interesting and important. The authors present convincing arguments for a bacteriophage capsid organization which involves split hexamers stabilized by an accessory protein, allowing for a novel bacteriophage capsid organization where the phage can package larger genomes without having to change the T-number. They formulated their discovery in the form of a "theorem" on the conditions when the capsid can have a hexamer centered on a true 3-fold axis which describes the empirical situation well. I believe the paper is well written and the conclusions well supported and clearly described. As such it should be accepted for publication.

RESPONSE 1.0: We thank again the reviewer for the thorough review and constructive comments. The changes listed below are highlighted in red in the manuscript.

REVIEWER 1.1: My two suggestion for an optional revision are: 1. that the authors include a reference to the review Physics Reports 847 (2020) 1–102 in the section "Icosahedral capsids that can be built with split or skew hexamers" which I believe is a fitting overview of the symmetry principles of icosahedral capsids, and 2. that they include a reference to J. Phys. Chem. B 2016, 120, 26, 6051–6060 when first mentioning the internal pressure since this reference describes in detail the mechanism and characteristics of the DNA repulsive pressure in bacteriophage capsids

RESPONSE 1.1: We appreciate the reviewer's recommendations regarding the two insightful articles that we had overlooked. We have cited them in the revised manuscript as requested.

CHANGES 1.1:
+ Line 274, Results section, reference added: Physics Reports 847 (2020) 1–102 is cited as Reference 29.
+ Line 53, Introduction section, reference added: J. Phys. Chem. B 2016, 120, 26, 6051–6060 is cited as Reference 5.

RESPONSE TO REVIEWER 2

REVIEWER 2.0: The manuscript by Podgorski et al describe the structures of the capsids of two related phages. The capsid has an interesting feature in that the hexameric capsomers contain a cleft down the center that is filled with a separate protein, gp4. The authors suggest that this allows this class of phage to package a larger genome.

The question of how to evolve different sized capsids is interesting. But the quality of this manuscript prevents publication as is. There are major issues with scientific rigor, presentation of the data, and interpretation that should prevent publication. It's a shame because I think that this could be a very interesting paper.

RESPONSE 2.0: We thank the reviewer for the thorough review and recommendations, which have improved the manuscript substantially. Below we describe the rationale of the changes implemented in response to each specific issues raised by the reviewer. The changes listed below are highlighted in red in the manuscript.

REVIEWER 2.1: There are major issues with scientific rigor. Further explanation of reconstruction methodology, results, model-building, model-validation are absolutely necessary. The reporting is not in line with the standard of the field. There are no micrographs nor 2D classes shown. The Section "Cryo-Electron Microscopy Data Analysis" has an insufficient description of the reconstruction methodology. Map assessment is missing as well; there are no FSC curves, nor local resolution maps shown. Description of Model building should be more complete, as well. How many amino acids of each of the proteins were built? The Materials and Methods section makes it sound as if the structural models were only refined against the map using Isolde. Is this correct? If so, is there a reason why this non-standard method was employed? Because using just Isolde (at least by my understanding) would not get the B-factors right and using this local refinement could mean some regions are better/worse refined than others. This is one of the reasons most atomic structures are refined using global refinement programs such as Phenix or Refmac. This raises real concerns as to the quality of the structures. Is there a reason the author's employed this building method? Compounding this issue, there is no assessment of the quality of the atomic models. It is standard to have a table describing the refinement of the atomic model and the stereochemistry of the model. This should be reported. Have these structures been deposited at the PDB and EMDB? Are there PDB validation reports? These are standard to provide to reviewers.

RESPONSE 2.1: The requests to improve the manuscript's rigor have been addressed. We have added a representative raw micrograph, the 2D classes as well as the FSC curves and local resolution maps as a new Supplementary Fig, 1.

The description of the model building method has been expanded, as well as descriptions of how many amino acids were built into the density. We apologize for the confusion about how Isolde was used. This has been corrected in the revised materials and methods. We used the standard method but incorporated Isolde since it is an excellent piece of software to get the model to fit well into the map during refinement. Refinement was done with phenix. We have updated Table 1 to show the assessment of the model.

We are sorry the reviewer did not have access to the validation reports and agree they are an important part of the review process. They were uploaded with the original manuscript, and we are not sure why the reviewer could not see them. If they can still not see them then they will have to follow it up with the editor. Models and maps were all deposited to PDB and EMDB as is standard before submission. The codes are present in section titled "Data deposition" on line 552 (prior version) and line 625 in the revised version. This section was present in the original submission.

REVIEWER 2.2: The description of structures is poor. In general, the figures do not show what the authors describe in the text. Figure 1 & 2 it is hard to see gp4. The cartoon is too thin to be seen. From the results section (specifically, the section entitled “The E-loop is over-extended in the major capsid proteins that bridge the 2-fold chasm”), it sounds like the E-loop is longer in some subunits. If so, please make this more clear both in text and in figure 6. Supp fig 6 is extremely hard to parse, as are most structural figures herein. But this is potentially an important finding and should be moved into main text, because this would be v relevant. The authors will need to show a close-up of some density as well, particularly in the metal binding region. In Figure 3, why are there two different types of hexamers? In a T=7 capsid, the seven subunits in the asymmetric unit are typically (always?) all six subunits within the hexamer itself and one of subunits in the pentamer (see next section). Did these different hexamers come out of 3D classification? How is this possible?

RESPONSE 2.2: Figure 1 and 2 have been modified to make gp4 more obvious. In Figure 1 a new box has been added that shows a zoomed in area of gp4 with cryo-EM density. Figure 2 has had the silhouette enhanced around gp4 to highlight the protein.

We have improved Supplementary Fig. 6 (now Supplementary Fig. 8) in an attempt to make it clearer. We have kept it in Supplementary information due to limits on the number of Figures.

Supplementary Fig. 10 has zoomed in boxes that show the model with the cryo-EM density overlaid as requested.

With regards to Figure 3 the two hexamers shown correspond, respectively, to Patience and Adjutor. The reviewer is correct about a T=7 capsid having 7 subunits that are the hexamer itself and one of the pentamer subunits. Each capsid analyzed displays one type of hexamer. However, the message of this figure is how gp4 interacts with the hexamer. It does not interact with the single pentamer subunit at all so adding it confuses the message. We are not claiming that the figure is the asymmetric unit. We have tried to add some language to make it clearer that it is only a hexamer that is being viewed and removed one of the hexamers that was showing just one copy of gp4. The hexamer was isolated in ChimeraX from the asymmetric model, not using 3D classification.

REVIEWER 2.2.1: More minor point: Whenever showing an Alphafold model, it is most informative to use color each residue by its associated Alphafold confidence value (which should already be in the B-factor column of the pdb file that alphafold spits out.) This would help to contextualize where Alphafold predictions are correct and where they are less correct.

RESPONSE 2.2.1: We have recolored the AlphaFold's reconstructions in Supplementary Fig. 11 to be colored by the alphafold prediction (so the values in the B-factor column of the pdb file). As can be seen the predictions are very good apart from the N-terminal regions.

REVIEWER 2.3: Insufficient analysis, interpretation and discussion of the data in the context of the field. There are several claims that are incorrect. Intro: “However, there has not been an empirical observation showing the adaptation of a specific stabilization mechanism to accommodate the evolution of a larger genome”. This is simply not true; the authors need not oversell to make the point that understanding the mechanisms of capsid size control is interesting. Intro and elsewhere: “Split or skewed hexamers are common in procapsids but were not previously observed in mature capsids, which usually contain 6-fold symmetry hexamers.” Again, this is not true. In a T=7 capsid, the hexamer is NOT 6-fold symmetric. The asymmetric unit of a T=7 capsid includes all six subunits constituting the hexamer, and one pentameric subunit (the pentons are actually 5-fold symmetric). The author’s mathematical analysis of capsid geometry has some nice ideas but these ideas are not new. I recommend the authors read the work of Mannige and Brooks, particularly their 2009 PNAS and 2010 PLOS One papers. The authors will see that

their ideas that have are in the germination stage in this manuscript are well developed in the Mannige papers. Capsid geometry and internal capacity have been studied for decades, but the authors don't put their results into context against this large body of work. For example, many studies (too numerous to list here) have shown that capsid size can be altered by modifying the T number or by transitioning between isometric and prolate assemblies. This type of transition can be easily accomplished through evolution, as single point mutations can drive these transitions. This work is not discussed, which is a shame because I think the author's point might be a good one if supported with solid data and robust comparisons. Moreover, recent work from the Antson and Kelch groups (Bayfield et al PNAS; Stone et al Nat Comm 2019) have uncovered a mechanism of altering capsid size without changing capsid T number or prolate/isometric states. The potential mechanism for capsid size control described here appears to be conceptually similar (make a bigger capsomer) but in a unique and interesting way. There are other important parallels that should be discussed (E-loop length, packing of A-domains, etc) that are not discussed here. A discussion of the mechanism identified by the authors (which is a cool mechanism!) should be placed in context against this rich and robust backdrop of prior knowledge.

RESPONSE 2.3: Regarding the criticism about our comment in the introduction, "However, there has not been an empirical observation showing the adaptation of a specific stabilization mechanism to accommodate the evolution of a larger genome", what we meant is that to our knowledge there was no direct comparison between closely related phages showing how a molecular mechanism would gradually accommodate a substantial increase of genome length. We are aware that different molecular mechanisms have been identified to reinforce capsids or alter the capsid size and T-number. However, the comparisons in those cases are between distant phages and it is not obvious how a molecular mechanism could gradually increase the genome length. However, as mentioned by the reviewer, an important case that we had overlooked and resonates directly with our research corresponds to the *Thermus* phages P23-45 and P74-26 (Bayfield et al PNAS 2019 and Stone et al Nat. Comm 2019). These two phages share a 95% genome similarity and have genomes that are only 900 bp different in length, representing a 1% increase in genome lengths. The structures have the same size. The major capsid proteins of both phages are identical. The decoration proteins that sit on the (quasi-)threefold axes share 79.45% similarity but are identical in length and display similar structures on the capsids. Thus, despite the difference in the amino-acid sequence of the minor proteins might be responsible for reinforcing the capsid for the additional 900 bp, the phages are too similar to infer a specific molecular mechanism that could be responsible for a more substantial genome length increase, like the 10% increase in genome length observed with the chasm-reinforcement mechanism reported for Adjutor and Patience in our manuscript. To reflect this insight in the revised manuscript, we have removed the comment mentioned by the reviewer from the introduction, added a new fragment in the introduction bringing up the case of the *Thermus* phages, and included also a fragment in the discussion section putting in context mechanisms associated with gradual genome increase versus the possibility of discrete genome increases associated with jumps between T-numbers (citing there references related to point mutations observed to change the T-number).

Regarding the criticism about the 6-fold symmetry hexamers, in the revised text we used the term near-6-fold symmetry, which is more precise and aligns with the language in the cryo-EM field, as is the case in one of the references mentioned by the reviewer (Bayfield et al PNAS 2019).

Regarding the criticism about our mathematical analysis, the skew hexamer framework we derived is compatible with the empirical observations reported in the article, while the framework mentioned by the reviewer (introduced by Mannige and Brooks) have predictions that are inconsistent with our observations. That framework investigated the different shapes of hexamers in icosahedral capsids based on the endo angle propagation from proteins forming the pentamers. It identified four types of hexamer shapes: flat (F), ruffled (R), wing (W), inverted wing (I), and single pucker (S). In our split hexamer framework, the F and R shapes would be incompatible with being skewed or split, while W, I, S are shapes compatible with skewed or split hexamers. For the capsid range analyzed in the endo angle framework, T = 1 to 243, only two had hexamer shapes compatible with all skewed hexamers, T=4 (W), and T=7 (S), which are also predicted in our framework. The reason the endo angle propagation

framework differs from our theorem for T>7 capsids is due to their more restrictive assumptions: The propagation of the pentameric endo angle and the use of rhomboid shape tiles, which impact on the direction of the angles being propagated. In fact, in the endo angle propagation framework, the T=12 capsid (Fig 2 in Mannige and Brooks 2010) is predicted to use hexamers with R, W, and I shape. The R shape coincides with our prediction of the 3-fold hexamer at the 3-fold axes, but the orientation of the R and W hexamers is incompatible with the orientation of the split hexamers in the 3D reconstruction of the SIO-2 procapsid (Figure 4c). One explanation for this discrepancy is that the orientation and rhomboid shape of the capsid tiles in the endo angle propagation framework was inspired by the single jelly-roll fold, while the SIO-2 procapsid is built from the HK97-fold which is oriented and shaped differently. Additionally, the interpretation of the T=12 procapsid of SIO-2 implies two types of hexamers (split and near six-fold), as predicted by our split hexamer framework, while the endo propagation framework predicts three types of hexamers for a T=12. We hypothesize that the lower number of hexamer types is because the constrain imposed by the pentamers may not propagate much further than the near capsomers in large capsids. Thus, the less restrictive skew framework proposed here is compatible with the capsids studied, while the endo propagation angle framework shows inconsistencies. However, a more thorough analysis with other capsids and protein folds would be necessary to determine if our predictions are as general as our current study suggests. We have adapted this comment as a new paragraph in the Discussion section.

CHANGES 2.3:
+ Line 65, Introduction section, fragment removed: "However, there has not been an empirical observation showing the adaptation of a specific stabilization mechanism to accommodate the evolution of a larger genome".
+ Lines 100–110, Introduction section, fragment added: "Phage Patience displayed a supersized T=7 icosahedral capsid [...] two recent *Thermus* phages P74-26 (83.3 kb) and P23-45 (84.2) described also to form supersized T=7 capsids reinforced by trimers (Stone et al. 2019 and Bayfield et al. PNAS). However, a novel intriguing feature of Patience was that its hexamers seem to display a large chasm along their 2-fold local symmetry line, effectively separating the hexamer into two halves."
+ Lines 409-438, Discussion section, fragment added: " The extension of the decoration protein in the chasm-reinforcement mechanism described here can accommodate [...] Investigating the relationship of packaging strategies and evolution capsid strategies is an important task to be pursued in future studies."
+ Line 111, Introduction section, edited fragment: "which usually contain near-6-fold symmetry hexamers."
+ Lines 449-480, Discussion section, "An approach to investigate the possible scenarios of capsid evolution is to consider [...] Alternatively, a mathematical approach, the endo angle propagation framework [...] However, a more thorough analysis with other capsids and protein folds would be necessary to determine if our predictions are as general as our current study suggests."
+ Figure 1 and 2 have been modified to make gp4 more obvious. In Figure 1 a new box has been added that shows a zoomed in area of gp4 with cryo-EM density. Figure 2 has had the silhouette enhanced around gp4 to highlight the protein.
+ Figure 3: one hexamer image removed due to confusion it was causing.
+ Supplementary Fig. 6 (now Supplementary Fig. 8) modified in an attempt to make it clearer.
+ Supplementary Fig. 10 has zoomed in boxes that show the model with the cryo-EM density overlaid as requested.
+ New S1 figure showing representative raw micrograph, the 2D classes as well as the FSC curves and local resolution maps
+ Expanded “Cryo-Electron Microscopy Data Analysis” (lines 526-551) and “Model Building” (599-618).
+ S11 HK97-folds re-colored to show AlphaFold confidence.

RESPONSE TO REVIEWER 3

REVIEWER 3.0: Podgorski et al submit a manuscript on their findings on bacteriophage that infect the Actinobacteria phylum. This is a continuation of their previous work and includes a refined structure of the Patience capsid to higher resolution than previously reported. The groups propose a novel capsid reinforcement mechanism that modifies the length of a stabilization protein to accommodate a larger genome while maintaining the same capsid size. Below are a few points to be considered and/or corrected prior to publication.

RESPONSE 3.0: We thank the reviewer for the thorough review and insightful comments. Below we describe the rationale and changes implemented in response to each of the specific issues raised by the reviewer. The changes listed below are highlighted in red in the manuscript.

REVIEWER 3.1: Page 5, Line 113: “Gp4 (Supplementary Fig. 1) is a small 102 amino acid…” Gp4 is not labeled in the figure. Is this the correct figure, if so, please label it.

RESPONSE 3.1: We have added the label “Gp4” above the gene box number 4 in Supplementary Fig. 2b so that it is clearer which gene is being discussed.

REVIEWER 3.2: Page 5, Line 113; Fig 1: “forms dimer within the hexamer”, it is hard to distinguish the monomers that compose the dimeric oligomers in Fig. 1 (image 4 of 5). Maybe it is a result of the lower resolution rendering in the provided PDF. Since the dimeric nature of gp4 is a major focus of the manuscript, could this be better emphasized either with alternate colors for members of a single dimer or blow up etc.

RESPONSE 3.2: Thank you for the kind suggestion. We have added a blow up of a single dimer on the right-hand side of Figure 1. We also added the fitted dimer model into the cryo-EM density.

CHANGES 3.2:
+ Modified S2 with gp4 label.
+ Modified Figure 1 by adding a zoomed in image of gp4.

REVIEWER 3.3: Page 6, Line 132: “The minor capsid protein lacks the peanut-like protein domain found in Patience.” The peanut-like protein domain is not described in the manuscript, could this be clarified and/or referenced?

RESPONSE 3.3: We have added a new Supplementary Fig. 6 that shows the minor capsid proteins of the two bacteriophages overlaid and the peanut-like domain labelled and referenced it in the text “The minor capsid protein lacks the peanut-like protein domain found in Patience (Supplementary Fig. 6).” We do not want to dwell on the peanut since it is not the focus of the paper. We have another paper in preparation that looks into these domains within the actinobacteriophage in more detail.

CHANGES 3.3:
+ Added S6.

REVIEWER 3.4: Page 6, Line 133: “However, the rest of the minor capsid protein of Adjutor is structurally very similar to Patience's (Supplementary Fig. 1)” Is this the correct figure callout? If so, similarity is not clear in this figure.

RESPONSE 3.4: Thank you for noticing this error. You are correct that it is the wrong figure. We have added Supplementary Figure 6 and referenced this correctly: “However, the rest of the minor capsid protein of Adjutor is structurally very similar to Patience's (Supplementary Fig. 6) and is part of the same protein family (Supplementary Fig. 2).”

CHANGES 3.4:
+ Added S6.

REVIEWER 3.5: Page 7, Line 169: “Patience and Adjutor display the same internal capsid diameter (63 nm), but Patience packs a larger genome (70.5 kb) than Adjutor (64.5 kb). This constitutes a 10% increase in genome density, which in tailed phages implies an increase of internal pressure. The concomitant increase of the gp4 length and number of interactions in Patience with respect to Adjutor, was attributed to this increase in internal pressure. We note that the genome of Adjutor is approximately 30% larger than other well studied T=7 phages, for example lambda, HK97 and L5. This suggests that gradual changes in gp4 length would help accommodate slight genome length increases in the evolution (or engineering) of tailed phages.” Are the volumes of HK97 and L5 the same as Patience and Adjutor? Regarding volume, if the same or not the same, how does this affect the hypothesis?

RESPONSE 3.5: The internal volumes of HK97, L5, and lambda are about 25-40% smaller than Adjutor and Patience. The larger volume and genome capacity of the supersized T=7 capsid associated with Adjutor and Patience is due to the presence of the trimers as in other supersized T=7 capsids like those of phages P74-26 and P23-45. This observation does not affect the hypothesis regarding the increase in genome capacity associated with the chasm-reinforcement mechanism. To avoid confusions, we have removed the comparison with classic T=7 capsids in the fragment mentioned by the reviewer. Instead, we have highlighted the differences in genome capacity with classic T=7 capsids when describing the supersized T=7 capsid of Patience in the Introduction section.

CHANGES 3.5:
+ Line 219, Results section, fragment removed: "We note that the genome of Adjutor is approximately 30% larger than other well studied T=7 phages, for example lambda, HK97, and L5.
+ Lines 100–108, Introduction section, fragment added: "Phage Patience displayed a supersized T=7 icosahedral capsid [...] capsids reinforced by trimers (Stone et al. 2019 and Bayfield et al. PNAS)."

REVIEWER 3.6: Page 8, Line 185; Fig. 3: “The two chains in Patience and Adjutor (A and D) have E-loops that bridge the 2-fold chasm (Fig. 3). This is achieved by straightening and elongating the E-loop starting at Val73 (Supplementary Fig. 6).” The E-loop is not labeled explicitly. While it is reasonably intuitive, a label would be helpful for the non-structural biologist. The E-loop is not defined in the text with respect to the protein and perhaps a pointer to Supplementary Fig. 7 would be appropriate. A note on the E-loop in the context of other structures such as HK97 and other viruses would also be appropriate. For example, in HK97 the E-loop forms interactions with the N-terminus of other subunits.

RESPONSE 3.6: We have made changes to this figure (now Supplementary Figure 8). We appreciate the comments and hope the new figure is clearer. We have added the entire HK97-fold and then zoomed into the relevant area so that the E-loop can be better contextualized. We have labelled it explicitly. Additionally, in the introduction we have mentioned the E-loop as one of the key domains of the HK97-fold, and, in the discussion, we have highlighted the role that the E-loop plays in Adjutor and Patience and related it to the variability of molecular stabilization strategies observed in the E-loop in other viruses, including P74-26, P23-45, BPP-1, HK97, and P22.

CHANGES 3.6:

+ Lines 73-75, Introduction section: Added sentence: " The HK97-fold is characterized by three structurally conserved domains: the axial domain (A-domain), the peripheral domain (P-domain), and the extended domain (E-loop)[7]."
+ Lines 438-447, Discussion section: Added fragment: " Such investigation would also benefit from including the way the E-loop stabilizes intra-capsomer and inter-capsomer connections [...] and the destabilization effect of single point mutations in the E-loop of P22 capsids[41]."

REVIEWER 3.7: Supplementary Fig. 6: Blue side chains on blue cartoon and green on green hard are to see. Hydrogens are not necessary, and the hydrogens make it cluttered and difficult to determine the residue in the figure. Residues should be labeled throughout, Ex. V58 and "Thus, the first salt bridge between Glu85 (E- loop) and Arg148 of the spine helix and the final salt bridge between Arg98 (E-loop) and Glu26 of the…": Are these residues relevant to Supplementary Fig. 6? Why is the E-loop cropped in the images?

RESPONSE 3.7: We have incorporated these changes into the figure. The hydrogens have been removed and the residues labelled. Hopefully they have made the figure clearer. "Thus, the first salt bridge between Glu85 (E- loop) and Arg148 of the spine helix and the final salt bridge between Arg98 (E-loop) and Glu26 of the…": These residues are not shown in Supplementary Fig. 8 (previously Supplementary Fig. 6) because they are further away than the elongated part of the E-loop. We are explaining that the contacts that the E-loop is making with other MCPs (the two salt bridges being discussed) are the same even though the distance between the two MCPs is greater across the chasm.

CHANGES 3.7:
+ Modified new S8 (previous S6) by removing hydrogens and attempting to make the figure clearer by reframing the HK97 fold.

REVIEWER 3.8: Page 8, Line 211: "It has a seven-strand beta-sheet…" Define "it".

RESPONSE 3.8: We have changed it to "The Patience-like major capsid proteins have a seven-strand beta-sheet in the A-domain".

CHANGES 3.8:
+ Changed line 259 to "The Patience-like major capsid proteins have a seven-strand beta-sheet in the A-domain"

REVIEWER 3.9: Page 12, Line 309: "This is consistent with the observation made computationally where about 2/3 of the capsids explored (12 out of the 20) could accommodate all skew/split hexamers while only about 1/3 (8 out of 20) displayed a 3-fold centered hexamer." What does this say about the larger/genome capacity? what about volume in split/non-split capsids? Presumably they are the same.

RESPONSE 3.9: The volume of capsids with non-split hexamers compared to the same capsid with split hexamers is expected to be the same. However, the chasm-reinforcing mechanism offers a molecular strategy to increase the genome length while conserving the capsid size, adding an evolutionary degree of freedom that could favor these capsids. In fact, a recent study identified that the distribution of genome lengths among isolated tailed phages had prominent peaks associated with genomes inferred to fit in T=4, T=7, and T=16 capsids, and when including metagenome-assembled genomes it included a peak in T=13 capsids (Luque et al 2020). These T-numbers coincide with the first group of capsids compatible with all skew/split hexamers, in favor of the evolutionary advantage in selecting capsids compatible with the chasm-reinforcing mechanism. We have included this information in the paragraph containing the quote highlighted by the reviewer.

CHANGES 3.9:
+ Lines 397-405, Results section, "The volume of a capsid with non-split hexamers compared to the same capsid [...] in favor of the evolutionary advantage in selecting capsids compatible with chasm-reinforcing mechanism."

REVIEWER 3.10: Can the authors provide more detail on the EM reconstructions including the reconstruction symmetry. Supplementary Table 1 lists I1. Were any additional symmetry considerations made for the split hexamers?

RESPONSE 3.10: We did try other symmetry considerations including I2, I3, I4 and C1 reconstructions and did not see any different in the final map apart from the resolution achieved. We did not try isolating the hexamers themselves using symmetry expansion, etc. We have added “A single class of particles was selected for high resolution 3D refinement applying I1 symmetry to the particles and using the de novo 3D model as the initial model. I2, I3, I4 and C1 reconstructions were also performed but we observed no difference in the final maps apart from the resolution achieved.” To the materials and methods.

REVIEWER 3.11: The document needs to be scanned for typographical errors: skey/split, etc

RESPONSE 3.11: The revised manuscript has been revised thoroughly to reduce any typographical errors.

RESPONSE TO REVIEWER 4

REVIEWER 4.0: I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

RESPONSE 4.0: We thank the reviewer for the contribution in evaluating the manuscript. We celebrate that Nature Comm recognizes the effort of early career researchers when co-reviewing an paper with more senior investigators.

The authors have made major improvements to the paper. However there are still several factors that should prevent publication as is.

1) The FSC shown for Adjutor in Fig S1 has red flags. The curve shown for the tight mask is extremely sharp, but this oddity is not explained in the text. Moreover, the FSC curve is not displayed in sufficient detail to ascertain data quality. The FSC curve should be shown to higher resolution than shown here so that the high-resolution correlation can be judged. Moreover, the image is so blurry that it is impossible to read the legend. The x-axes should be identical units for Adjutor and Patience.

We don't have an explanation for the oddity. The workflow was very standard in relion and mask creation was done using the standard protocol described in the relion tutorial provided by the Scheres group. It may be that the mask was too tight. All the raw data has been deposited into the various databases so other scientists can double-check the interpretations. We added the following to the text so that readers are aware of the oddity.

"We note that the curve corresponding to the tight mask exhibits a pronounced sharpness (Supplementary Fig. 1), which may indicate potential inaccuracies in the mask generation process.".

We have made the FSC legend better resolution and changed the x-axis units. We agree that the FSC curve would be better if it went to higher resolution. However, this is a common issue when working with large (relatively speaking) capsids. We are hitting the Nyquist limit with the resolution. This is because we had to balance the box size with memory usage on the memory card. The raw data is on EMPIAR so other scientists with more modern GPUs, with more memory, could extract a larger box size and overcome our issue. We have added the following text so that readers are aware of this limitation:

"We also note the FSC curve for Adjutor does not go to high enough resolution so that the correlation can be judged; this is likely due to issues with box size arising from memory limitation on the GPUs used for the reconstruction meaning that the Nyquist limit reached."

2) In the discussion, the authors mention the thermophilic phages P74-26 and P23-45, as previously requested. But I don't think that the authors fully understand the findings from the thermophilic phage structures. These pair of papers on nearly identical phages show that these phages have vastly larger internal capacity than a typical T=7 phage because the E-loop is greatly extended. The longer E-loop allows for much larger hexamers and pentamers than a typical HK97 fold. Instead, the authors chose to compare P23-45 and P74-26 against each other. P23-45 and P74-26 are essentially identical phages and thus the comparison between them is not interesting nor meaningful towards the discussion. The relevant comparison is between the thermophilic phages and other t=7 phages, such as lambda. Thus, the increase in genome capacity rises from ~48kb (lambda) vs ~83kb (P23-45), equivalent to nearly doubling the internal capacity and not 1% as the authors claim. This needs to be fixed.

Following the reviewers' suggestions in the Discussion section, we have clarified that P23-45 and P74-26, as well as Adjutor and Patience, form expanded T=7 capsids due to the formation of larger hexamers and pentamers. We have also emphasized that the capsids of these expanded T=7 phages accommodate nearly double packaging capacity compared to classical T=7 phages capsids. We have also clarified that the comparison regarding the 1% internal capacity is precisely to compare what is learned molecularly when comparing P23-45 and P74-26 in contrast with Adjustor and Patience. The bottom line is that in the case of P23-45 and P74-26, the differences are too small to appreciate how the capsid accommodates molecularly the additional genome length, while in the case of Adjutor and Patience (10% difference between them in genome length), one can appreciate how the reinforcement molecule in the chasm has a different length that correlates with the change in genome length.

The new fragment reads as follows: " These thermophilic phages adopt, like Adjutor and Patience, supersized T=7 capsids with a packaging capacity that nearly doubles the capacity of canonical T=7 capsids like lambda phage. In both cases, this increase in volume is due to the longer E-loops allowing larger hexamers and pentamers than for classic HK97-fold capsids. The trimeric minor proteins reinforcing the capsids, present in P74-26, P23-45, Adjutor, and Patience, were hypothesized to offer a molecular mechanism facilitating a gradual expansion of the hexamers and the internal capsid volume. However, in the case of P74-26 and P23-45, it was not possible to appreciate the specific molecular variations adding the gradual change in genome between them; the differences in genome length between phages P74-26 (84.2 kb) and P23-45 (83.3 kb) were only about 1% (a difference of 900 bp), their major capsid proteins were identical, and their minor proteins were 79% similar but had the same length and adopted similar structures."

Additionally, the authors should remove the claim in the Discussion that: "there was no direct evidence on how the molecular mechanism associated with the trimers facilitates evolution of a larger capsid", as the molecular

mechanism described above is already apparent.

Following the reviewers' suggestion, we have the removed the claim in the Discussion.

Because the authors are also predicating their proposed mechanism of capsid enlargening on an extended E-loop, the authors should put their observation the context that the thermophilic capsid structures display a longer E-loop.

Following the reviewer's suggestion, we have emphasized the role of the E-loop in the enlargening of P74-26 and P23-45 capsids: "This increase in volume is due to the longer E-loops allowing larger hexamers and pentamers than for classic HK97-fold capsids."

3) The authors have cleaned up most of their errors in describing capsid symmetry. However a few remain. For example, in Figure 2A and in the text, the authors refer to chasm running down the centerline of the hexamer as the “2-fold chasm”. However, this should not be a true 2-fold axis, but rather a pseudo-2fold axis. Afterall, all six subunits of the hexamer are part of the asymmetric unit, indicating that the two sides of the chasm are not truly identical.

We have corrected Figure 2A to say “pseudo 2 fold chasm”. We have also corrected throughout the text where the 2-fold chasm is mentioned to read “pseudo 2-fold chasm”.

Minor points
1) I like the addition of Supplemental Figure 13. To make the author’s point more clear though, I think that they should point out the pseudo-2 fold axis for

We have added a label and a dashed line to highlight the pseudo 2-fold axis for the Patience hexamer in this figure.

2) Supplemental Fig 8 is a nice addition, wherein the authors make the claim that the E-loop is extended in some subunits versus others. The figure shows one E-loop that looks longer than the other, but it is hard to be sure because this could be simply a difference in perspective. The authors should quantify this putative extension, both to unambiguously establish that the E-loop does indeed extend, and also to provide the reader with a sense of how much the E-loop extends. This could be done through a simple calculation such as a measurement of the distance between the Ca at the tip of the E-loop to the Center of Mass of the rest of the HK97 fold (not counting the E-loop).

Thank you for the suggestion. We have carried out that analysis and added the following to the text: “We used the distance measurement tool in ChimeraX to measure the distance between the C-alpha of Glu93 (the residue at the tip of the E-loop) and Thr70 (a residue within the core of the HK97-fold) and found that the distance within Chain A, with an elongated E-loop, and Chain B, without an elongated E-loop. We observed a difference of 4 angstroms between the two (Chain A: 64.4 angstroms and Chain B: 60.4 angstroms).”