

Preferred evolutionary routes of convergence in *Klebsiella pneumoniae* favor siderophore acquisition over hypervirulence

Corresponding Author: Dr Francois Lebreton

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)

Preferred evolutionary routes of convergence in *Klebsiella pneumoniae* favor siderophore acquisition over hypervirulence

In this study 1,468 isolates and 7,520 plasmids were analyzed. This reviewer agrees with the authors that the dominant route among the “analyzed plasmids” involved multidrug-resistant classical *K. pneumoniae* acquiring conjugative IncFIB(Mar)/IncHI1B (Mar) plasmids carrying an incomplete set of virulence biomarkers and that the acquisition of the aerobactin siderophore locus was the common feature in this collection and that isolates consistently lack the hypervirulent phenotype in vivo.

However, this reviewer suggests more specificity. It is not acceptable that in a study of this relevance the plasmids are named using the list of replicons (reported incorrectly sometimes) and that there are not linear or circular maps that can help the readers to understand if they have the same plasmid or not in their hands.

Major points

1-PlasmidFinder contains 3 IncHI1B replicons.

IncHI1B(R27) is one of the 3 replicons that with IncHI1A and IncFIA(HI1) identifies the reference plasmid R27 (AF250878) prototype of all the IncHI1 plasmids that have 3 replicons HI1A-HI1B and FIA(specific of the HI1).

IncHI1B(pNDM-CIT) and IncHI1A(NDM-CIT) identify the HI1-like plasmids prototype pNDM-CIT (JX182975, homology <94% with the others HI1)

IncFIB(Mar)/IncHI1B(Mar) identify the plasmids relative to the first described with these replicons called pNDM-Mar (JN420336, homology <94% with the others HI1). HI1B(MAR) but not FIB (Mar) was also identified in the hvKp virulence plasmid pLVPK, 219385 bp, AY378100. It was not associated with FIB(Mar) but with repB_KLEB_VIR. pLVPK was not conjugative. The acquisition of the FIB(Mar) replicon is coincident with the acquisition of a transfer locus and blaCTX-M-15 (Chen YT et al. Gene 337, 189-198 (2004); Arcari, G., & Carattoli, A. (2023).

<https://doi.org/10.1080/20477724.2022.2121362>).

considering the complexity of the HI1 plasmids and their replicon type B, it is mandatory that the nomenclature IncFIB(K)/IncHI1B is changed with IncFIB(Mar)/IncHI1B(Mar) or FIB-HI1B-Mar

2-virulence markers are not correctly defined. rmpA2 is out of frame in most the FIB-HI1B-Mar plasmids. This impacts on the virulence but has not been clearly stated in the paper. Biomarkers are searched on genomes must it should be carefully checked if the database also indicates partial coverage of the product and potential pseudogenes. The concept of virulence may vary for mutations in the rmpA and rmpA2 genes but also because of mutations in wzi (Ernst et al., Nat Med. 2020 May;26(5):705-711. doi: 10.1038/s41591-020-0825-4.)

Line 295: pC-2 (IncFIB(Mar)/IncHI1B(Mar) hybrid plasmid carrying 4 virulence biomarkers, including iuc1, and the NDM-1 carbapenemase. RmpA2 is out of frame?

3-Most of the collection is in contigs and only 53 are complete plasmid sequences. The general definition of the pC groups is

done on replicons and biomarkers. Is it possible that resistance or biomarkers are within the same bacterial gene content but are not physically located on one single plasmid scaffold? Please check and present a single accession no. and/or a map of the reference for each of the most relevant PC type. This avoids mistakes, example at line 104:

"IncFIB(Mar)/IncHI1B dual replicon type carrying resistance (blaNDM-1, armA) and hvKp-specific virulence (iuc, rmpADC, rmpA2, and/or peg-344) biomarkers that are found in globally epidemic nosocomial MDR lineages such as ST-307 and ST-147 (26,27)".

NOT TRUE for ST147: blaNDM was not on the FIB-HIB(Mar) plasmid but on a separate FIB(Qil) plasmid. From reference 26: While the blaNDM-1 gene was carried by an IncFIB-type plasmid, our investigations revealed that the ST-147 lineage from Italy also acquired a hybrid IncFIB/IncHI1B-type plasmid carrying the 16S methyltransferase armA gene as well as key virulence biomarkers often found in hypervirulent isolates.

Is it true that the pC10 has incL-oxa-48 and armA on the same scaffold? Is it IncM?

Reviewer #2

(Remarks to the Author)

This study by Sapre, Martin, and colleagues presents a comprehensive genomic and phenotypic analysis of the convergence of antimicrobial resistance and virulence in *Klebsiella pneumoniae*. By analyzing a large, global collection of isolates and performing deep plasmidomics, the authors demonstrate that convergence predominantly occurs through MDR-cKp acquiring plasmids that carry an incomplete set of virulence genes, with the aerobactin locus being the universal common denominator. Crucially, they provide in vivo evidence that these convergent isolates exhibit enhanced siderophore production but do not display the hypervirulent phenotype, challenging the assumption that genomic convergence equals phenotypic hypervirulence. The work is of high significance, however, there are some concerns of this study.

Major concerns:

1. The authors state that pC2 plasmids carry an incomplete set of virulence biomarkers and are more frequently associated with the carriage of only iuc and rmpA2. However, the manuscript lacks analysis of the evolutionary relationship between pC2 and pC4 plasmids. This is a critical gap in understanding the origin of these successful convergent vectors. Do pC2 plasmids represent degenerated variants of the canonical pC4 (pLVPK-like) virulence plasmids that have lost iro, peg-344, and other markers over time? Alternatively, did they arise through independent acquisition events where the iuc and rmpA2 loci were incorporated into a non-virulence IncFIB(Mar)/IncHI1B plasmid backbone? These two scenarios have fundamentally different implications for understanding the evolution of convergence.
2. The authors repeatedly emphasize the "conjugative" nature of pC2 plasmids and the "non-conjugative" nature of pC4 plasmids, using this to explain their differential distribution. However, direct experimental evidence for these claims is lacking.
3. The authors argue that pC2 plasmids are more successful in MDR-cKp backgrounds due to their "conjugative nature" and "lack of hypervirulence phenotype" compared to pC4 plasmids. However, critical factors include plasmid stability, fitness cost imposed on the host, host range, and compatibility with resident plasmids all influence plasmid success. The authors do not provide data on any of these parameters.
4. 53 isolates were selected for long-read sequencing and complete plasmid resolution. The criteria for selection are not clearly defined. If isolates were selected based on specific genetic features without a systematic sampling strategy, the plasmid diversity captured may not be representative of the broader population.
5. The study conclusively shows that enhanced siderophore production, driven by iuc acquisition, does not confer systemic hypervirulence in the murine model. However, existing literature suggest that iron acquisition provides ecological advantages beyond systemic invasion. The discussion of these trade-offs is currently underdeveloped.

Other specific comments:

1. Line 121, "The frequency of isolate collection varied by year and country." Please provide a brief explanation for this variation. More importantly, discuss the potential consequences of this sampling bias on the conclusions.
2. Line 124, Please clarify whether the collection includes any duplicate isolates. Additionally, provide more detail on the strain isolation methods and inclusion criteria in the Materials and Methods section.
3. Line 138, "This largely correlated with the sampling of clonal, geographically clustered isolates." Please state this correlation more specifically.
4. Line 180, Please define "MRSN"
5. Line 205-206, The figure legend appears to be incorrectly labeled. Please verify that the colors and labels in panels C and D correspond correctly to virulence and resistance gene carriage as described in the text.
6. Please maintain consistent formatting with plasmid cluster designations, for example PC2 and PC-2.
7. Line 259, Should be an enrichment...
8. Line 313, Should be pC33.
9. Line 329, The legend appears to have an error.
10. Line 444-445. "clast allele assignment" appears to be a typo. Please specify the exact procedure used for cgMLST allele assignment and minimum spanning tree generation.

Reviewer #3

(Remarks to the Author)

The manuscript by Sapre et al. investigates the evolutionary pathways and phenotypic characteristics of *Klebsiella pneumoniae* strains that acquire both antimicrobial resistance and virulence traits, highlighting the predominance of siderophore production over classical hypervirulence in convergent lineages. The study is timely and relevant, particularly given the emergence of multidrug-resistant strains carrying hypervirulent traits, and it complements prior work on this topic (PMID: 41512896). While the study is interesting and the findings are valuable, several aspects of the experimental design and the interpretation of results require clarification before the conclusions can be confidently generalized. Some comments are provided for the authors' consideration.

Major comments

1. Sampling. The study relies on a highly specific set of 1458 isolates, predominantly from LMICs and a limited number from the United States ($n = 39$), seemingly drawn from a single context (MRSN). The criteria for this sampling strategy should be clearly explained and justified, as the broad conclusions risk overgeneralization. Heterogeneous temporal coverage, geographic bias, host population age/gender, and the potential inclusion of outbreak-associated or otherwise non-representative isolates could influence the observed distribution of clonal lineages. Clarification of how and why isolates were selected, their temporal and geographic distribution, individual epidemiological characteristics, and whether they reflect routine surveillance or targeted collections is essential to assess the robustness and generalizability of the findings.

2. Use of "convergence" terminology. The use of the term "convergence" in this work is challenging and may cause confusion for different readers throughout the text. Convergent evolution is traditionally defined as the process by which different species independently evolve similar traits. When similar outcomes occur within populations of the same species, the term generally used is parallel evolution. In this study, the focus is on the acquisition of both antibiotic resistance and virulence traits. While the co-carriage of these traits could be described as phenotypic convergence, other adaptive traits located in the plasmids analyzed have not been examined and might also contribute to adaptation (a limitation that should be mentioned). Also, it is confusing to describe isolates or lineages themselves as "convergent" (lines 67, 70, 101, 156, 160, 177, and elsewhere), because convergence refers to evolutionary processes or traits, not to individual isolates or lineages. Mechanistically, plasmid acquisition is primarily classified as horizontal gene transfer (HGT) rather than genomic convergence. Nonetheless, the description of the outcome at the population or phenotype level as convergent adaptation is frequent in microbial evolutionary genomics (see ref #38 for discussion about "convergent *Klebsiella pneumoniae*"), particularly relevant when distinct alleles of the same antibiotic resistance genes (e.g., NDM, CTX-M) or virulence genes (e.g., iuc) are involved as they may differ at the sequence level but produce the same adaptive phenotype. However, the process described here and other papers may be more accurately described as "convergent adaptation via HGT", "parallel acquisition of adaptive genes via plasmids", or "mobile gene pool-mediated adaptation". Strict genomic convergence requires independent mutations or events that evolve separately in different populations (e.g., gyrA mutations conferring quinolone resistance). In contrast, in this study, genes are imported from the mobilome rather than evolving de novo, so evolution occurs through the mobile genetic element pool rather than independently in each lineage. See previous comment about the sample.

3. Convergence events. The treatment of convergence events in the manuscript requires a more detailed description and a precise definition. From the text, it appears that the authors consider the five putative virulence genes previously linked to hypervirulent *K. pneumoniae* as a "unit of selection," noting that this set can be "incomplete" in some isolates. However, the coexistence of these disparate virulence determinants on a single plasmid is likely exceptional and primarily reflects the selection and transmission of the plasmid as a unit. Other scenarios are also possible, as the authors mention, including the acquisition of these genes individually or in different combinations by different plasmids. Therefore, it is important to clarify whether convergence events are being defined at the gene level or at the plasmid level (see how convergence is approached in reference 38). For antibiotic resistance genes, allelic variants of a single gene (e.g., bla_{NDM}) are treated as distinct genes. Yet some allelic variants may share a single evolutionary origin, rather than representing independent acquisition events, and conversely, the same allele can be repeatedly acquired by different lineages via widespread plasmids. Clarifying these points is essential to accurately defining what constitutes a convergence event in this dataset.

4. Novelty and context. The manuscript could better situate its findings within existing literature. Previous studies have documented convergence of virulence and resistance traits in *K. pneumoniae* via plasmid hybridization (PMID: 41512896, reference 38). The present work provides complementary insights, showing that less than 5% of plasmids were hybrids, demonstrating the relevance of acquisition by independent plasmids rather than hybridization events. However, reference 38 is mentioned only once, and a more thorough comparison would strengthen the discussion of novelty and relevance. Additionally, isolates from wounded individuals in the Ukrainian conflict have been reported, with results similar to those reported here (PMID: 41177264), which could be relevant for contextualizing lineage diversity and sampling, tying back to the first comment on sampling.

Other comments.

Line 180, define the acronym MRSN and indicate the number of plasmid sequences involved.

Line 348 – Recurring and polyphyletic nature of convergence: Consider clarifying what is meant by "recurring" and "polyphyletic" in this context to ensure readers understand how convergence is being defined.

Lines 187-199, plasmid description. The description of the plasmids here is currently unclear and could benefit from additional context. Introducing the type of each plasmid cluster (PC) at the beginning of the section would help the reader understand the rationale for the analysis. They only appear in the last sentences (e.g., pC10 as an IncL/M carrying bla_{OXA-48}). Also, it would also be helpful to clarify how these plasmids relate to previously described pOXA plasmids. Similarly, providing information on the relationships of pC38 and pC39 to other widespread plasmids would strengthen the interpretation of their significance. Finally, specify which bla_{CTX-M} alleles are being referred to, as this is critical for

assessing the functional and epidemiological relevance of these plasmids.

Lines 477–485: Analyses described here are part of the study's results and should be moved to the Results section.

General: Check spelling throughout the text (e.g., line 139) and ensure consistent nomenclature for antibiotic resistance genes and proteins. Use the acronym MDR consistently. Specify the number of isolates or plasmids corresponding to this study in the different sections for clarity.

Reviewer #4

(Remarks to the Author)

Sapre and colleagues conducted a large-scale analysis of the co-occurrence of resistance and virulence genes in *Klebsiella pneumoniae* (Kp) strains and associated plasmids. The study focused on 1,468 Kp isolates collected across nine countries between 2011 and 2021, complemented by a broader and more representative dataset of 7,520 plasmids retrieved from NCBI.

The authors identified 25 independent convergence events, with acquisition of the aerobactin siderophore system emerging as the only consistent feature across all events. Convergent strains produced higher levels of siderophores compared to the control *ckp1* strain; however, in vivo experiments did not confirm a corresponding increase in virulence. Based on these findings, the authors suggest that siderophore acquisition may contribute primarily to adaptive fitness rather than directly enhancing virulence.

In my opinion the major limits of this work are that it is focussed on a very specific Kp dataset without a rationale (or needs) and it considers resistance and virulence genes only. Furthermore, other minor modifications are needed.

i) The genomic dataset:

The Kp dataset includes mostly strains isolated from Thailand, Georgia, Peru and Jordan. Thus it is very biased for these countries, that in my knowledge can not be considered as representative for some reason. Just an example: In figure 1A, the number of Virulence Biomarkers appears to be not fully independent from the isolation country, which in turn could be associated with the local epidemiology.

The dataset is also strongly biased for what concerns isolation time, with the 2018 over-represented. Considering the dynamic nature of the Kp epidemiology, a time bias can drastically affect the results.

One of the effects of bias sampling is evident considering the very low representation of the pandemic ST258, while the less diffused ST258-derived ST512 is well represented (Figure S1). This bias is particularly important considering that the sampling period was characterised by the diffusion of the ST258.

I suggest to the authors to repeat the analysis including also Kp genomes retrieved from NCBI, paying attention to geographical/time stratification.

ii) Genes:

The authors focused the attention of virulence and resistance genes. This approach can drastically limit their ability to investigate the evolutionary trajectories. In particular, even if Kleborate is probably the most used tool for Kp typing, there are other tools/dataset for the identification of virulence genes. Unfortunately, the definition of virulence is not simple as for the antibiotic resistance, and thus a larger gene dataset could provide a more precise description of the virulence of the strains. Indeed, the authors did not find any correlation between virulence gene content and real in vivo virulence.

As part of the phylogenomics analysis flow, the authors used Roary. I suggest they analyse the orthology groups distribution among the strains to identify putative genes associated with virulent phenotypes.

Point by point suggestions:

line 136: I suggest to perform a correlation analysis between time and the rate of CP-producing isolates

Legend 2: C and D letters are inverted

line 226: in figure 2 we can see that within the cluster pC2 there are two subclusters, one including mostly virulence genes and one mostly (or better a lot of) resistance genes.

line 259: typing error

Figure S5: the image is unreadable to me (I think too low DPI)

Figure 4: I really appreciate this figure, I suggest the author make a similar figure of each phylogenetic lineage for which they have convergent routes.

Figure 5: in the legend the letters C and B are inverted. I also suggest performing a regression between [suderophore] and LD50.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This reviewer acknowledges the effort made by the authors to correct the manuscript according to the suggestions. The unique point to revise is in the replicons of Figure S5. Heatmap of plasmid diversity for isolates with at least one virulence biomarker but lacking an acquired ESBL or carbapenemase. In this figure replicons have not been revised accordingly with the current nomenclature adopted in the text.

Reviewer #2

(Remarks to the Author)

I would like to thank the authors for their comprehensive and considerate replies to the points raised in my earlier review. I am satisfied that all concerns have now been sufficiently addressed.

Reviewer #4

(Remarks to the Author)

I appreciate the revisions made by the authors. However, I found it very difficult to assess the changes because the line numbers cited in the rebuttal do not correspond to those in the revised manuscript.

Thus, I would like to ask to correct the line numbers reported in the rebutal file.

made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

Preferred evolutionary routes of convergence in *Klebsiella pneumoniae* favor siderophore acquisition over hypervirulence

In this study 1,468 isolates and 7,520 plasmids were analyzed. This reviewer agrees with the authors that the dominant route among the “analyzed plasmids” involved multidrug-resistant classical *K. pneumoniae* acquiring conjugative IncFIB(Mar)/IncHI1B (Mar) plasmids carrying an incomplete set of virulence biomarkers and that the acquisition of the aerobactin siderophore locus was the common feature in this collection and that isolates consistently lack the hypervirulent phenotype in vivo.

However, this reviewer suggests more specificity. It is not acceptable that in a study of this relevance the plasmids are named using the list of replicons (reported incorrectly sometimes) and that there are not linear or circular maps that can help the readers to understand if they have the same plasmid or not in their hands.

Authors' responses: We thank Reviewer #1 for affirming the main conclusions of our study and for providing detailed, expert commentary on plasmid nomenclature and classification. All four major points have been addressed as described below including the design of a new figure (Figure S4) to provide maps/reference isolates for the most relevant pC types to enhance the readers' ability to translate our findings into their own plasmids.

Major points

1-PlasmidFinder contains 3 IncHI1B replicons.

IncHI1B(R27) is one of the 3 replicons that with IncHI1A and IncFIA(HI1) identifies the reference plasmid R27 (AF250878) prototype of all the IncHI1 plasmids that have 3 replicons HI1A-HI1B and FIA(specific of the HI1). IncHI1B(pNDM-CIT) and IncHI1A(NDM-CIT) identify the HI1-like plasmids prototype pNDM-CIT (JX182975, homology <94% with the others HI1).

IncFIB(Mar)/IncHI1B(Mar) identify the plasmids relative to the first described with these replicons called pNDM-Mar (JN420336, homology <94% with the others HI1).

HI1B(MAR) but not FIB (Mar) was also identified in the hvKp virulence plasmid pLVPK, 219385 bp, AY378100. It was not associated with FIB(Mar) but with repB_KLEB_VIR. pLVPK was not conjugative.

The acquisition of the FIB(Mar) replicon is coincident with the acquisition of a transfer locus and blaCTX-M-15 (Chen YT et al. Gene 337, 189-198 (2004); Arcari, G., & Carattoli, A. (2023). <https://doi.org/10.1080/20477724.2022.2121362>).

considering the complexity of the HI1 plasmids and their replicon type B, it is mandatory that the nomenclature IncFIB(K)/IncHI1B is changed with IncFIB(Mar)/IncHI1B(Mar) or FIB-HI1B-Mar.

Authors' responses: We thank the reviewer for this essential clarification on plasmid replicon nomenclature. We have revised it throughout the text and figures in coordination with your comments. Specifically, in Table S3 we included the PlasmidFinder data in the second tab with the raw output including replicon coordinates, accession, and % identity. We confirmed that PlasmidFinder identified all IncHI1B replicons from pNDM-Mar plasmid accession JN420336 and we changed the naming of IncHI1B to IncHI1B(Mar). However, the average percent identity of IncHI1B(Mar) found on pC4 plasmid was much lower (91.3%) than the average percent identity of IncHI1B(Mar) found on pC2 plasmids (avg. 99.6%).

In agreement with the reviewers comments, the pC4, pLVPK-like plasmids in this study carried IncHIB(Mar) but not IncFIB(Mar). Instead these plasmids carried IncFIB(K) (repB), accession

JN233704, >91% identity. Changes to indicate IncFIB(K) were made throughout the revised manuscript.

In Figure 3, we indicated that the IncHI1B replicons are IncHI1B(Mar). We confirmed that the pC4 hybrid plasmid found in isolate 474269 is indeed carrying replicons IncFIB(K)/IncHI1B(Mar). The revised Figure S7 plasmid alignment shows MRSN474269_p1 shares 99% identity to the pLVPK plasmid and that IncFIB(Mar) is not present.

2-virulence markers are not correctly defined. *rmpA2* is out of frame in most the FIB-HI1B-Mar plasmids. This impacts on the virulence but has not been clearly stated in the paper. Biomarkers are searched on genomes must it should be carefully checked if the database also indicates partial coverage of the product and potential pseudogenes. The concept of virulence may vary for mutations in the *rmpA* and *rmpA2* genes but also because of mutations in *wzi* (Ernst et al., Nat Med. 2020 May;26(5):705-711. doi: 10.1038/s41591-020-0825-4.)

Authors' responses: We agree this is an important precision.

First, regarding *rmpA/rmpA2* truncations. We now indicate that “The *rmpA* regulator of mucoviscosity and capsule synthesis was detected in 41% (n=91 including 11 with truncated alleles) while *rmpA2* was present in 44% isolates (with 94 of 98 isolates carrying a truncated version)” in the result section (lines 160-162). However, no direct functional evidence of the impact of naturally occurring truncations is available. In a previous publication (PMID: 38231533), we observed truncations in *rmpA2* at similar frequencies in phenotypically hypervirulent (87%) and non-hypervirulent (88%) isolates. In the same study, truncations in *rmpA* were detected in 27% of isolates carrying this biomarker despite all being categorized as hvKp phenotypically. The above statements are now discussed in the revised manuscript (lines 472 to 477).

Second, regarding *wzi* and other virulence modulators: we agree with the reviewer that factors beyond the 5 virulence biomarkers modulate virulence in Kp. As demonstrated by Ernst et al. (PMID: 32284589), *wzi* mutations that alter capsule type can influence virulence phenotypes. However, this manuscript's scope is deliberately focused on defining the phenotypic consequences of the specific plasmid acquisition events associated with the five established hypervirulence biomarkers. We have added a sentence to the Discussion acknowledging that capsule diversity, *wzi* variation, and other factors can modulate pathogenic potential in ways not captured by the five-biomarker model, and that these represent important areas for future investigation (lines 478 to 481).

Line 295: pC-2 (IncFIB(Mar)/IncHI1B(Mar) hybrid plasmid carrying 4 virulence biomarkers, including *iuc1*, and the NDM-1 carbapenemase. *RmpA2* is out of frame?

Authors' responses: We indicated that the *rmpA2* is truncated in lines 346 and 348.

3-Most of the collection is in contigs and only 53 are complete plasmid sequences. The general definition of the pC groups is done on replicons and biomarkers.

Authors' responses: We welcome this opportunity to clarify an important methodological point. The mge-cluster analysis was performed on 7,347 complete plasmids retrieved from NCBI and 173 complete plasmids sequenced in this study. mge-cluster performs whole-sequence, unitig-based pairwise Jaccard distance comparisons across the entire plasmid sequence, not just replicons or gene content. This produces a plasmid cluster (pC) classification rooted in overall sequence relatedness. This distinction is now explicit in the revised Results (lines 189-193) Methods section (lines 564-569).

Is it possible that resistance or biomarkers are within the same bacterial gene content but are not physically located on one single plasmid scaffold?

Authors' responses: For all isolates that we have long read sequencing we have identified the location of the resistance and biomarkers found on either the plasmid contigs (TableS3) or the chromosome (TableS1). Furthermore, Table S4 precisely lists the location of virulence and resistance genes on the 7,347 plasmids contigs we retrieved from NCBI. Additionally, in Figure 3 the location of virulence and AMR genes is indicated in the grey boxes. For example, in isolate 862392 the *iuc* biomarker is found on a pC-23 plasmid while *iro*, *rmpA*, and *peg-344* are found on "C" or circularized chromosome.

Please check and present a single accession no. and/or a map of the reference for each of the most relevant PC type.

Authors' responses: Thank you for this suggestion, we have substantially expanded the description of the major pCs, in the Result section (lines 204-225). We now provide a reference/accession for each. In addition, we also created including designed a new Figure S4, where we provide a circular map of all the MRSN plasmid sequences within the 5 most prevalent pCs (pC2, pC4, pC10, pC38, and pC39). Isolate names, plasmid reference sizes, virulence/resistance genes, and replicons are identified to enhance the readers' ability to visually compare to their own plasmids.

This avoids mistakes, example at line 104:

"IncFIB(Mar)/IncHI1B dual replicon type carrying resistance (*bla*NDM-1, *armA*) and *hvkP*-specific virulence (*iuc*, *rmpADC*, *rmpA2*, and/or *peg-344*) biomarkers that are found in globally epidemic nosocomial MDR lineages such as ST-307 and ST-147 (26,27)".

NOT TRUE for ST147: *bla*NDM was not on the FIB-HIB(Mar) plasmid but on a separate FIB(Qil) plasmid. From reference 26: While the *bla*NDM-1 gene was carried by an IncFIB-type plasmid, our investigations revealed that the ST-147 lineage from Italy also acquired a hybrid IncFIB/IncHIB-type plasmid carrying the 16S methyltransferase *armA* gene as well as key virulence biomarkers often found in hypervirulent isolates.

Authors' responses: Thank you for pointing to this possible inconsistency. In our previous study (reference 26), we did show that two configurations exist in ST-147 outbreak isolates from Italy: isolates carrying *bla*NDM on a separate FIB(Qil) plasmid but also isolates carrying NDM-1, *armA*, *iuc*, *rmpADC*, *rmpA2*, and *peg-344* on a hybrid IncFIB(Mar)/IncHI1B(Mar) plasmid. Please refer to Figure S3A from reference 26. To avoid any ambiguity, we have revised the sentence to read: "...biomarkers that are found in subsets of globally epidemic nosocomial MDR lineages such as ST-307 and ST-147 (26,27)" (lines 108-109)."

Is it true that the pC10 has *incL-oxa-48* and *armA* on the same scaffold? Is it IncM?

Authors' responses: Yes. PlasmidFinder analysis of the pC10 plasmid contigs from this study confirmed that OXA-48 and *armA* co-locate on the same contig, with 88% identity to the IncL/M reference accession AF550415. This is consistent with previously described pOXA-48-like plasmids that have acquired *armA* via transposition. These results are shown in Figure 3 and Table S3.

Reviewer #2 (Remarks to the Author):

This study by Sapre, Martin, and colleagues presents a comprehensive genomic and phenotypic analysis of the convergence of antimicrobial resistance and virulence in *Klebsiella pneumoniae*. By analyzing a large, global collection of isolates and performing deep plasmidomics, the authors demonstrate that convergence predominantly occurs through MDR-cKp acquiring plasmids that carry an incomplete set of virulence genes, with the aerobactin locus being the universal common denominator. Crucially, they provide in vivo evidence that

these convergent isolates exhibit enhanced siderophore production but do not display the hypervirulent phenotype, challenging the assumption that genomic convergence equals phenotypic hypervirulence. The work is of high significance, however, there are some concerns of this study.

Authors' responses: We thank Reviewer #2 for recognizing the high significance of our work (and in particular the critical value of combining genomics and in vivo phenotypic testing) and for the detailed, constructive major and minor comments. All concerns have been addressed as described below including a new figure (Figure S.8) and a detailed analysis of the virulence-carrying region and the evolutionary relationship between pC2 and pC4 plasmids.

Major concerns:

1. The authors state that pC2 plasmids carry an incomplete set of virulence biomarkers and are more frequently associated with the carriage of only *iuc* and *rmpA2*. However, the manuscript lacks analysis of the evolutionary relationship between pC2 and pC4 plasmids. This is a critical gap in understanding the origin of these successful convergent vectors. Do pC2 plasmids represent degenerated variants of the canonical pC4 (pLVPK-like) virulence plasmids that have lost *iro*, *peg-344*, and other markers over time? Alternatively, did they arise through independent acquisition events where the *iuc* and *rmpA2* loci were incorporated into a non-virulence IncFIB(Mar)/IncHI1B plasmid backbone? These two scenarios have fundamentally different implications for understanding the evolution of convergence.

Authors' responses: We thank the reviewer for identifying this as a critical gap and we fully agree. We have performed a detailed comparative analysis of the plasmid sequences and, crucially, the genomic regions flanking each virulence locus across all convergent isolates (Figure S.8). The results unambiguously favor the second scenario: pC2 plasmids are not degenerate derivatives of canonical pC4 plasmids that have lost virulence genes. Instead, they arose through the insertion of virulence gene-carrying mosaic regions into a non-virulence IncFIB(Mar)/IncHI1B(Mar) backbone.

Specifically, when compared to a pC2 backbone (i.e. the original pNDM-Mar plasmid NC_016980.1 lacking any virulence gene), pC2 plasmids from convergence events G-N carried a mosaic virulence region with sizes ranging from 115 to 78 kb. This identified pC2 as potential recipients.

When compared to pLVPK (canonical pC4 virulence plasmid), a similar modular region with all 5 virulence biomarkers, was identified with a size of 111kb and an average nucleotide identity of 99.99%. This identified pC4 as potential donors.

The mosaic virulence region was bounded by an IS3 and IS5 family transposases. The boundaries were conserved; however, no target site duplication was identified. Instead, the mosaic virulence region was directly flanked by two homologous regions shared by pNDM-Mar NC_016980.1 (and all other pC2 backbones) and pLVPK (pC4 donor). Upstream (Homologous Region Left or HR-L), a small 33 bp region sharing 91% nucleotide identity. Downstream (Homologous Region Right or HR-R), a large 21 kb region containing the tellurite operon involved in gut colonization (PMID: 33930099) and sharing 98.88% average nucleotide identity between NC_016980.1 and pLVPK.

We have included this new data in the result section of the manuscript (lines 298-317), in a newly created figure (Figure S8) and in the discussion (lines 422-430). Further studies (including comprehensive phylogenetic analysis of regions of all pC2 backbones, pC2 virulence

plasmids, and pC4 virulence plasmids) will be needed to reconstruct the stepwise evolution of these plasmids, determine the mechanism of transfer for the virulence mosaic region (transposition or homologous recombination), and infer if these originated from a single or multiple acquisition events between pC4 donor and pC2 recipient plasmids.

2. The authors repeatedly emphasize the "conjugative" nature of pC2 plasmids and the "non-conjugative" nature of pC4 plasmids, using this to explain their differential distribution. However, direct experimental evidence for these claims is lacking.

Authors' responses: We agree and have revised the manuscript to explicitly support these statements with published experimental evidence. Specifically, the non-conjugative nature of canonical pC4 plasmids (pLVPK-like) is supported by: their lack of transfer genes as shown by complete genome sequencing and annotation (PMID: 15276215) and demonstrated inability to transfer by conjugation in experimental settings (PMID: 35863038). The conjugative nature of pC2 (IncFIB(Mar)/IncHI1B(Mar)) plasmids is supported by: experimental conjugation of the prototype pNDM-Mar plasmid (PMID: 39888998) and in silico identification of a complete *tra* transfer locus (PMID: 37971980). These citations are now included in the revised manuscript in the Discussion (lines 411-419).

3. The authors argue that pC2 plasmids are more successful in MDR-cKp backgrounds due to their "conjugative nature" and "lack of hypervirulence phenotype" compared to pC4 plasmids. However, critical factors include plasmid stability, fitness cost imposed on the host, host range, and compatibility with resident plasmids all influence plasmid success. The authors do not provide data on any of these parameters.

Authors' responses: We agree that these factors should be acknowledged. We have revised the Discussion to explicitly acknowledge that the observed predominance of pC2-mediated convergence events likely reflects a combination of factors, of which conjugative ability is only one. Specifically, we now discuss "Besides their high mobility (with conjugation frequencies of 10⁻¹ to 10⁻⁴), their stability and compatibility with other frequent resident plasmids in MDR-cKp (such as IncR and IncL/M) likely account for their success (PMID: 39888998) despite a relatively narrow host range and an expected fitness cost (PMID: 38906484) "(lines 411-419).

4. 53 isolates were selected for long-read sequencing and complete plasmid resolution. The criteria for selection are not clearly defined. If isolates were selected based on specific genetic features without a systematic sampling strategy, the plasmid diversity captured may not be representative of the broader population.

Authors' responses: We appreciate this comment and recognize the importance of transparent selection criteria. The 53 isolates were selected using a two-step strategy, now described explicitly in the revised Methods section (lines 513-517). First, at least one isolate was selected to represent each of the 25 independent convergence events identified by parsimony analysis of the full cohort (Fig. 4 and S6). Second, within each convergence event, additional isolates were selected to represent each distinct pattern of virulence and resistance gene presence/absence (as shown in Fig. S5), ensuring that the full within-event micro-evolution of plasmid acquisitions was captured. By design, this strategy maximizes genomic diversity both between and within convergence events, though we acknowledge in the Discussion that it may not capture the full frequency distribution of plasmid types in the broader population (line 469).

5. The study conclusively shows that enhanced siderophore production, driven by *iuc* acquisition, does not confer systemic hypervirulence in the murine model. However, existing

literature suggest that iron acquisition provides ecological advantages beyond systemic invasion. The discussion of these trade-offs is currently underdeveloped.

Authors' responses: We thank the reviewer for this comment. We have expanded the discussion paragraph (lines 441-450), drawing on the recent study from Lim et al. (PMID: 40184910), who demonstrated in isogenic hvKp mutants that aerobactin is the key siderophore for intestinal colonization, not systemic virulence. Genomic data from ~4,000 virulence-carrying plasmids and strains corroborate this: fecal isolates predominantly carry *iuc* alone. We now explicitly suggest that the universal *iuc* acquisition across convergence events is likely due to colonization and transmission fitness rather than enhanced invasiveness.

Other specific comments:

1. Line 121, "The frequency of isolate collection varied by year and country." Please provide a brief explanation for this variation. More importantly, discuss the potential consequences of this sampling bias on the conclusions.

Authors' responses: We have revised the sentence to read "All collected isolates were sequenced but frequency of isolate collection varied by year and partner sites, influenced by local laboratory capacity and surveillance funding cycles rather than underlying infection burden." (lines 124-126).

More generally, we have revised the Results (lines 121-126) and Methods (lines 500-503) section to explain that the variation in isolate collection frequency reflects the structure of the MRSN/GEIS surveillance network, in which sampling intensity at each partner site varied according to local laboratory capacity, active surveillance protocols, and annual funding cycles rather than underlying disease burden.

Finally, we have revised the Discussion to further discuss the limitations due to the sampling strategy that was however necessitated as we needed physical isolates for *in vivo* phenotypic testing. (lines 466-472).

2. Line 124, Please clarify whether the collection includes any duplicate isolates. Additionally, provide more detail on the strain isolation methods and inclusion criteria in the Materials and Methods section.

Authors' responses: Duplicate isolates were not intentionally collected; each isolate in the collection represents a distinct clinical specimen from an individual patient at the time of clinical care. Isolate selection was based on MDR phenotype (non-susceptibility to ≥ 3 antibiotic classes) as the inclusion criterion. These criteria are now stated explicitly in the Methods section (lines 493-503).

3. Line 138, "This largely correlated with the sampling of clonal, geographically clustered isolates." Please state this correlation more specifically.

Authors' responses: The sentence has been revised to read: "This temporal increase in carbapenemase-producing isolates was largely attributable to the expanded sampling of geographically and clonally clustered isolates in the later years of surveillance, particularly MDR-Kp ST-16 and ST147 from Thailand and ST-395 from Georgia which comprised 82-95% of carbapenemase-producing isolates" (line 143-146).

4. Line 180, Please define "MRSN"

Authors' responses: Corrected. At first use, MRSN is now defined as the Multidrug-Resistant Organism Repository and Surveillance Network (line 122).

5. Line 205-206, The figure legend appears to be incorrectly labeled. Please verify that the colors and labels in panels C and D correspond correctly to virulence and resistance gene carriage as described in the text.

Authors' responses: Thank you. Confirmed and corrected.

6. Please maintain consistent formatting with plasmid cluster designations, for example PC2 and PC-2.

Authors' responses: All plasmid cluster designations have been standardized to the lower-case format pC followed by the cluster number (e.g., pC2, pC4) throughout the text, figures, and tables.

7. Line 259, Should be an enrichment...

Authors' responses: Corrected. "am enrichment" has been changed to "an enrichment." (line 282)

8. Line 313, Should be pC33.

Authors' responses: No change was made. Route Y is associated with a pC23 (not pC33) plasmid.

9. Line 329, The legend appears to have an error.

Authors' responses: Thank you. Corrected in the revised manuscript.

10. Line 444-445. "clast allele assignment" appears to be a typo. Please specify the exact procedure used for cgMLST allele assignment and minimum spanning tree generation.

Authors' responses: This was a typographic error for "cgMLST allele assignment." The revised sentence reads: "Core genome MLST (cgMLST) allele assignment and minimum spanning tree generation were performed with SeqSphere (47) with the *K. pneumoniae* /*variicola*/quasipneumoniae cgMLST scheme (<https://www.cgmlst.org/ncs>). To be included in the analysis, isolates had to contain 90% of the 2,358 genes included in the cgMLST scheme" (lines 538-541).

Reviewer #3 (Remarks to the Author):

The manuscript by Sapre et al. investigates the evolutionary pathways and phenotypic characteristics of *Klebsiella pneumoniae* strains that acquire both antimicrobial resistance and virulence traits, highlighting the predominance of siderophore production over classical hypervirulence in convergent lineages. The study is timely and relevant, particularly given the emergence of multidrug-resistant strains carrying hypervirulent traits, and it complements prior work on this topic (PMID: 41512896). While the study is interesting and the findings are valuable, several aspects of the experimental design and the interpretation of results require clarification before the conclusions can be confidently generalized. Some comments are provided for the authors' consideration.

Authors' responses: We thank Reviewer #3 for acknowledging the timeliness of this work and for the thoughtful, detailed comments. All four major concerns have been addressed.

Major comments

1. Sampling. The study relies on a highly specific set of 1458 isolates, predominantly from LMICs and a limited number from the United States ($n = 39$), seemingly drawn from a single context (MRSN). The criteria for this sampling strategy should be clearly explained and justified, as the broad conclusions risk overgeneralization. Heterogeneous temporal coverage, geographic bias, host population age/gender, and the potential inclusion of outbreak-associated or otherwise non-representative isolates could influence the observed distribution of clonal lineages. Clarification of how and why isolates were selected, their temporal and geographic distribution, individual epidemiological characteristics, and whether they reflect routine surveillance or targeted collections is essential to assess the robustness and generalizability of the findings.

Authors' responses: We agree with these limitations and have addressed them comprehensively in the revised manuscript. First, in the Method section, we now provide an explicit description of the surveillance network structure, including the criteria for site participation and isolate inclusion (lines 495-503). Second, we clarify that Figure S1/S2 and the opening Results paragraphs are meant to transparently communicate the geographic and temporal biases in the dataset (lines 124-126). Third, we have added a dedicated paragraph in the Discussion that directly addresses the generalizability of our findings. The key point we now make explicit is: this study was not designed as a global epidemiological survey of convergent Kp (comprehensive and rigorous population-level analyses for this purpose already exist and are now cited, including Gibbon and Couto et al and the European EARS-Net investigation (Linkevicius et al.) (lines 466-472).

The unique design of this study was to combine genomic and plasmidomic analysis with in vivo phenotypic characterization, which required physical access to cultured isolates and constrained us to our available collection. We emphasize that the biological conclusions regarding the phenotypic consequences of acquiring an incomplete vs. complete set of virulence biomarkers, and the dominant role of the aerobactin locus are derived from the 25 evolutionary independent events and do not depend on their geographic or temporal representation being proportional to global prevalence.

2. Use of “convergence” terminology. The use of the term “convergence” in this work is challenging and may cause confusion for different readers throughout the text. Convergent evolution is traditionally defined as the process by which different species independently evolve similar traits. When similar outcomes occur within populations of the same species, the term generally used is parallel evolution. In this study, the focus is on the acquisition of both antibiotic resistance and virulence traits. While the co-carriage of these traits could be described as phenotypic convergence, other adaptive traits located in the plasmids analyzed have not been examined and might also contribute to adaptation (a limitation that should be mentioned). Could be but these are two defined phenotypes. Also, it is confusing to describe isolates or lineages themselves as “convergent” (lines 67, 70, 101, 156, 160, 177, and elsewhere), because convergence refers to evolutionary processes or traits, not to individual isolates or lineages. Mechanistically, plasmid acquisition is primarily classified as horizontal gene transfer (HGT) rather than genomic convergence. Nonetheless, the description of the outcome at the population or phenotype level as convergent adaptation is frequent in microbial evolutionary genomics (see ref #38 for discussion about “convergent *Klebsiella pneumoniae*”), particularly relevant when distinct alleles of the same antibiotic resistance genes (e.g., NDM, CTX-M) or virulence genes (e.g., *iuc*) are involved as they may differ at the sequence level but produce the same adaptive phenotype. However, the process described here and other papers may be more

accurately described as “convergent adaptation via HGT”, “parallel acquisition of adaptive genes via plasmids”, or “mobile gene pool–mediated adaptation”. Strict genomic convergence requires independent mutations or events that evolve separately in different populations (e.g., *gyrA* mutations conferring quinolone resistance). In contrast, in this study, genes are imported from the mobilome rather than evolving *de novo*, so evolution occurs through the mobile genetic element pool rather than independently in each lineage. See previous comment about the sample.

Authors’ responses: We agree with the reviewer's analysis of the terminology and appreciate the opportunity to clarify our usage. The term “convergence” in the context of *K. pneumoniae* harboring both AMR and virulence genes has been used consistently in the clinical microbiology literature for over a decade (PMID: 28864028) and is used in authoritative reviews (PMID: 36089853 and PMID: 33558323) as well as in the very recent literature cited in this manuscript. Departing from this established usage would risk creating confusion. To remove any ambiguity with the evolutionary biology definition, however, we have revised the Introduction to provide an explicit definition at first use (line 100-102).

3. Convergence events. The treatment of convergence events in the manuscript requires a more detailed description and a precise definition. From the text, it appears that the authors consider the five putative virulence genes previously linked to hypervirulent *K. pneumoniae* as a “unit of selection,” noting that this set can be “incomplete” in some isolates. However, the coexistence of these disparate virulence determinants on a single plasmid is likely exceptional and primarily reflects the selection and transmission of the plasmid as a unit. Other scenarios are also possible, as the authors mention, including the acquisition of these genes individually or in different combinations by different plasmids. Therefore, it is important to clarify whether convergence events are being defined at the gene level or at the plasmid level (see how convergence is approached in reference 38). For antibiotic resistance genes, allelic variants of a single gene (e.g., *bla*_{NDM}) are treated as distinct genes. Yet some allelic variants may share a single evolutionary origin, rather than representing independent acquisition events, and conversely, the same allele can be repeatedly acquired by different lineages via widespread plasmids. Clarifying these points is essential to accurately defining what constitutes a convergence event in this dataset.

Authors’ responses: We agree and have substantially expanded the description of our methodology in both the Results (lines 267-269) section and Methods section (lines 571-582). Independent convergence events were identified using a parsimony-based approach applied to the full 1,468-isolate phylogeny. Briefly, each orthologous group of resistance genes (carbapenemases and ESBL alleles, grouped at >98% nucleotide identity) and virulence biomarker loci was mapped onto the phylogenetic tree. Gain and loss events were inferred by maximum parsimony using PAUP* (<https://paup.phylosolutions.com/>), with an asymmetric penalty matrix applying a 3:1 cost for gene gain vs. gene loss. This probabilistic reconstruction assigns gain events to the most parsimonious ancestral nodes. A “convergence event” was defined as a node where both an ESBL or carbapenemase gene and ≥1 virulence biomarker were inferred to have been acquired (or co-exist) since divergence from the most recent common ancestor, before the next speciation event. A new Figure 4 displays the full phylogeny annotated with all 25 inferred events (detailed figure provided as revised Figure S6), and the corresponding plasmid in the 25 representative isolates tested *in vivo*.

4. Novelty and context. The manuscript could better situate its findings within existing literature.

Previous studies have documented convergence of virulence and resistance traits in *K. pneumoniae* via plasmid hybridization (PMID: 41512896, reference 38). The present work provides complementary insights, showing that less than 5% of plasmids were hybrids, demonstrating the relevance of acquisition by independent plasmids rather than hybridization events. However, reference 38 is mentioned only once, and a more thorough comparison would strengthen the discussion of novelty and relevance. Additionally, isolates from wounded individuals in the Ukrainian conflict have been reported, with results similar to those reported here (PMID: 41177264), which could be relevant for contextualizing lineage diversity and sampling, tying back to the first comment on sampling.

Authors' responses: We have expanded the Discussion to include additional comparisons with Gibbon, Couto et al. (Lancet Microbe 2026) (lines 430 and 472). That study performed a global population genomic analysis of *Kp* plasmids carrying the *iuc3* virulence locus, complementing our findings in several ways: i) it demonstrates the global molecular epidemiology of *iuc3*-carrying plasmids using NCBI genomic data, and ii) it documents that less than 5% of plasmids in both studies are hybrid (carrying both AMR and virulence loci), highlighting that co-resident independent plasmids remain the dominant mechanism for co-carriage. Our study extends this picture by systematically characterizing all 5 virulence biomarkers across 25 independent convergence events, including *iuc1*, *iuc3*, and *iuc5* variants, not just *iuc3*; and most importantly providing *vivo* phenotypic validation for representative isolates of every convergence event.

Other comments.

Line 180, define the acronym MRSN and indicate the number of plasmid sequences involved.

Authors' responses: At first use, MRSN is now defined as the Multidrug-Resistant Organism Repository and Surveillance Network (line 122). The number of plasmids is now indicated (line 190).

Line 348 – Recurring and polyphyletic nature of convergence: Consider clarifying what is meant by “recurring” and “polyphyletic” in this context to ensure readers understand how convergence is being defined.

Authors' responses: The sentence has been revised to read: " At the lineage level, 25 independent convergent events distributed across 22 sequence types were identified, highlighting the recurring and polyphyletic (i.e. multiple independent occurrences across phylogenetically distinct lineages) nature of convergence." (lines 398-400)

Lines 187-199, plasmid description. The description of the plasmids here is currently unclear and could benefit from additional context. Introducing the type of each plasmid cluster (PC) at the beginning of the section would help the reader understand the rationale for the analysis. They only appear in the last sentences (e.g., pC10 as an IncL/M carrying *bla*_OXA-48). Also, it would also be helpful to clarify how these plasmids relate to previously described pOXA plasmids. Similarly, providing information on the relationships of pC38 and pC39 to other widespread plasmids would strengthen the interpretation of their significance. Finally, specify which *bla*_CTX-M alleles are being referred to, as this is critical for assessing the functional and epidemiological relevance of these plasmids.

Authors' responses: Thank you, this section has been substantially expanded and rewritten (lines 186-225). Each plasmid cluster is now introduced with its defining replicon type and representative reference plasmid before discussing the gene content in our collection.

Lines 477–485: Analyses described here are part of the study's results and should be moved to the Results section.

Authors' responses: We agree with this suggestion and this paragraph, initially in the methodology section was moved to the result section of the revised manuscript (lines 189-196).

General: Check spelling throughout the text (e.g., line 139) and ensure consistent nomenclature for antibiotic resistance genes and proteins. Use the acronym MDR consistently. Specify the number of isolates or plasmids corresponding to this study in the different sections for clarity.

Authors' responses: Spelling was checked throughout. Specific corrections include: "hvkp" to "hvKp" consistently; "Carabapenemase" to "Carbapenemase"; "trycylcer" to "Trycycler"; MDR is used as an acronym throughout and defined at first use; and all plasmid cluster references use lowercase pC consistently.

Reviewer #4 (Remarks to the Author):

Sapre and colleagues conducted a large-scale analysis of the co-occurrence of resistance and virulence genes in *Klebsiella pneumoniae* (Kp) strains and associated plasmids. The study focused on 1,468 Kp isolates collected across nine countries between 2011 and 2021, complemented by a broader and more representative dataset of 7,520 plasmids retrieved from NCBI.

The authors identified 25 independent convergence events, with acquisition of the aerobactin siderophore system emerging as the only consistent feature across all events. Convergent strains produced higher levels of siderophores compared to the control ckp1 strain; however, in vivo experiments did not confirm a corresponding increase in virulence. Based on these findings, the authors suggest that siderophore acquisition may contribute primarily to adaptive fitness rather than directly enhancing virulence.

In my opinion the major limits of this work are that it is focussed on a very specific Kp dataset without a rationale (or needs) and it considers resistance and virulence genes only. Furthermore, other minor modifications are needed.

Authors' responses: We thank Reviewer #4 for the careful review and constructive suggestions. The two major concerns regarding the genomic dataset and gene focus are addressed below, followed by point-by-point responses to all minor suggestions. including the design of a new figure (Figure 4) summarizing all convergence events.

i) The genomic dataset:

The Kp dataset includes mostly strains isolated from Thailand, Georgia, Perù and Jordan. Thus it is very biased for these countries, that in my knowledge cannot be considered as representative for some reason. Just an example: In figure 1A, the number of Virulence Biomarkers appears to be not fully independent from the isolation country, which in turn could be associated with the local epidemiology.

The dataset is also strongly biased for what concerns isolation time, with the 2018 over-represented. Considering the dynamic nature of the Kp epidemiology, a time bias can drastically affect the results.

One of the effects of bias sampling is evident considering the very low representation of the pandemic ST258, while the less diffused ST258-derived ST512 is well represented (Figure S1).

This bias is particularly important considering that the sampling period was characterised by the diffusion of the ST258.

I suggest to the authors to repeat the analysis including also Kp genomes retrieved from NCBI, paying attention to geographical/time stratification.

Authors' responses: We fully acknowledge the geographic and temporal biases in the dataset, which we designed Figure S1/S2 and the opening Results paragraphs to transparently communicate. We have now reinforced this with explicit text in the Methods (lines 493-503) Results (lines 121-126) and a new Discussion paragraph dedicated to limitations (lines 466-472).

The specific suggestion to repeat the analysis with NCBI genomes warrants careful consideration. The plasmidome analysis already incorporates 7,347 complete Kp plasmids from NCBI representing 306 STs and ≥47 countries. What is specific to our collection is the requirement for physical bacterial isolates for *in vivo* phenotypic testing. This phenotypic characterization, testing 25 independent convergence events in the murine subcutaneous model, is the primary value that this study adds to what can be derived from genomic databases alone, and it is not possible with NCBI sequences. We therefore maintain that expanding the genomic collection to all NCBI Kp assemblies would not advance the core objectives of this study.

Regarding ST-258 underrepresentation: this reflects the geographic focus of the surveillance network (predominantly LMICs and the Asia-Pacific region) during the sampling period, and is acknowledged as a limitation (lines 466-472). The 25 convergence events identified here include diverse carbapenemase types (NDM, OXA-48, ESBL) and do not depend on ST-258 for any of the major conclusions. Our findings are complementary to, and consistent with, the extensive literature on KPC-producing ST-258.

ii) Genes:

The authors focused the attention of virulence and resistance genes. This approach can drastically limit their ability to investigate the evolutionary trajectories. In particular, even if Kleborate is probably the most used tool for Kp typing, there are other tools/dataset for the identification of virulence genes. Unfortunately, the definition of virulence is not simple as for the antibiotic resistance, and thus a larger gene dataset could provide a more precise description of the virulence of the strains. Indeed, the authors did not find any correlation between virulence gene content and real *in vivo* virulence.

As part of the phylogenomics analysis flow, the authors used Roary. I suggest they analyse the orthology groups distribution among the strains to identify putative genes associated with virulent phenotypes.

Authors' responses: We respectfully maintain that the scope of this study, focused on the five established biomarkers of hvKp and their associated phenotypic consequences, is well-defined and justified by the existing literature. As extensively documented, hypervirulence in Kpn is driven by the virulence plasmid (we direct the reviewer to PMID: 39178743 - when the plasmid is lost, the hypervirulent phenotype is lost), and the five biomarkers studied here are the most predictive of that phenotype (PMID: 38231533). This focused approach was necessary to conduct the parallel *in vivo* validation for each of the 25 convergence events.

Genome-wide approaches to identify additional genetic or epigenetic virulence markers have been pursued by others (PMID: 42204287; PMID: 38112443) and remain to be functionally

validated. We acknowledge in the revised Discussion that additional genomic features not captured by the five-biomarker model may contribute to the fitness of convergent lineages in hospital settings (line 478-482).

Point by point suggestions:

line 136: I suggest to perform a correlation analysis between time and the rate of CP-producing isolates

Authors' responses: Thank you. Done as suggested. The increased prevalence of CP-producing isolates over time is now substantiated with a correlation analysis (lines 141-142).

Legend 2: C and D letters are inverted

Authors' responses: Thank you. Corrected in revised manuscript.

line 226: in figure 2 we can see that within the cluster pC2 there are two subclusters, one including mostly virulence genes and one mostly (or better a lot of) resistance genes.

Authors' responses: We agree with the reviewer, this is a very interesting pattern consistent with the known biology of IncFIB(Mar)/IncHI1B(Mar) plasmids. We have added a full analysis of pC2 backbone (pNDM-Mar) compared to virulence-carrying pC2. We found that pC2 arose through the insertion of 115-78 kb virulence gene-carrying mosaic regions, homologous to pC4 plasmids from hvKp, into a non-virulence IncFIB(Mar)/IncHI1B(Mar) backbone. The mosaic virulence region was bounded by an IS3 and IS5 family transposases. The boundaries were conserved; however, no target site duplication was identified. Instead, the mosaic virulence region was directly flanked by two homologous regions shared by pNDM-Mar NC_016980.1 (and all other pC2 backbones) and pLVPK (pC4 donor). Upstream (Homologous Region Left or HR-L), a small 33 bp region sharing 91% nucleotide identity. Downstream (Homologous Region Right or HR-R), a large 21 kb region containing the tellurite operon involved in gut colonization (PMID: 33930099) and sharing 98.88% average nucleotide identity between NC_016980.1 and pLVPK. These results are described lines 298-317 and in a new figure S8.

We finally we added a sentence in the discussion proposing that the analysis of the macro and micro evolution of pC2 plasmids (e.g. gene content, fusion, recombination, mutations etc.) through time and lineages would be a valuable direction for future study (line 427-430).

line 259: typing error

Authors' responses: Thank you. Corrected in revised manuscript.

Figure S5: the image is unreadable to me (I think too low DPI)

Authors' responses: A revised high-resolution image for Figure S6 (originally Figure S5) is provided in the revised manuscript.

Figure 4: I really appreciate this figure, I suggest the author make a similar figure of each phylogenetic lineage for which they have convergent routes.

Authors' responses: Thank you for this suggestion. Instead of 20+ individual figures, we designed a new figure (Figure 4 in the revised manuscript) showing the entire phylogenetic tree, the 25 convergence events, and the corresponding plasmid configurations in the 25 representative isolates for which phenotypic testing was done.

Figure 5: in the legend the letters C and B are inverted. I also suggest performing a regression between [siderophore] and LD50.

Authors' responses: Thank you. We have corrected the figure legend (now Figure 6). We agree that evaluating the relationship between siderophore production and LD50 is valuable. However, because no mortality was observed at the highest tested doses for most non-pC4 isolates, these LD50 values represent lower-bound, right-censored data points rather than exact, continuous measurements. Performing a regression on censored data would introduce parameter bias and violates the assumption of normally distributed residuals. Instead, we provide side-by-side visual comparisons in Figure 6 (Panels A and C) for all isolates.