nature portfolio

Peer Review File

Plasmid dynamics driving carbapenemase gene dissemination in healthcare environments: a nationwide analysis of closed Enterobacterales genomes

Corresponding Author: Dr Oon-Tek Ng

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

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for the opportunity to review this manuscript by Koh et al. that presents a comprehensive analysis of 5 years of carbapenemase plasmids from Singapore. This is a very unique dataset and makes an important contribution to the field. However, there are several issues with the manuscript that need to be addressed.

Major points:
Comment 1
The number of isolates included make the dataset extremely rich but much of the work that the paper presents is highly descriptive. At the core is a simple plasmid clustering and description of the plasmids. This descriptive nature limits the novelty and usefulness of the paper.

Comment 2
The paper focuses on 'successful' plasmids but does not give a robust definition of what this means. As far as can be ascertained, the authors equate 'successful' plasmids with those that are most frequent/prevalent. If this is the case, this should be stated explicitly. The work that follows notes that 'trim' genomes are associated with 'success' but there isn't a detailed analysis to further demonstrate this. For example, there is no functional work to demonstrate conjugation efficiency of the different plasmid clusters and therefore quantify 'success'. Furthermore, there was no attempt to look at other similar publicly available datasets to assess whether this finding is generalisable. These datasets are included in the references by the authors and could be used to test the hypothesis that 'trim' genomes lead to 'successful' plasmids.

Comment 3
To cluster plasmids, the authors use a k-mer based approach and Jaccard similarity. This is internally consistent but creates arbitrary plasmid clusters that the authors do not attempt to place in a larger context. It is only in the Discussion that we find that PC1 is a 'hybrid of pSA20021456.2-like plasmids (GenBank accession no. CP030221) and pKPCAPSS-like plasmids (GenBank accession no. KP008371)', and PC2 was 'was none other than pNDM-ECS01 (GenBank accession no. KJ413946)'. These observations should be included in the results along with a comprehensive attempt to place the plasmids in the dataset in a broader context. This was outlined to a certain degree in Supplementary Table 6 but it would also be helpful to include other plasmid typing approaches (e.g. MOBsuite clusters, plasmid taxonomic units). The key plasmid types could also be searched for in public data using BLAST. The second issue with the plasmid clustering approach is that it largely ignores any structural aspects. As an example, while two plasmids may only share 70% k-mer similarity, this could be because one of the plasmids incorporated a large block of genetic material from another plasmid. However, the remainder of the plasmid remains identical to the other plasmid and is clearly related. This is reflected in the findings of the relationships between PC1/PC7/PC43 and PC2/PC20/PC49.

Comment 4
The dataset seems to be somewhat underutilised. While it is reasonable to focus on the two most prevalent plasmid clusters, it would be helpful to at least make some key comments regarding the other plasmid types. Similarly, the dataset contains five years of data and it would be helpful to get a sense of temporal trends. PC1 and PC2 appear to be endemic, but are there outbreaks occurring of some of their 'subvariants' e.g. PC7 or other plasmid types. Was there any signal for other

plasmids potentially moving towards endemicity?

Comment 5
The figures are at times very difficult to interpret. Please see specific comments below but this relates to the need for clearer visualisation, as well as better explanation in the legends.

Minor Comments:
- Line 114: Is it possible to comment on the differing epidemiology between study sites and what the potential drivers of the differences are?
- Line 125: While there are more details contained in the methods, it would be helpful to contain a brief description of the clustering method that was adopted.
- Line 127. As per Comment 3 above, the Excel spreadsheet as supplementary material would benefit from further characterisation of each of the plasmids, including rep types, MOBsuite clusters, PTUs etc.
- Line 136-147: This section is very descriptive and repetitive. Is there any possibility of capturing these data in a table or figure and using text to state salient points eg. regarding diversity of K. pneumoniae isolates.
- Line 154: It would be helpful to briefly state how it was that transmission was 'genomically determined' to be plasmid or clonal. While a reference is provided, it is difficult for the reader to then have to search for this in Reference 11. In Reference 11, it appears that PlasmidSeeker was used but I'm wondering whether perhaps the k-mer plasmid clustering approach was used in the current paper? Please clarify.
- Line 167: Were there other isolates with multiple carbapenemase genes? If so, can these findings be summarised briefly?
- Line 182: Why are the authors comparing the core genomes of plasmids that are already known to be clearly different (i.e. carriage of different carbapenemase genes and already characterised as different by the clustering approach). It is not clear what the message from this analysis is.
- Line 223: How are 'clonal clusters' defined?
- Line 287: It would be helpful to discuss some of the other limitations of the work, e.g. methodological issues of the clustering approach, the absence of conjugation work etc.
- Line 301: Some further details regarding how isolates were collected would be helpful. What proportion were clinical isolates vs 'surveillance' isolates. How is a 'surveillance' isolate defined? Were these rectal swabs done to detect colonisation?
- Line 316: Some additional details regarding the quality parameters used to accept/reject assemblies are needed.
- Line 330: Was there a specific algorithm used for clustering (e.g. hierarchical clustering)? If so, can this be specified including whether there was a specific software tool used to do this?
- Figure 1C: this is difficult to interpret in its current form. Perhaps a network graph may be more useful with edges drawn between members of the same cluster?
- Figure 2: The message in this figure is not entirely clear. What does the y-axis represent? Is each tick a separate plasmid cluster? Is the intention to show that clonal transmission only contributed a small portion to the total number of plasmids transmitted? It is very difficult to interpret the species and hospitals involved in the current figure. Please consider representing this in a different way.
- Figure 4: This figure appears to show that PC7 and PC20 are subvariants of PC1 and PC2, respectively, with additional genetic material. Would consider making this a supplemental figure.
- Figure 5A and B. Is it possible to use two more differentiated palettes for the two panels? They currently seem very similar, leading to confusion between PC1 vs PC2 for example. These two panels are also very difficult to interpret as the labelling of the axes is not clear. How is time represented? What are each of the clonal clusters? Are they different species/ST combinations? Is the intention to show that most clonal clusters carry only one plasmid cluster?
- Figure 5C: The patient plots are very difficult to follow. Is there any way that the visualisation can be simplified?

Reviewer #2

(Remarks to the Author)
"Plasmid dynamics driving carbapenemase gene dissemination in healthcare environments: a nationwide analysis of closed Enterobacterales genomes" by Koh et al. is an interesting manuscript about an extensive study of sequenced CPE isolates from Singapore. In a previous study, the complete genomes of more than 1000 isolates were compared with Illumina short-read sequencing. In the current study they describe the use of Nanopore long-read sequencing in order to analyse the complete carbapenemase plasmids of these isolates. The paper is well written and give a thorough analysis of the data. Overall, the figures are small and contain too many panels. Specifically for figure 3 and 5 this needs to be addressed. It is a shame because the amount of data in these figures is great.
Line 30: How did you asses if genes were acquired via clonal transmission?
Line 45 and 283: What do you mean by 'maintenance of trim genomes'. I understand that these genomes are plastic and require only a low number of core genes, but I've not come across this term before. It is not explained in the study you reference.
Line 53: Although many CRE are resistant to most beta-lactams, some enzymes such as certain OXA-48-like enzymes, GES, IMI and FRI do not confer resistance to cephalosporins. This leads to underestimation of these genes when selective isolation is performed using cephalosporins. Please amend this sentence.
Line 68: Although I agree that plasmids cannot be assembled easily from short-read data, this data can be used more broadly than reference-based mapping, as you've shown in your previous study.
Line 131: In the provided reference I do not see any information about the functional replicon trfA or the classification of IncPe1.
Line 241-244: Please expand this section. The finding of highly similar plasmids in the same patient in diverse species

warrants additional details.
Line 312:Please include the full specifics on the Nanopore sequencing. Which sequence kit did you use. How was the basecalling performed. I am surprised that so many of your IncF multireplicon plasmids were fully assembled into circular contigs, for us these are difficult to get fully assembled without manual curation. Please include data on the depth of sequencing (highest, lowest, average).
Line 456: please adapt the author list to include the name of the first author..
Line 579: Please correct carbapenemase.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for providing a detailed and comprehensive response to my queries. I feel that the authors have done an outstanding job in trying to address them, with the manuscript being strengthened in turn.

I have no further comments at this point.

Reviewer #2

(Remarks to the Author)
Koh et al. have submitted a new version of their manuscript “Plasmid dynamics driving carbapenemase gene dissemination in healthcare environments: a nationwide analysis of closed Enterobacterales genomes”.
The authors have critically evaluated their manuscript and really improved the manuscript based on the previous suggestions made by myself and reviewer 1.

One of my concerns regarding the size of the figures has been addressed by reshuffling the panels. Although I recognise the amount of work that has gone into the preparation, I still believe that some figures are quite small for the reader to understand the message they convey. Most importantly, Figure 5 is still to crowded and panel C should be made into a separate figure.

We thank the reviewers for their thoughtful and constructive comments, which have helped us to improve the clarity and rigor of our manuscript. Below, we provide point-by-point responses to each comment and detail the corresponding revisions made. We are grateful for the opportunity to revise our work and hope that the changes address all concerns satisfactorily.

Line numbers refer to the clean manuscript.

Reviewer 1

Thank you for the opportunity to review this manuscript by Koh et al. that presents a comprehensive analysis of 5 years of carbapenemase plasmids from Singapore. This is a very unique dataset and makes an important contribution to the field. However, there are several issues with the manuscript that need to be addressed.

Major comments:

R1-C1

The number of isolates included make the dataset extremely rich but much of the work that the paper presents is highly descriptive. At the core is a simple plasmid clustering and description of the plasmids. This descriptive nature limits the novelty and usefulness of the paper.

R1-C2(a)

The paper focuses on 'successful' plasmids but does not give a robust definition of what this means. As far as can be ascertained, the authors equate 'successful' plasmids with those that are most frequent/prevalent. If this is the case, this should be stated explicitly.

>> Reply:

We have made amendments to the manuscript to address this. As recommended, we have explicitly stated our definition of "success": "PC1 and PC2 were considered evolutionarily successful plasmids due to their high prevalence, with each accounting for more than 25% of plasmids in the cohort (Table 1)" [lines 294-295]. The term "success", when mentioned, is also more overtly aligned with "evolutionary success" which is commonly quantified by abundance in a population [lines 50, 279, 485]. For clarity, we have also replaced the term "successful" with "predominant" where possible [lines 387, 426, 455].

R1-C2(b)

The work that follows notes that 'trim' genomes are associated with 'success' but there isn't a detailed analysis to further demonstrate this. For example, there is no functional work to demonstrate conjugation efficiency of the different plasmid clusters and therefore quantify 'success'. Furthermore, there was no attempt to look at other similar publicly available datasets to assess whether this finding is generalisable. These datasets are included in the references by the authors and could be used to test the hypothesis that 'trim' genomes lad to 'successful' plasmids.

>> Reply:

For clarity, have replaced the term "trim" with "conserved" [lines 52, 487]. As suggested, we looked at the datasets referenced in the manuscript and identified a similar institutional collection of carbapenem-resistant isolates systematically collected from 2002 to 2020, that had a sufficient number of closed plasmids (n=154 circularized *bla*$_{\text{IMP-4}}$ plasmids) required for this analysis (Macesic *et al*., 2023). The plasmids in this comparable dataset were mostly from clinical isolates, and did not include surveillance isolates. We analysed the data using the same plasmid clustering approach and observed a similar pattern in which a subset of highly conserved plasmid clusters dominate the cohort. We have added this to the manuscript: "We applied the same plasmid clustering approach to a different, publicly-available dataset of circularized *bla*$_{\text{IMP-4}}$ plasmids (n=154)… (Supplementary Figure 2)" [lines 176-181]. We have also added the corresponding figure (appended Supplementary Figure 2) to the Supplementary appendix.

R1-C3(a)

To cluster plasmids, the authors use a k-mer based approach and Jaccard similarity. This is internally consistent but creates arbitrary plasmid clusters that the authors do not attempt to place in a larger context. It is only in the Discussion that we find that PC1 is a 'hybrid of pSA20021456.2-like plasmids (GenBank accession no. CP030221) and pKPCAPSS-like plasmids (GenBank accession no. KP008371)', and PC2 was 'was none other than pNDM-ECS01 (GenBank accession no. KJ413946)'. These observations should be included in the results along with a comprehensive attempt to place the plasmids in the dataset in a broader context. This was outlined to a certain degree in Supplementary Table 6 ***[note: this has been moved to the main manuscript and renamed Table 1]*** but it would also be helpful to include other plasmid typing approaches (e.g. MOBsuite clusters, plasmid taxonomic units). The key plasmid types could also be searched for in public data using BLAST.

>> Reply: Thank you for this suggestion. We have moved what was previously known as Supplementary Table 6 to the main manuscript as Table 1. To place the plasmids in the dataset in a

broader context, we have performed BLAST on all plasmid clusters against the PLSDB database and identified the representative best hit for each plasmid cluster. Plasmid hits are listed with details such as ascension number and geographical location in Supplementary Table 7 (appended) and discussed in the results section of the manuscript: "To contextualize the plasmids in our dataset in terms of known plasmid diversity, we performed BLAST on all 48 plasmid clusters against the PLSDB database. Thirty-five plasmid genotypes matched (weighted average identity >99%) a previously described plasmid outside of our dataset (appended Supplementary Table 7), and 13 were putative novel plasmids" [lines 168-171]

We have also compared our plasmid clustering approach with other approaches such as using MOB-suite to assign MOB-suite clusters (Robertson and Nash, 2018) and COPLA to define plasmid taxonomic units (Redondo-Salvo *et al*., 2021). Our clustering approach demonstrates good agreement with MOB-suite, which also utilizes an alignment-free method. All of our plasmid clusters are assigned to a single MOB-suite cluster (with the sole exception of PC11), but MOB-suite clusters could be further subdivided into multiple closely-related plasmid clusters, suggesting that our method is higher in resolution (appended Supplementary Figure 4). COPLA was unable to classify 35.7% of our clustered plasmids into plasmid taxonomic units, suggesting previously unsampled diversity in our dataset (appended Supplementary Figure 5). We have also appended this discussion in the results section of the manuscript: "We also compared our plasmid clustering approach with other approaches such as using MOB-suite to assign MOB-suite clusters[16] and COPLA to define plasmid taxonomic units[17]… (Supplementary Figure 5)" [lines 171-180]

R1-C3(b)

The second issue with the plasmid clustering approach is that it largely ignores any structural aspects. As an example, while two plasmids may only share 70% k-mer similarity, this could be because one of the plasmids incorporated a large block of genetic material from another plasmid. However, the remainder of the plasmid remains identical to the other plasmid and is clearly related. This is reflected in the findings of the relationships between PC1/PC7/PC43 and PC2/PC20/PC49.

>> Reply:

You are right that gain (or loss) of large fragments of genetic material plasmids could lead to separate classification of plasmids that are otherwise related; recombination events are common in plasmids. These structural differences are visualised by the gene cluster comparisions between representative PC1/PC7/PC43 plasmids and PC2/PC20/PC49, in Figure 3.

In addition, we have addressed the limitations of our clustering approach more clearly in the discussion section: “There are also inherent limitations in our clustering strategy based on k-mer similarity. It may not reliably distinguish phylogenetically related plasmids, as recombination can lead to convergence or divergence in gene content that causes unrelated plasmids to cluster together or closely related ones to appear distinct. The method also does not account for gene order or synteny, which can offer important clues about shared plasmid backbones. Additionally, the similarity threshold used for clustering, while empirically selected, is ultimately arbitrary. It serves to provide a useful descriptive landscape of carbapenemase plasmid diversity but may still imply relationships that do not reflect true evolutionary history” [lines 358 to 365].

R1-C4

The dataset seems to be somewhat underutilised. While it is reasonable to focus on the two most prevalent plasmid clusters, it would be helpful to at least make some key comments regarding the other plasmid types. Similarly, the dataset contains five years of data and it would be helpful to get a sense of temporal trends. PC1 and PC2 appear to be endemic, but are there outbreaks occurring of some of their ‘subvariants’ e.g. PC7 or other plasmid types. Was there any signal for other plasmids potentially moving towards endemicity?

>> Reply:

Thank you for your suggestions. To provide a better sense of temporal trends particularly for the other plasmid types, we have added similar graphs tracking plasmid-mediated and clonal transmission over the study period for isolates carrying plasmids from the next two largest clusters, *bla*$_{\text{OXA-181}}$-positive PC3 (n=37 isolates) and *bla*$_{\text{OXA-48}}$-positive PC4 (n=31 isolates), as well as from PC7 (n=25 isolates), the next-largest *bla*$_{\text{KPC-2}}$-positive plasmid cluster after PC1, and PC5 (n=29 isolates), the next-largest *bla*$_{\text{NDM-1}}$-positive plasmid cluster after PC2 (appended Supplementary Figure 6). As with PC1 and PC2, clonal outbreaks were similarly relatively short-lived and spread across fewer institutions compared to plasmid-mediated transmission of carbapenemase genes.

Additionally, we have added a plot of the number of plasmids (y-axis) over time (x-axis) for PCs accounting for >1% of all plasmids (i.e. PC1 to PC12) to the supplementary appendix (appended Supplementary Figure 7). Apart from PC1 and PC2, none of the other plasmids showed signs of potentially moving towards hyperendemicity.

We have added this discussion to the results section in the manuscript: “Similar trends were observed for isolates carrying plasmids from the next two largest clusters, *bla*$_{\text{OXA-181}}$-positive PC3 (n=37) and *bla*$_{\text{OXA-48}}$-positive PC4 (n=31), as well as from PC7 (n=25), the next-largest *bla*$_{\text{KPC-2}}$-positive plasmid

cluster after PC1, and PC5 (n=29), the next-largest *bla*$_{\text{NDM-1}}$-positive plasmid cluster after PC2, in which clonal lineage-dependent vertical transmission limited carbapenemase gene spread across fewer institutions compared to plasmid-mediated horizontal transmission of carbapenemase genes (Supplementary Figure 6)" [lines 206-211] and "Our data suggest that plasmids are important drivers in the mobilization of carbapenemase genes between species and between geographic niches (different institutions), and also in the persistence of carbapenemase genes over time in the population. Based on the temporal trends of plasmid clusters that each accounted for >1% of all plasmids (n=12), many plasmid genotypes are stably maintained in the population for years, although apart from PC1 and PC2, none of the other plasmids showed signs of potentially moving towards hyperendemicity within our period of study (Supplementary Figure 7)" [lines 214-220]

R1-C5

The figures are at times very difficult to interpret. Please see specific comments below but this relates to the need for clearer visualisation, as well as better explanation in the legends.

>> Reply:

Thank you for highlighting this, we will do our best to address your specific comments below.

Minor comments:

R1-C6

Line 114: Is it possible to comment on the differing epidemiology between study sites and what the potential drivers of the differences are?

>> Reply:

The study sites were multi-disciplinary acute-care public hospitals, providing approximately 80% of inpatient medical care in Singapore. I have added the following to the manuscript: "Differences in epidemiology between study sites could potentially be due to varying capacity per site, ranging from 300 to 1600 beds. Two hospitals were academic medical centers with solid organ and stem cell transplant units and four were teaching hospitals with academic affiliations; these factors, in addition to location of the study site (e.g. proximity to mature housing estates, city centre or industrial areas) could have some influence on inpatient demographics and case mix." [lines 111-116]

R1-C7

Line 125: While there are more details contained in the methods, it would be helpful to contain a brief description of the clustering method that was adopted.

>> Reply:

As advised, we have added more details on the plasmid clustering method to the manuscript in the results section: "Carbapenemase-encoding plasmids were clustered based on pairwise k-mer (21bp) similarity. To assign plasmids to clusters, we built an undirected similarity network in R with igraph (v1.6.0): each plasmid was represented as a node and an edge was drawn between any two plasmids whose 21-mer Jaccard similarity was ≥ 0.90. Clusters correspond to the connected components of this network (single-linkage grouping)" [lines 136-140]

We have also appended the additional information to the methods section [lines 412-418].

R1-C8

Line 127. As per Comment 3 above, the Excel spreadsheet as supplementary material would benefit from further characterisation of each of the plasmids, including rep types, MOBsuite clusters, PTUs etc.

>> Reply:

We have done so as advised, please refer to my reply to your above comment, **R1-C3(a)**. We agree that adding these analyses enhances the report by providing more context to the work presented.

R1-C9

Line 136-147: This section is very descriptive and repetitive. Is there any possibility of capturing these data in a table or figure and using text to state salient points eg. regarding diversity of K. pneumoniae isolates.

>> Reply:

We have moved the data to the supplementary appendix as (appended) Supplementary Table 6, and only report salient points highlighting the wide diversity of host species for PC1 and PC2 in text, as follows: "PC1 was distributed across 10 species, predominantly *K. pneumoniae* (42.9%, n=167) and *E. coli* (33.2%, n=129). Among *K. pneumoniae* isolates carrying PC1, there were 60 assigned STs, most commonly ST231 (6.7%, n=11). Among *E. coli* isolates carrying PC1, there were 57 assigned STs, the most common being ST131 (11.9%, n=14). PC2 was also distributed across 10 species,

predominantly *E. coli* (38.7%, n=110) and *K. pneumoniae* (27.5%, n=78). Among *E. coli* isolates carrying PC2, there were 31 known STs, with ST131 (43.2%, n=45) being the most common. Among *K. pneumoniae* isolates carrying PC2, there were 28 known STs, the most common being ST34 (15.8%, n=12) (Supplementary Table 6).” [lines 160-167]

R1-C10

Line 154: It would be helpful to briefly state how it was that transmission was ‘genomically determined’ to be plasmid or clonal. While a reference is provided, it is difficult for the reader to then have to search for this in Reference 11. In Reference 11, it appears that PlasmidSeeker was used but I’m wondering whether perhaps the k-mer plasmid clustering approach was used in the current paper? Please clarify.

>> Reply:

Yes, the k-mer plasmid clustering approach was used in the current paper. We apologize for being unclear. In addition, while revising our manuscipt, we have made other changes to provide more clarity on the methods used to define “clonal”:

In the results subsection “Transmission dynamics of predominant carbapenemase-encoding plasmids”, we overtly state our aim to investigate the relative impact of vertical versus horizontal transmission on the dissemination of carbapenemase-encoding plasmids. For this purpose, we explain that “a carbapenemase-encoding plasmid was considered vertically acquired by an isolate if the host met pairwise clonal linkage criteria with an earlier isolate. Briefly, a pair of isolates were determined to be clonally linked if they shared the same ST-cluster, same carbapenemase gene allele and had a pairwise single-nucleotide polymorphism (SNP) count (based on the recombination-filtered core gene alignments) below the BEAST-derived mutation rate threshold, assuming a Poisson distribution for the accumulation of mutations[11]. A carbapenemase-encoding plasmid was considered horizontally acquired by an isolate if the host was not clonally linked to another isolate” [lines 186-193]. We subsequently report the relative proportion of isolates that putatively acquired the plasmid via horizontal transmission versus clonal lineage-dependent vertical transmission. More information on BEAST-derived mutation rates and SNP thresholds are also added in the Methods [lines 450-457].

In the results subsection “Evolution of plasmids within the same clonal transmission cluster and within the same patient”, we aim to examine plasmid evolution along the same clonal lineage between-patients and across multiple species within the same patient. For this purpose, we explain that “a clonal transmission cluster comprised index isolates that met pairwise clonal linkage criteria with at least one earlier index isolate from another patient. An index isolate was defined as the first-

detected isolate carrying a unique carbapenemase gene in a patient during the study period. More than one index isolate (carrying different carbapenemase genes) could be associated with a patient if they shared the same date of culture. Only index isolates were considered for construction of clonal transmission clusters to ensure that clusters reflect between-patient transmission events" [lines 288-294]

R1-C11

Line 167: Were there other isolates with multiple carbapenemase genes? If so, can these findings be summarised briefly?

>> Reply:

The 1,115 closed carbapenemase-producing plasmids in this dataset were identified from 1,088 isolates. We have added the following to the results section in the manuscript: "Of the 1,088 isolates, 20 isolates carried two different carbapenemase genes on separate plasmids, and 2 isolates carried two different carbapenemase genes on the same plasmid. Details on carbapenemase gene co-carriage can be found in the Supplementary Excel File" [lines 100-103]

The Supplementary Excel file is also provided.

R1-C12

Line 182: Why are the authors comparing the core genomes of plasmids that are already known to be clearly different (i.e. carriage of different carbapenemase genes and already characterised as different by the clustering approach). It is not clear what the message from this analysis is.

>> Reply:

To clarify, in Figure 3A (revised manuscript), the core genome of *bla*$_{KPC-2}$-positive PC1 is compared to all other *bla*$_{KPC-2}$-positive IncN plasmids. Likewise in Figure 3B (revised manuscript), the core genome of *bla*$_{NDM-1}$-positive PC1 is compared to all other *bla*$_{KPC-2}$-positive plasmids. The aim of these figures is to highlight the most closely related plasmids to PC1 and PC2.

We have rearranged the subfigures in Figures 3 and 4 as well as revised the manuscript accordingly to improve clarity and flow. Related to revisions in Figure 3:

"PC1 and PC2 were considered evolutionarily successful plasmids due to their high prevalence, with each accounting for more than 25% of plasmids in the cohort (Table 1). The PC1 core genome of 97 genes that were present in 95% of PC1 plasmids was found to be highly conserved (>90%) in PC7

and PC43 compared to other *bla*$_{KPC-2}$-positive IncU plasmid genotypes (Figure 3A). The PC2 core genome of 52 genes that were present in 95% of PC2 plasmids was found to be most highly conserved (>90%) in PC20 and PC49 compared to other *bla*$_{NDM-1}$-positive IncN genotypes (Figure 3B). Gene cluster organisation of representative PC1 (Figure 3C) and PC2 (Figure 3D) plasmids, visualized by Clinker, emphasizes the structural conservation of a distinct backbone of core gene loci, and reveals divergence from the PC1 and PC2 genetic settings through insertion events." [lines 223-232]

Related to revisions in Figure 4:
"To investigate if there were any specific genes that could be associated with the likelihood or failure to achieve hyperendemicity, we compared the full gene complement of PC1 (34.9% of all plasmids) with that of plasmids with high core genome similarity but significantly lower prevalence (Supplementary Figure 3), such as PC7 (2.2% of all plasmids) and PC43 (0.2% of all plasmids) (Figure 4A). Likewise, we compared the full gene complement of PC2 (25.5% of all plasmids) with that of PC20 (0.4% of all plasmids) and PC49 (0.2% of all plasmids) (Figure 4B). It could be ruled out that the less prevalent plasmids were poorly represented in the cohort as a result of late emergence towards the end of the sampling time frame (Supplementary Figure 8)." [lines 246-254]

We hope the edits more clearly reflect the purpose of Figures 3 and 4, which show the structural conservation of a distinct backbone of core gene loci in predominant PC1 and PC2 plasmids, and which additional accessory genes are present in less prevalent plasmids.

R1-C13
Line 223: How are 'clonal clusters' defined?

>> Reply:
We apologise for not providing a clear definition. Per your related comment **R1-C10** above, we have made several changes to provide more clarity on the methods used to define "clonal":

In the results subsection "Transmission dynamics of predominant carbapenemase-encoding plasmids", we overtly state our aim to investigate the relative impact of vertical versus horizontal transmission on the dissemination of carbapenemase-encoding plasmids. For this purpose, we explain that "a carbapenemase-encoding plasmid was considered vertically acquired by an isolate if the host met pairwise clonal linkage criteria with an earlier isolate. Briefly, a pair of isolates were determined to be clonally linked if they shared the same ST-cluster, same carbapenemase gene allele and had a pairwise single-nucleotide polymorphism (SNP) count (based on the recombination-filtered core gene alignments) below the BEAST-derived mutation rate threshold, assuming a Poisson distribution for

the accumulation of mutations[11]. A carbapenemase-encoding plasmid was considered horizontally acquired by an isolate if the host was not clonally linked to another isolate" [lines 186-193]. We subsequently report the relative proportion of isolates that putatively acquired the plasmid via horizontal transmission versus clonal lineage-dependent vertical transmission. More information on BEAST-derived mutation rates and SNP thresholds are also added in the Methods [lines 450-457].

In the results subsection "Evolution of plasmids within the same clonal transmission cluster and within the same patient", we aim to examine plasmid evolution along the same clonal lineage between-patients and across multiple species within the same patient. For this purpose, we explain that "a clonal transmission cluster comprised index isolates that met pairwise clonal linkage criteria with at least one earlier index isolate from another patient. An index isolate was defined as the first-detected isolate carrying a unique carbapenemase gene in a patient during the study period. More than one index isolate (carrying different carbapenemase genes) could be associated with a patient if they shared the same date of culture. Only index isolates were considered for construction of clonal transmission clusters to ensure that clusters reflect between-patient transmission events" [lines 288-294]

R1-C14

Line 287: It would be helpful to discuss some of the other limitations of the work, e.g. methodological issues of the clustering approach, the absence of conjugation work etc.

>> Reply:

We have further discussed the limitations of the work as follows:

"There are also inherent limitations in our clustering strategy based on k-mer similarity. It may not reliably distinguish phylogenetically related plasmids, as recombination can lead to convergence or divergence in gene content that causes unrelated plasmids to cluster together or closely related ones to appear distinct. The method also does not account for gene order or synteny, which can offer important clues about shared plasmid backbones. Additionally, the similarity threshold used for clustering, while empirically selected, is ultimately arbitrary. It serves to provide a useful descriptive landscape of carbapenemase plasmid diversity but may still imply relationships that do not reflect true evolutionary history. We did not conduct *in vitro* conjugation assays for plasmids other than PC1, which we had previously demonstrated to impose low fitness costs, have high conjugation frequencies and high retention rates in multiple Enterobacterales species" [lines 358-368]

R1-C15

Line 301: Some further details regarding how isolates were collected would be helpful. What proportion were clinical isolates vs 'surveillance' isolates. How is a 'surveillance' isolate defined? Were these rectal swabs done to detect colonisation?

>> Reply:

We apologise for not being clear about this. We have added the definition to the manuscript: "Surveillance cultures to detect gut colonization were from rectal swab or stool samples taken from asymptomatic carriers identified via screening of high-risk patients and epidemiologically linked contacts." [lines 380-382]

R1-C16

Line 316: Some additional details regarding the quality parameters used to accept/reject assemblies are needed.

>> Reply:

We have added the following to the manuscript: "Hybrid assemblies that contained closed carbapenemase gene contigs less than 500kb (assumed to be non-chromosomal) that contained at least one replicon were included in the final analysis." [lines 398-400]

R1-C17

- Line 330: Was there a specific algorithm used for clustering (e.g. hierarchical clustering)? If so, can this be specified including whether there was a specific software tool used to do this?

>> Reply:

We have added more details in the results and methods section of the manuscript: "To assign plasmids to clusters, we built an undirected similarity network in R with igraph (v1.6.0): each plasmid was represented as a node and an edge was drawn between any two plasmids whose 21-mer Jaccard similarity was ≥ 0.90. Clusters correspond to the connected components of this network (single-linkage grouping)." [lines 137-140 and 414-418]

R1-C18

Figure 1C: this is difficult to interpret in its current form. Perhaps a network graph may be more useful with edges drawn between members of the same cluster?

>> Reply:

We have moved key information on the plasmid clusters from the Supplementary appendix to the main manuscript as Table 1, which we hope will complement Figure 1C. The purpose of the heatmap in Figure 1C is to visualise the relative size of the plasmid clusters and inter-cluster partitioning. We have also expanded the dendrogram slightly to make it more visible, and made some edits to the figure legend. We hope these changes help with interpreting Figure 1C and welcome further suggestions.

We have attempted to construct network graphs using Cytoscape but the network graphs were not more informative that the current heatmap, in terms of visualising the inter-cluster relatedness beyond a fixed similarity threshold.

R1-C19

Figure 2: The message in this figure is not entirely clear. What does the y-axis represent? Is each tick a separate plasmid cluster? Is the intention to show that clonal transmission only contributed a small portion to the total number of plasmids transmitted? It is very difficult to interpret the species and hospitals involved in the current figure. Please consider representing this in a different way.

>> Reply:

In Figure 2, each point represents the incidence of a **(A)** *bla*$_{\text{KPC-2}}$-positive PC1 plasmid or **(B)** *bla*$_{\text{NDM-1}}$-positive PC2 plasmid. The points are ordered along the x-axis based on the date of detection, with displacement along the y-axis for clarity. Plasmids that occurred in isolates that met pairwise clonal linkage criteria are visually connected by pink horizontal lines. Species and hospitals associated with the plasmids are listed. PC1 plasmids were detected across 10 species and five different hospitals, while PC2 plasmids were detected across 10 species and all six study sites. For clonal lineage, the following information is annotated in this format: "abbreviated species; ST; number of isolates; number of hospitals where the clone was detected". We decided to remove the y-axis label as it may be misleading, and we have provided more information in the figure legend in lines 664 to 672 for clarity. More details on the clonal clusters are also provided in Supplementary Tables 8 and 9, which are referenced in the results section as well as the figure legend.

The figure is intended to show that the widespread dissemination of *bla*$_{\text{KPC-2}}$ and *bla*$_{\text{NDM-1}}$ across multiple species and institutions was primarily driven by horizontal gene transfer mediated by PC1 and PC2 plasmids respectively. In contrast, clonal lineage-dependent vertical transmission contributed

not only a smaller portion of the total number of plasmids transmitted but was unable to account for the spread of PC1 and PC2 across multiple hospitals.

R1-C20

Figure 4: This figure appears to show that PC7 and PC20 are subvariants of PC1 and PC2, respectively, with additional genetic material. Would consider making this a supplemental figure.

>> Reply:

As advised, we have moved PC7 and PC20 of this figure to the Supplementary appendix as Supplementary Figure 9 (appended). We have also reorganised the other elements of Figure 4 (and Figure 3) and revised the manuscript accordingly.

R1-C21

Figure 5A and B. Is it possible to use two more differentiated palettes for the two panels? They currently seem very similar, leading to confusion between PC1 vs PC2 for example. These two panels are also very difficult to interpret as the labelling of the axes is not clear. How is time represented? What are each of the clonal clusters? Are they different species/ST combinations? Is the intention to show that most clonal clusters carry only one plasmid cluster?

>> Reply:

Thank you for the constructive feedback. For clarity, we have changed the color palette of the panel in Figure 5A to differentiate if from Figure 5B. We have also further annotated the figures with the number of isolates per clonal cluster as well as number of days between the earliest and latest isolate detected, and revised the figure legend as follows: “Plasmid profiles of **(A)** *bla*$_{\text{KPC-2}}$-positive and **(B)** *bla*$_{\text{NDM-1}}$-positive clonal transmission clusters. Clonal transmission clusters comprising three or more isolates are shown. Each column represents a clonal transmission cluster comprising index isolates that met pairwise clonal linkage criteria with at least one earlier index isolate from another patient. An index isolate was defined as the first-detected isolate carrying a unique carbapenemase gene in a patient during the study period. For each cluster, carbapenemase plasmid genes are arranged along the y-axis and isolates are arranged along the x-axis according to date of culture, from earliest to latest (left to right). Heat map indicates gene presence (colored) or absence (greyed). The number of isolates per clonal transmission cluster as well as number of days between the earliest and latest isolate detected are shown” [lines 697-706]

R1-C22

Figure 5C: The patient plots are very difficult to follow. Is there any way that the visualisation can be simplified?

>> Reply:

Figure 5C visualizes the host range and persistence of plasmids found in 22 patients with five or more isolates. We have added more information to the figure legend and manuscript as follows: “Twenty-two individuals (of 705 patients) were determined to be the source of five or more isolates. In 20 of 22 patients, the same plasmid was found in two or more different species; in five of 22 patients, the same plasmid was found in three or more different species. The median number of species in which PC1 and PC2 were found were 2 (IQR, 2–2 and IQR, 1–2, respectively). For remaining PCs, the median number of host species was 1 (IQR 1–1)” [lines 312-316].

Reviewer 2

"Plasmid dynamics driving carbapenemase gene dissemination in healthcare environments: a nationwide analysis of closed Enterobacterales genomes" by Koh et al. is an interesting manuscript about an extensive study of sequenced CPE isolates from Singapore. In a previous study, the complete genomes of more than 1000 isolates were compared with Illumina short-read sequencing. In the current study they describe the use of Nanopore long-read sequencing in order to analyse the complete carbapenemase plasmids of these isolates. The paper is well written and give a thorough analysis of the data.

R2-C1

Overall, the figures are small and contain too many panels. Specifically for figure 3 and 5 this needs to be addressed. It is a shame because the amount of data in these figures is great.

>> Reply:

Thank you for the constructive feedback. We have reorganised Figures 3 and 4, and added more information to Figure 5 for clarity. We have also revised the figure legends accordingly.

R2-C2

Line 30: How did you asses if genes were acquired via clonal transmission?

>> Reply:

We apologize for being unclear. We have edited the manuscript to provide more clarity on the methods used to define "clonal":

In the results subsection "Transmission dynamics of predominant carbapenemase-encoding plasmids", we overtly state our aim to investigate the relative impact of vertical versus horizontal transmission on the dissemination of carbapenemase-encoding plasmids. For this purpose, we explain that "a carbapenemase-encoding plasmid was considered vertically acquired by an isolate if the host met pairwise clonal linkage criteria with an earlier isolate. Briefly, a pair of isolates were determined to be clonally linked if they shared the same ST-cluster, same carbapenemase gene allele and had a pairwise single-nucleotide polymorphism (SNP) count (based on the recombination-filtered core gene alignments) below the BEAST-derived mutation rate threshold, assuming a Poisson distribution for the accumulation of mutations[11]. A carbapenemase-encoding plasmid was considered horizontally acquired by an isolate if the host was not clonally linked to another isolate" [lines 186-193]. We

subsequently report the relative proportion of isolates that putatively acquired the plasmid via horizontal transmission versus clonal lineage-dependent vertical transmission. More information on BEAST-derived mutation rates and SNP thresholds are also added in the Methods [lines 450-457].

In the results subsection "Evolution of plasmids within the same clonal transmission cluster and within the same patient", we aim to examine plasmid evolution along the same clonal lineage between-patients and across multiple species within the same patient. For this purpose, we explain that "a clonal transmission cluster comprised index isolates that met pairwise clonal linkage criteria with at least one earlier index isolate from another patient. An index isolate was defined as the first-detected isolate carrying a unique carbapenemase gene in a patient during the study period. More than one index isolate (carrying different carbapenemase genes) could be associated with a patient if they shared the same date of culture. Only index isolates were considered for construction of clonal transmission clusters to ensure that clusters reflect between-patient transmission events" [lines 288-294]

R2-C3

Line 45 and 283: What do you mean by 'maintenance of trim genomes'. I understand that these genomes are plastic and require only a low number of core genes, but I've not come across this term before. It is not explained in the study you reference.

>> Reply:

For clarity, have replaced the term "trim" with "conserved" [lines 46, 350].

R2-C4

Line 53: Although many CRE are resistant to most beta-lactams, some enzymes such as certain OXA-48-like enzymes, GES, IMI and FRI do not confer resistance to cephalosporins. This leads to underestimation of these genes when selective isolation is performed using cephalosporins. Please amend this sentence.

>> Reply: We have amended the sentence from "CRE are resistant to all β-lactam antibiotics…" to: "CRE are resistant to most β-lactam antibiotics…" [line 54]

R2-C5

Line 68: Although I agree that plasmids cannot be assembled easily from short-read data, this data can be used more broadly than reference-based mapping, as you've shown in your previous study.

>> Reply:

We agree that our original statement was not accurate and we have amended it in the manuscript, from “short-read sequencing, while empowering genomic surveillance of high-risk bacterial clones by means of reference-based mapping, do not enable the same high-resolution tracking of plasmids due to their relative plasticity and recombinant nature”, to: “Short-read sequencing, although widely used for genomic surveillance of high-risk bacterial clones through approaches such as reference-based mapping, is suboptimal for high-resolution tracking of plasmids due to their structural plasticity and recombinant nature, which complicate accurate assembly” [lines 69-72]

R2-C6

Line 131: In the provided reference I do not see any information about the functional replicon trfA or the classification of IncPe1.

>> Reply:

Thank you for bringing this to our attention. We had cited the wrong paper (from the same group) and we have corrected the reference [line 150]

R2-C7

Line 241-244: Please expand this section. The finding of highly similar plasmids in the same patient in diverse species warrants additional details.

>> Reply:

We have added the following to the section: “Twenty-two individuals (of 705 patients) were determined to be the source of five or more isolates. In 20 of 22 patients, the same plasmid was found in two or more different species; in five of 22 patients, the same plasmid was found in three or more different species. The median number of species in which PC1 and PC2 were found were 2 (IQR, 2–2 and IQR, 1–2, respectively). For remaining PCs, the median number of host species was 1 (IQR 1–1). Our data suggests that in the context of a singular human host, $bla_{\text{KPC-2}}$ and $bla_{\text{NDM-1}}$ are more stably maintained over time and across multiple species (and STs) in PC1 and PC2 plasmids, respectively, compared to other genetic settings (Figure 5C).” [lines 312-319]

R2-C8

Line 312: Please include the full specifics on the Nanopore sequencing. Which sequence kit did you use. How was the basecalling performed. I am surprised that so many of your IncF multireplicon plasmids were fully assembled into circular contigs, for us these are difficult to get fully assembled without manual curation. Please include data on the depth of sequencing (highest, lowest, average).

>> Reply:

We have added the following additional details to the manuscript: “Libraries for long-read sequencing were prepared for the same isolates using the ONT Rapid Barcoding Kit (RBK004) and sequenced on R9.4.1 flowcells on a GridION platform (Oxford Nanopore Technologies). Basecalling was performed using Guppy version 3.2.6 (https://nanoporetech.com/software/other/guppy) which was integrated with the MinKNOW software, and demultiplexing was performed using qcat (https://github.com/nanoporetech/qcat). The average Nanopore data throughput was 1.2Gb per isolate (range: 0.002-5.4 Gb)” [lines 391-397]

R2-C9

Line 456: please adapt the author list to include the name of the first author..

>> Reply:

Thank you for bringing this to our attention, we have edited the reference to properly credit the first author of the cited maunscript [line 552]

R2-C10

Line 579: Please correct carbapenemase.

>> Reply:

We have corrected “carbapenmase” to “carbapenemase” [line 702]