

# In-host evolution of *Yersinia enterocolitica* during a chronic human infection

Corresponding Author: Dr Cyril Savin

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript of Savin et al is an enormous effort to characterize the in vivo evolution of a bacterial pathogen in a chronically infected patient. I have to confess that I could not evaluate the correctness of all the details and conclusions presented in the manuscript, however, the manuscript in the big picture is an excellent demonstration of the possibilities the modern scientific tools allow.

Overall the manuscript is well written and I have only a few minor comments:

L173-180. The text is somewhat confusing as if some words are missing. Please revise.

L226. ... in early versus late isolate ...

L409. add space after "and"

L443. use OD600 or similar throughout the text

Table S8. Reference column. Replace "This sudy" with "This study"

Reviewer #2

(Remarks to the Author)

This is an interesting article that describes a bacterial adaptive evolution story on a case of chronic infection caused by *Yersinia enterocolitica*. The study postulates that the persistent infection in the patient was likely caused by vegetation that developed on the atrial lead of pacemaker post implantation, although this vegetation has not been detected until 2006. Four isolates of *Y. enterocolitica*, designated Ye.1-4, were obtained from the patient and their genomes were sequenced and combined with 259 additional *Y. enterocolitica* genomes of the same biotype of Ye.1-4 (*Y. enterocolitica* bioserotype 4/O:3). The results demonstrated that the multiple infection episodes were a result of chronic colonization rather than separate, independent infections. The authors assessed the susceptibility of the four isolates to 33 different antibiotics or antibiotic combinations and determined their growth rate. An extensive in-host evolution study of the bacterium was carried out using genomic, proteomic and transcriptomic approaches, resulting in identification of mutations in key genes that accounted for the observed antibiotic resistance and the abundant differences of proteins that could promote the adaptability within the host.

The data of this manuscript is well-presented and the results are logically organized. Nevertheless, I am a little concerned about the robustness of evidence supporting some of the authors' claims about mutations conferring resistance to specific antibiotic. Additionally, the effects of mutations on the activities of the proteins in question, such as DksA among others, requires further experimental validation and/or the provision of supporting references to substantiate these claims.

Major points:

1. It is regrettable that the isolate Ye.X, obtained in August 2001, was not collected and included in the genomics, proteomics and the phenotypic analysis. The absence of this isolate from the study impairs the comprehensiveness of the evolutionary trajectory of bacterium, particularly under condition that Ye.1 and Ye.2 are very close in isolation time point (only apart from less than 1 or 2 months; note that the isolation time of Ye. 2 is reported as January 2000 in the text, yet it is Feb 15, 2000, in Fig S3), and isolation time points for Ye.3 and Ye.4 are only 9 month apart. Ye.X, having evolved for 2

years from Ye.1 and Ye.2, could have served as a critical bridge connecting the endpoints of in-host evolution trajectory, potentially revealing a more complete picture of micro-evolutionary changes occurring within the host.

2. Except examining the antibiotic resistance characteristics, have the authors compared the virulence of the early and late isolates? The authors stated that 'comparative proteomics of Ye isolates reveals an extensive metabolic rewiring, reduced virulence determinant expression and modulation of antibiotics resistance.' The determination of virulence of the 4 isolates will provide evidences to corroborate those finding from omics studies.
3. The growth rate of the late isolates is obviously slowed down compared with the two early isolates. Have the authors observed the differences between the mutation rates of early isolate and late isolates? Line 177-180, dose the least-squares dating tool used here has taken the variations of the isolates' growth rate into account, when estimating the occurrence of diverging into Ye.3 and Ye. 4?
4. The accumulated SNPs for the four Ye strains, including those on internal branches, appear be restricted to those within CDS-region (Lines180-181, Figure 3A, and TableS4-SNP-specific). According to the methods described, Snippy was used to call core-genome SNPs, which typically yields results including both CDS and intergenic regions. Please consider to include results from the intergenic regions in the manuscript for a more comprehensive analysis.
5. Similar to the SNP analysis, it appears that the analysis of small insertion and deletion analysis excludes the intergenic regions, which may affect the conclusions drawn in Line 170. Additionally, the Methods section lacks some essential descriptions on indel analysis, for instance, the thresholds used in this study (e.g.,  $\leq 50$  bp, supported by at least a specified number of reads, and the allele frequency, defined as the number of reads supporting the mutation divided by the total reads at the site).
6. The assertion that "Truncation of DksA is one of the main drivers of proteome changes" may not be fully substantiated. In Lines 267-268, the authors state that 'early vs. late isolate comparisons had a correlation coefficient ranging from -0.29 to -0.43 with comparisons of a Ye WT vs.  $\Delta$ dksA mutant.' However, in Figure 6 and Table S9, the late vs. late isolate comparison (Ye.3 vs. Ye.4, Number 6) also shows a correlation coefficient of -0.29 with the Ye WT vs.  $\Delta$ dksA mutant comparison (Number 51). If I understand correctly, truncation of DksA would be just one of the possible factors impacting proteome changes, rather than a primary driver. Since other truncated genes in Table 2 are not included in the 57 publicly available transcriptome datasets, the discussion appears to overemphasize DksA without addressing other potential contributing factors.

#### Minor points:

1. Line 142-144, it appears that the ciprofloxacin resistance of Ye strains depends heavily on the culture medium, and the increasing resistance to ciprofloxacin from Ye.1 to 4 can only be seen when they were grown in LB. Could the authors please potential explanations for this phenomenon, such as mutations in gene associated with carbon and nitrogen sources utilization and metabolism?
2. Lines 190-200, it is interesting that positive selection signals were found in genes encoding acyl-CoA thioesterase II and phosphate transporter. I propose, if feasible, the construction of mutant strains or proteins for these genes. Comparing the activity between the wild-type and the mutants would elucidate the biological relevance of these mutations. These results would make the conclusion more robust and convinced.
3. Line 93, the patient received only 18 months of ciprofloxacin therapy (up to July 2001), yet from Ye. 3 to Ye. 4, ciprofloxacin resistance increased significantly in the absence of ongoing ciprofloxacin selection pressure. Could the authors please propose a potential mechanism or driving force underlying this phenomenon.
4. Line 134, the author stated that an unexpected AG MIC decrease was observed for the late isolates Ye.3 and Ye.4 compared to the early isolates Ye.1 and Ye.2, could they please make some explanation on this point.
5. Line 382, this suggests that the late slow-growing isolates accumulated less mutations and evolved even slower than expected, while most of the in-host evolving strains previously described showed increased mutation rate. The rationale behind this conclusion is not immediately clear. Could the authors please provide a detailed explanation? Additionally, is it possible that the slow-growing and longer doubling time is the main reason responsible for the lower mutation rate of the late isolates?
6. Lines 205-210, explaining the genetic basis of the resistance to quinolones in Ye.3 and Ye.4....., explain cefoxitin and ceftriaxone resistance. These conclusive statements need supporting from the experimental results. Currently, the authors provide only phenotypic resistance and comparative genomic analysis of the isolates. Without experimental validation specifically designed to test this variant, the conclusion cannot be directly drawn. Additionally, the analysis focused solely on coding regions, without examining potential mutations in the upstream intergenic regions of the gene. A similar issue arises with the claim regarding ompF-mediated cephalosporin resistance. It would be beneficial for the authors to cite relevant published literature, if available, that documents a direct relatedness between specific types of mutations and antibiotic resistance.
7. Line 227, please explain 'the loss of the pYV virulence plasmid is a known laboratory artifact'. How did the authors confirm that pYV was present in the Ye isolates when they were initially isolated? Additionally, were the CDS on pYV absent from the genome sequencing data?
8. Figure S2, the legend is confusing. While the blue circles in 'a Ye.1' represent the inhibition radius for Ye 1, how do green circles represent the inhibition radius for Ye 2 in the same panel? The same question is for 'b Ye.2'. Why the assay results for Ye. 1 appeared in plate for Ye.2. Please provide some additional description to make it clear.
9. Line 127-128, Figure S3 is for Ye. 1 and Ye. 2, thus Fig S3-4 in braces should be Fig S3, please correct.
10. Line 718, legend for Fig. 2, "a,f Ye.1" should be "a,e Ye.1", please correct.
11. Figure 3b, how were the strains selected? Were they chosen based on the shortest genetic distance to the four Ye strains? Additionally, were any temporal signals evident before estimating the date, possibly using tools like TempEst?
12. The title of Table 3, Log2(foldchange) of the proteome pairwise comparisons for selected proteins. I suggest to provide the selecting criteria for the proteins shown in this table.
13. In the tables and supplementary tables, decimal points are consistently shown as commas.

#### Reviewer #3

(Remarks to the Author)

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

#### Reviewer #4

(Remarks to the Author)

This is an interesting paper on the evolution of a commensal/pathogen *Yersinia enterocolitica* during 14 years within a patient with bacteremia. This research is important from bio-medical (antibiotic resistance, development of antibiotic therapies, host-pathogen interactions, pathogen persistence evolution) and fundamental prospective (e. g. pangenome evolution).

This manuscript is interesting for publication, however, requires some clarifications and more context. Below my comments. Comments:

1. Reported pangenome consists of 3,824 genes (line 464), however genomes of *Yersinia enterocolitica* appear to have around 3,900-4,000 genes or more, so pangenome of 4 strains has to be at least 3900 genes. Would it be possible to clarify following questions:
  - Provide assembly statistics values like N50 for strains sequenced in this study, genome completeness and contamination obtained for example by checkM.
  - What are the thresholds for gene clustering in pangenome building? Perhaps *Yersinia* has a lot of paralogs which were combined together, or a lot of genes were lost during evolution during bacteremia and pangenome became smaller than average genome of *Yersinia*. Perhaps there are other scenarios/explanations?
  - Did core genome size decline during the 14 years of evolution?
2. It may be somewhat unexpected that strains from this study are on much longer branches in the tree (Figure 3). How pure was the culture, could there be within species divergence in the cultures which were sequenced? Is mutation rate higher than expected? Does the tree have an outgroup from a different species? How was the tree rooted? (Is it rooted?)
3. If the genome of these strains is significantly reduced what are the origins of the closest public strains? Are there other bacteremia/antibiotic treatment strains or extra-host environment? I think it would be beneficial to plot metadata from Table S3 and bootstrap support values on the tree.
4. To which fraction of the pangenome do genes belong that change their expressions in proteomics (and/or public transcriptomic data)? Are there more core genes which change expression or gain deleterious mutations?
5. Are there any new genes that were acquired in later strains (e.g Ye4) but were absent in Ye1?
6. Genomes are deposited to ENA but associated metadata about those genomes (e.g. timepoints of sampling) does not seem to appear on the site.

Version 1:

Reviewer comments:

#### Reviewer #2

(Remarks to the Author)

The authors supplemented additional experiments and conducted detailed structural analyses of protein mutants, which effectively addressed the concerns I raised. The revised manuscript now exhibits a more rigorous presentation, with data that better substantiate the research conclusions. However, a few minor issues remain and should be addressed as follows:  
Line 169, "Genomic analysis supports a chronic Ye infection" is not an appropriate title for the content that followed.  
Line 190, "..... indicating that evolution had already started within the first months of infection". Evolution is going on all the time. Please consider to revise this sentence.  
Line 278-279, "affecting two cytoplasmic, one 278 non cytoplasmic and three transmembrane domains". This description is not appropriate, please correct. They are not three domains, but different regions in one domain.  
Line 306-309, "....., we assessed cell invasivity which is independent of the type 3 secretion system encoded by pYV genes" is not correct. Several Yop effectors, like YopE, YopT and YpkA, show enzymatic activity to proteins that are involved in cytoskeletal remodeling, which will influence the invasion capability of the bacteria. Supplementary Fig. 11, Cell invasion assay, according to the description of "Gentamycin protection assay" in the Supplementary methods, it appears to be determining the living bacterial inside the cells but not the bacteria that invade the host cells. Ye.1 and Ye.2 exhibit much higher ability to replicate inside macrophages. Please consider to correct the relevant descriptions.

#### Reviewer #3

(Remarks to the Author)

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

#### Reviewer #4

(Remarks to the Author)

Authors have addressed my questions and suggestions. I would like to recommend the manuscript for publication.

**Open Access** This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were 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 COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript of Savin et al is an enormous effort to characterize the in vivo evolution of a bacterial pathogen in a chronically infected patient. I have to confess that I could not evaluate the correctness of all the details and conclusions presented in the manuscript, however, the manuscript in the big picture is an excellent demonstration of the possibilities the modern scientific tools allow.

**We thank the reviewer for this positive comment.**

Overall the manuscript is well written and I have only a few minor comments:

L173-180. The text is somewhat confusing as if some words are missing. Please revise.

**Done. A verb was missing.**

L226. ... in early versus late isolate ...

**Corrected. A “s” was removed.**

L409. add space after “and”

**Done. We added a space after “and”.**

L443. use OD600 or similar throughout the text

**Done. We added “600nm” as previously written in the manuscript.**

Table S8. Reference column. Replace “This sudy” with “This study”

**Done. We added a “t” to “sudy” in the actual **Supplementary Data 7**.**

Reviewer #2 (Remarks to the Author):

This is an interesting article that describes a bacterial adaptive evolution story on a case of chronic infection caused by *Yersinia enterocolitica*. The study postulates that the persistent infection in the patient was likely caused by vegetation that developed on the atrial lead of pacemaker post implantation, although this vegetation has not been detected until 2006. Four isolates of *Y. enterocolitica*, designated Ye.1-4, were obtained from the patient and their genomes were sequenced and combined with 259 additional *Y. enterocolitica* genomes of the same biotype of Ye.1-4 (*Y. enterocolitica* bioserotype 4/O:3). The results demonstrated that the multiple infection episodes were a result of chronic colonization rather than separate, independent infections. The authors assessed the susceptibility of the four isolates to 33 different antibiotics or antibiotic combinations and determined their growth rate. An extensive in-host evolution study of the bacterium was carried out using genomic, proteomic and transcriptomic approaches, resulting in identification of mutations in key genes that accounted for the observed antibiotic resistance and the abundant differences of proteins that could promote the adaptability within the host.

The data of this manuscript is well-presented and the results are logically organized. Nevertheless, I am a little concerned about the robustness of evidence supporting some of the authors' claims about mutations conferring resistance to specific antibiotic. Additionally, the effects of mutations on the activities of the proteins in question, such as DksA among others, requires further experimental validation and/or the provision of supporting references to substantiate these claims.

We thank the reviewer for these positive comments and the suggestions to strengthen our manuscript.

Major points:

1. It is regrettable that the isolate Ye.X, obtained in August 2001, was not collected and included in the genomics, proteomics and the phenotypic analysis. The absence of this isolate from the study impairs the comprehensiveness of the evolutionary trajectory of bacterium, particularly under condition that Ye.1 and Ye.2 are very close in isolation time point (only apart from less than 1 or 2 months; note that the isolation time of Ye. 2 is reported as January 2000 in the text, yet it is Feb 15, 2000, in Fig S3), and isolation time points for Ye.3 and Ye.4 are only 9 month apart. Ye.X, having evolved for 2 years from Ye.1 and Ye.2, could have served as a critical bridge connecting the endpoints of in-host evolution trajectory, potentially revealing a more complete picture of micro-evolutionary changes occurring within the host.

We fully agree with the reviewer that the study of the strain we named Ye.X would have been a great opportunity to investigate the evolutionary trajectory as a “midpoint” between early and late isolates. Unfortunately, the hospital team taking care of the patient at the time did not estimate necessary to keep this strain, it was not kept in a collection nor sent to the French *Yersinia* National Reference Centre, precluding any possible analysis.

We thank the reviewer for her/his accuracy regarding an error in the isolation date of Ye.2. The correct date is written now in **Supplementary Data 1**. We corrected the isolation date in the text (**line 88 and line 164**) as “February” “instead of “January”.

2. Except examining the antibiotic resistance characteristics, have the authors compared the virulence of the early and late isolates? The authors stated that ‘comparative proteomics of Ye isolates reveals an extensive metabolic rewiring, reduced virulence determinant expression and modulation of antibiotics resistance.’ The determination of virulence of the 4 isolates will provide evidences to corroborate those finding from omics studies.

Indeed, determining and comparing the virulence of the four isolates would have provided evidence to support the findings from the omics studies. Unfortunately, 3 of the 4 isolates lost their virulence plasmid as mentioned **line 289**. This plasmid was lost during subcultures in the laboratory and its loss correlates with loss of virulence (**Cornelis et al., 1998**), thus preventing the possibility of analyzing the virulence of the isolates in animal models, for example. We nevertheless could address cell invasion by gentamicin protection assay (**Figure 1 below**), and we demonstrate that Ye.4 displays a reduced invasivity in HeLa cells, probably due to Inv inactivation (as predicted from our genomic analyses). We now incorporated these results to our manuscript (**new Supplementary Fig. 11**).

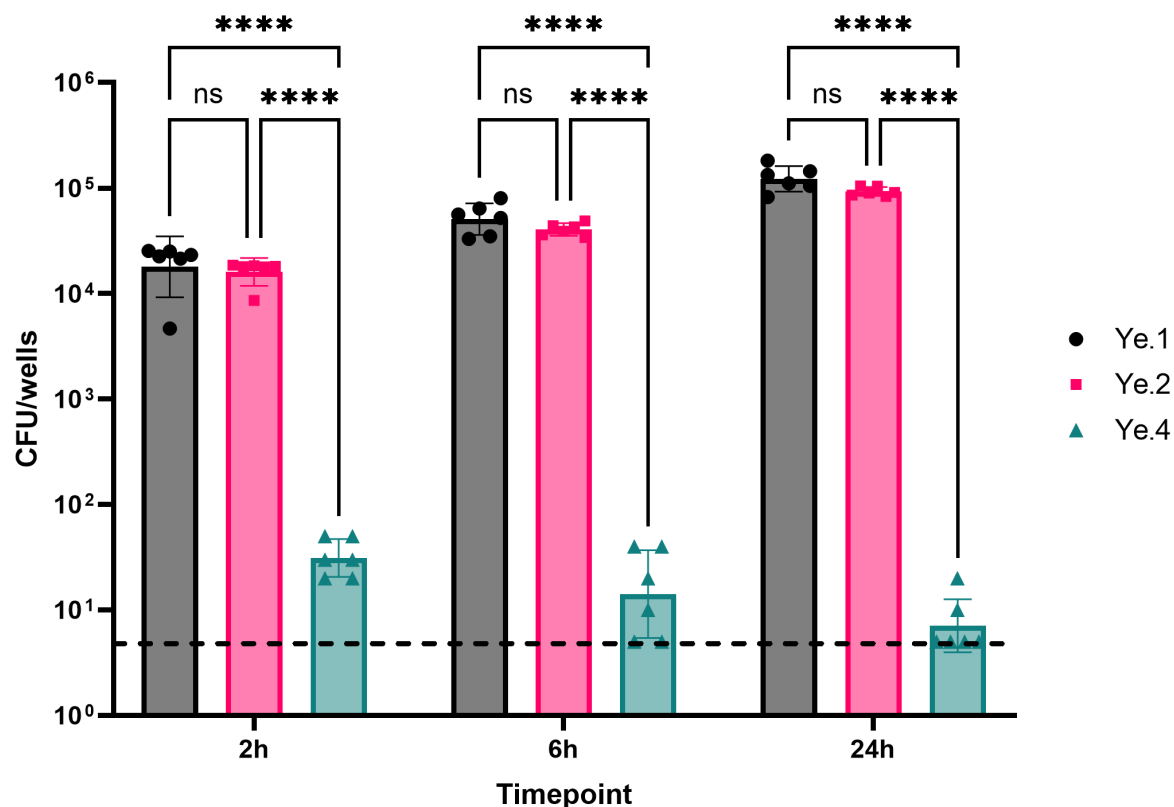


**Figure 1.** Cell invasion assay. CFU counts of intracellular Ye after contact with HeLa cells, at different timepoints. Statistical significance is calculated on log-transformed values by t-test corrected with Šidák correction. ns: nonsignificant. \*\*\*\*: p-value < 0.0001.

3. The growth rate of the late isolates is obviously slowed down compared with the two early isolates. Have the authors observed the differences between the mutation rates of early isolate and late isolates? Line 177-180, dose the least-squares dating tool used here has taken the variations of the isolates' growth rate into account, when estimating the occurrence of diverging into Ye.3 and Ye. 4?

We thank the reviewer for this excellent suggestion. To approach the question regarding the evolutionary rates between the early isolates and late isolates from the patient, we conducted a time-calibrated phylogenetic inference using a Bayesian approach implemented in BEAST v1.10.4 software. The approach uses the isolation dates from each recovered isolate and the variants between their genomes to calibrate the molecular clock and infer the evolutionary rates for the different branches in the samples' phylogeny. **Figure 2 (below)** displays the consensus phylogeny obtained from the time-calibrated phylogenetic analysis, which infers evolutionary rates for the branches leading to each of the four isolates of the order of  $10^{-5}$  and  $10^{-6}$  substitutions x site x year, comparable to the rates observed in other bacterial species (**Duchêne et al., 2016**) as well as in other in-host evolving bacterial strains (**Didelot et al., 2016**). We now incorporate these results to our manuscript (**lines 156-160** and **new Supplementary Fig. 3a**) and in the discussion (**lines 469-472**).

We note that the branches leading to the late isolates show 60-80% lower inferred evolutionary rates compared to the branches leading to the early isolates, suggesting that the slow growth rates of the later isolates may have been reflected into subsequently lower evolutionary rates. Although these results are based on a gold standard approach in time-calibrated phylogenetic

inference, we stress out that the sampling of only 4 isolates makes it difficult to assess the ability of molecular dating approaches to properly compare the rates between late and early isolates, especially since the early isolates are separated by only one month. A more complete sampling including intermediate isolates would have helped clarifying the question of the evolutionary rates between the different isolates.

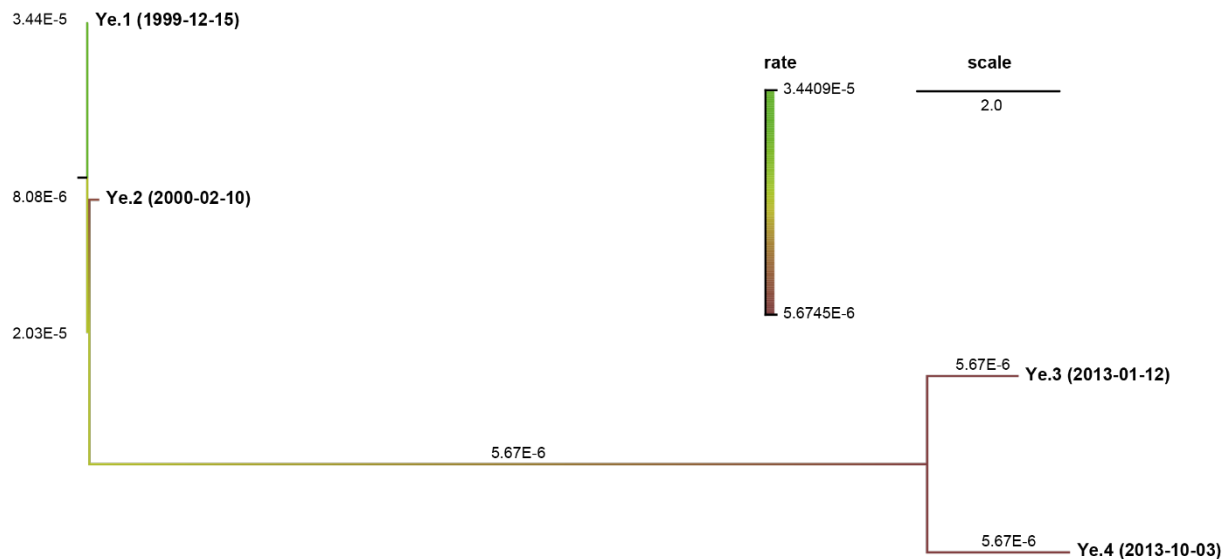


**Figure 2.** Maximum clade credibility tree of the 4 clinical isolates reconstructed from the time-calibrated phylogenetic inference using BEAST v1.10.4. Evolutionary rates in substitutions x site x year are annotated over each branch in the phylogeny. Branches are colored according to median rates and their value range is indicated in the legend on the top right.

Regarding the least-square dating approach (LSD, <https://github.com/tothuhien/lsd2>) used to infer the divergence dates for the clinical isolates from the study and closer *Y. enterocolitica* genomes, the framework does not incorporate complex data like phenotypic growth rates into their models. However, the approach does allow the different lineages to have different evolutionary rates, similar to relaxed molecular clock models in Bayesian phylogenetic inferences. This is because the expected branch distances (in time units) are calculated based on the branch lengths and the estimated rates for each branch (To et al., 2016). Thus, the approach allows evolutionary rates to vary between branches, and therefore the method can better fit the data when there is evidence of rate heterogeneity (which is common when comparing bacterial lineages).

4. The accumulated SNPs for the four Ye strains, including those on internal branches, appear be restricted to those within CDS-region (Lines180-181, Figure 3A, and TableS4-SNP-specific). According to the methods described, Snippy was used to call core-genome SNPs, which typically yields results including both CDS and intergenic regions. Please consider to include results from the intergenic regions in the manuscript for a more comprehensive analysis.

Thank you for raising this point. Indeed, Snippy identifies variants (SNPs and indels) regardless of their annotation (now mentioned in the Materials and Methods, lines 530-532). Variants in the intergenic regions are present in the analyses but their annotation was wrong. Thank you for your accuracy, we now corrected the annotations in **Supplementary Data 2**.

5. Similar to the SNP analysis, it appears that the analysis of small insertion and deletion

analysis excludes the intergenic regions, which may affect the conclusions drawn in Line 170. Additionally, the Methods section lacks some essential descriptions on indel analysis, for instance, the thresholds used in this study (e.g.,  $\leq 50$  bp, supported by at least a specified number of reads, and the allele frequency, defined as the number of reads supporting the mutation divided by the total reads at the site).

Thank you for raising this point. Indeed, we did not mention the thresholds used in this study. SNPs and indels were detected with a minimum coverage of 10X together with an allele frequency  $> 0.9$ . The size of the indels goes up to 50 bp. We have added this precision to the Methods section (**lines 530-532**).

As mentioned for the SNP analysis (see point 4 above), indels in the intergenic regions were not excluded from the analysis, they were identified regardless of their annotation (in CDS or intergenic regions). It turned out that we did not identify any indels in the intergenic regions from the 4 isolates.

6. The assertion that "Truncation of DksA is one of the main drivers of proteome changes" may not be fully substantiated. In Lines 267-268, the authors state that 'early vs. late isolate comparisons had a correlation coefficient ranging from -0.29 to -0.43 with comparisons of a Ye WT vs.  $\Delta dksA$  mutant.' However, in Figure 6 and Table S9, the late vs. late isolate comparison (Ye.3 vs. Ye.4, Number 6) also shows a correlation coefficient of -0.29 with the Ye WT vs.  $\Delta dksA$  mutant comparison (Number 51). If I understand correctly, truncation of DksA would be just one of the possible factors impacting proteome changes, rather than a primary driver. Since other truncated genes in Table 2 are not included in the 57 publicly available transcriptome datasets, the discussion appears to overemphasize DksA without addressing other potential contributing factors.

We thank the reviewer for her/his remark. First, we reanalyzed our data using a Spearman correlation coefficient  $r$  instead of Pearson correlation coefficient  $R$ , as multi-omics comparisons across experiments are not necessarily linearly correlated and we wanted to assess monotonic relationships. This yielded slightly higher  $r$  values than corresponding  $R$  values (**Source Data of Figure 6**). However, this did not change the fact that indeed, Ye.2 vs. Ye.3 and Ye.3 vs. Ye.4 displayed the same correlation concerning the WT vs.  $\Delta dksA$  comparison ( $r = -0.32$  and  $-0.33$  respectively). We thus decided to remove the claim that DksA is the main driver of proteome changes and modified our approach: we clustered the correlations (**Figure 3, below**), then highlighted that early vs late showed a good correlation with the "stringent response (SR) cluster". Late isolates thus have a perturbed stringent response which could be due to purine metabolism shutdown, or DksA or GppA truncation. We also highlighted that Ye.4 correlated better with the SR cluster, and that the Ye.3 vs Ye.4 comparison also correlated. Thus Ye.4 presumably has a more perturbed stringent response than Ye.3, possibly due to specific mutations in Ye.3 or Ye.4, such as OmpF truncation in Ye.4. We now incorporate these results to our manuscript (**lines 334-345 and new Figure 6**) and in the discussion (**lines 443-454**).

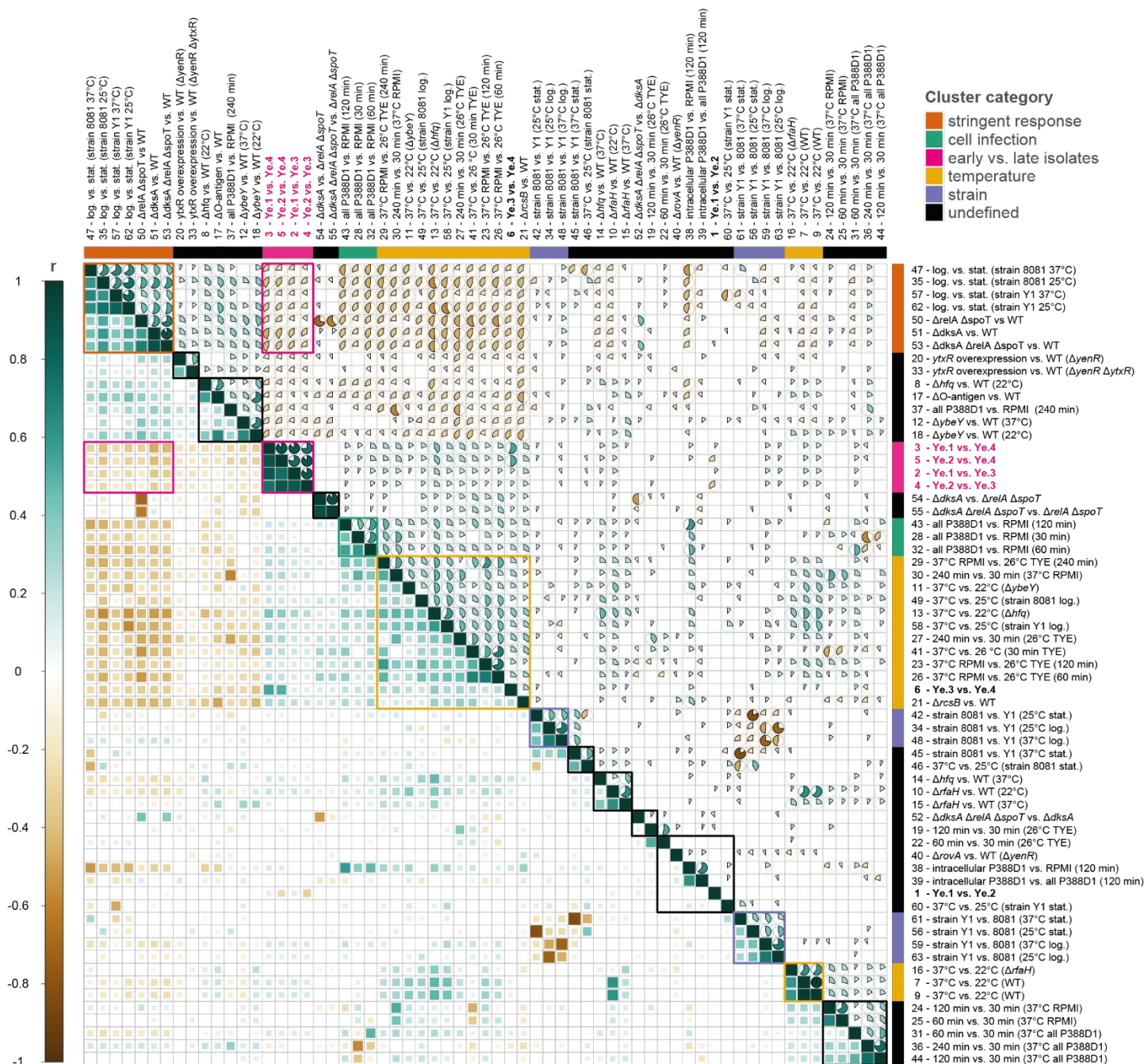


**Figure 3.** Heatmap of omics comparison correlations. Spearman correlation of 63 Ye omic experiment comparisons, clustered by hierarchical clustering and showing 15 clusters (squares). Color code of manually annotated clusters is shown on the top and right of the heatmap, and cluster squares are colored accordingly. Ye.1-Ye.4 isolate proteome comparisons are in bold, with early vs. late isolate comparisons colored in magenta and their correlations with the stringent response cluster highlighted by two magenta rectangles. Upper part of the heatmap is a pie representation (the correlation coefficient is represented by color and filling of the pie) and lower part is a square representation (the correlation coefficient is represented by color and size of the square). Only correlation coefficient with a p-value < 0.001 are shown. Coefficient color scale is shown on the left.

Minor points:

1. Line 142-144, it appears that the ciprofloxacin resistance of Ye strains depends heavily on the culture medium, and the increasing resistance to ciprofloxacin from Ye.1 to 4 can only be seen when they were grown in LB. Could the authors please potential explanations for this phenomenon, such as mutations in gene associated with carbon and nitrogen sources utilization and metabolism?

We thank the reviewer for this remark. Indeed, we only observed a ciprofloxacin MIC increase in liquid LB and not on MH agar + blood. However, LB is a non-standardized medium, unsuitable for MIC determination, and we cannot exclude interfering effects of components of the medium with the antibiotics. Thus, metabolic interpretation of this MIC would be risky and would need careful validation. We have added this information to the results (**lines 407-409**) and removed accordingly our metabolic interpretation from the discussion.

2. Lines 190-200, it is interesting that positive selection signals were found in genes encoding acyl-CoA thioesterase II and phosphate transporter. I propose, if feasible, the construction of mutant strains or proteins for these genes. Comparing the activity between the wild-type and the mutants would elucidate the biological relevance of these mutations. These results would make the conclusion more robust and convinced.

We thank the reviewer for this suggestion. To address this question, we replaced the phosphate transporter gene (*pitA*) in Ye.1 with the Ye.3 *pitA* allele. We tested intracellular survival in macrophages, as the gene is downregulated intracellularly, but could not observe any effect on survival. We tested the antibiotic resistance profile, and a slight decrease in the aminoglycoside MIC was observed, suggesting a role for the deficiency in the PitA transporter as described in **Zhao et al., 2022**. (see **new Table 2** in the revised manuscript). We incorporated these results to our manuscript (**lines 235-246**) and in the discussion (**lines 465-468**).

Similarly, we swapped the Ye.1 acyl-CoA thioesterase II (*tesB*) allele with the *tesB* allele from IP29610, the closest isolate we had in our collection. We tested the thioesterase II activity of both Ye.1 and IP29610 alleles swapped in both genetic backgrounds and could clearly observe that the Ye.1 allele had a decreased activity compared to the IP29610 allele, confirming that the mutation close to the catalytic triads and the dimer interface influenced the enzyme's activity. We also modeled the 3D structure of the proteins leading to the hypothesis that a hydrogen-bond involving the Ser 280 hydroxyl chain could interfere with the hydrogen network of TesB active site (**Figure 4** below). We now incorporated these results to our manuscript (**lines 247-270, new Fig. 4c, 4d and 4e, new Supplementary Fig. 6 and 7**) and in the discussion (**lines 455-465**).

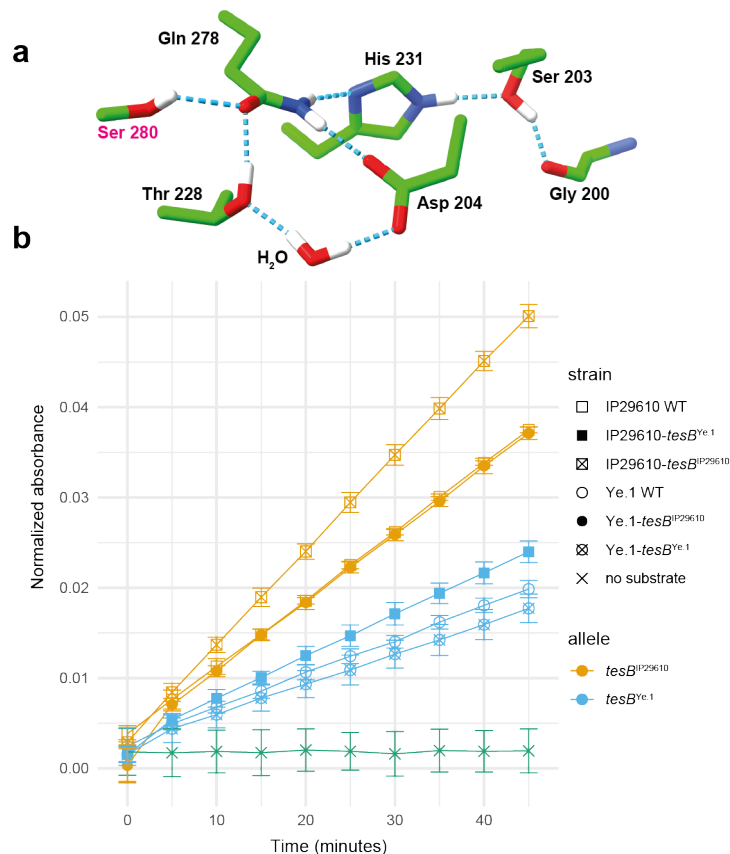


**Figure 4. a** Structure of the TesB active site of *E. coli* with a G280S mutation, annotated in magenta, showing an additional interfering hydrogen bond from the serine side chain. **b** Thioesterase II activity of IP29610 and Ye.1 mutant harboring different *tesB* alleles, measured by the CoA release from decanoyl-CoA. Each point shows the mean and standard deviation of 3 replicated wells in one experiment.

3. Line 93, the patient received only 18 months of ciprofloxacin therapy (up to July 2001), yet from Ye. 3 to Ye. 4, ciprofloxacin resistance increased significantly in the absence of ongoing ciprofloxacin selection pressure. Could the authors please propose a potential mechanism or driving force underlying this phenomenon.

To validate the implication of several mutations in the antibiotic resistance phenotypes, we introduce alleles of several Ye.3 and Ye.4 genes into the early strain Ye.1. Upon replacement of *ompF*<sup>Ye.1</sup> with *ompF*<sup>Ye.4</sup>, we observed a slight increase for the ciprofloxacin MIC. The Ye.4-specific *ompF* mutation could thus partially explain the increase in ciprofloxacin MIC observed for the late isolates, and the Ye.4-specific *ompF* mutation could synergistically increase Ye.4 ciprofloxacin MIC (**new Table 2**). The mechanism, whereas direct (decrease in ciprofloxacin influx) or indirect (i.e. metabolic rewiring due to *ompF* mutation) remains unknown. Ye.4 also harbors a second mutation in *gyrA* (D87N) involved in quinolone resistance and which could participate in the ciprofloxacin MIC increase. It remains however unknown why this mutation appeared in Ye.4 in the absence of ciprofloxacin selection pressure.

4. Line 134, the author stated that an unexpected AG MIC decrease was observed for the late isolates Ye.3 and Ye.4 compared to the early isolates Ye.1 and Ye.2, could they please make some explanation on this point.

We apologize for the lack of clarity in our formulation. Considering our new results with the different Ye.1 mutants we constructed, we decided to remove this sentence. We elaborate instead on the observed decrease in AG MIC which could be due to the accumulation of several individual mutations such as *dksA* and *pitA* (lines 230-233, 243-246 and 398-402)

5. Line 382, this suggests that the late slow-growing isolates accumulated less mutations and evolved even slower than expected, while most of the in-host evolving strains previously described showed increased mutation rate. The rationale behind this conclusion is not immediately clear. Could the authors please provide a detailed explanation? Additionally, is it possible that the slow-growing and longer doubling time is the main reason responsible for the lower mutation rate of the late isolates?

We thank the reviewer for this remark. This sentence resulted in a misinterpretation of the data from our part, as if an *mfd* mutation could reduce the mutation rate (contrary to the hypermutator phenotype sometimes found in in-host evolving strains). We decided to remove the sentence from the discussion. On the other hand, we calculated the evolutionary rates of the isolates and found rates concordant with in-host evolution (see our answer to major point 3 above, and lines 156-159 in the manuscript).

Concerning the slow-growing phenotype and low mutation rate, there are indeed described cases of some pathogens during persistent infections, with reduced replication or latency periods, where their low replication rate is thought to be translated into low mutation rate in the subsequent genome analysis of the isolated strains. This was shown for persister strains of Ebola virus (Keita *et al.*, 2021), in *M. tuberculosis* (Colangeli *et al.*, 2014) or in *S. aureus* (Young *et al.*, 2012) which appear to show steady rates in isolates through 30 years of persistent infection. We incorporated this information in the discussion (lines 469-475).

6. Lines 205-210, explaining the genetic basis of the resistance to quinolones in Ye.3 and Ye.4....., explain cefoxitin and ceftriaxone resistance. These conclusive statements need supporting from the experimental results. Currently, the authors provide only phenotypic resistance and comparative genomic analysis of the isolates. Without experimental validation specifically designed to test this variant, the conclusion cannot be directly drawn. Additionally, the analysis focused solely on coding regions, without examining potential mutations in the upstream intergenic regions of the gene. A similar issue arises with the claim regarding ompF-mediated cephalosporin resistance. It would be beneficial for the authors to cite relevant published literature, if available, that documents a direct relatedness between specific types of mutations and antibiotic resistance.

Regarding quinolone resistance, we identified a mutation in the quinolone resistance determining region (QRDR) of the *gyrA* gene. This mutation has not been identified in other studies because they usually consist of substitutions. Here it is based on a deletion of 3 nucleotides at 2 different codons, resulting in a single glycine instead of aspartic acid and serine, still in the QRDR. Since no other mutations are observed in the *gyrA* gene and since all known mutations are located in the QRDR of *gyrA* (Capilla *et al.*, 2004), we assumed that this new mutation is associated with quinolone resistance.

To validate this claim, we introduced the Ye.3 *gyrA* mutation in Ye.1 and we could indeed observe the resistance to nalidixic acid as well as a mild increase in ciprofloxacin and levofloxacin MIC. We now incorporate these results to our manuscript (lines 202-214 and new Table 2).

Regarding cefoxitin and ceftriaxone resistance, we cited two articles in the discussion to support our claim: “Abrogation of OmpF porin production is a widely shared strategy to reduce membrane permeability to antibiotics and was shown to be induced in *E. coli* upon exposure to ceftriaxone (Ching & Zaman, 2021). A study showing OmpF (YOMP-C in this previous work) mutations selected by cefoxitin treatment in Ye and leading to increased  $\beta$ -lactam resistance supports the antibiotic import via this porin (Brzostek & Hrebenda, 1988)” (lines 384-388). To validate this, we constructed an Ye.1 mutant harboring the SNP present in Ye.4 *ompF*. This mutant showed a strong increase in cefoxitin MIC (cefoxitin being known as the most susceptible  $\beta$ -lactam to membrane permeation) but only a mild increase in ceftriaxone MIC (lines 215-225 and new Table 2). We reasoned that the synergy between *ompF* mutation and the  $\beta$ -lactamase upregulation was responsible for Ye.4 ceftriaxone resistance. We incorporated these results to our manuscript and rewrote the discussion accordingly (lines 383-393). We looked at all the other impacting mutations inside and outside of genes specific to Ye.4 (considering mutations which could be in 5’ or 3’ end of a transcript) but could not find any genes related to  $\beta$ -lactam resistance, and recreating all the mutations is out of the scope of our manuscript.

7. Line 227, please explain ‘the loss of the pYV virulence plasmid is a known laboratory artifact’. How did the authors confirm that pYV was present in the Ye isolates when they were initially isolated? Additionally, were the CDS on pYV absent from the genome sequencing data?

All pathogenic *Y. enterocolitica* carry the virulence plasmid (pYV) encoding a Type 3 Secretion System (T3SS) that enables them to cause disease. Unfortunately, this plasmid is easily lost during subculture in vitro and this loss correlates with a loss of virulence (Cornelis et al, 1998). We can assume that the pYV was present in all 4 isolates, but that it was lost in the laboratory in the last 3 strains. Otherwise, the isolates would not have been able to cause disease in the host. In addition, we confirmed that the CDSs of the pYV are not present in the 3 genomes lacking the pYV.

8. Figure S2, the legend is confusing. While the blue circles in ‘a Ye.1’ represent the inhibition radius for Ye 1, how do green circles represent the inhibition radius for Ye 2 in the same panel? The same question is for ‘b Ye.2’. Why the assay results for Ye. 1 appeared in plate for Ye.2. Please provide some additional description to make it clear.

We apologize for not making this clearer. The circles were intended to allow comparisons between the different strains. We now moved the images to the source data files and removed the annotations.

9. Line 127-128, Figure S3 is for Ye. 1 and Ye. 2, thus Fig S3-4 in braces should be Fig S3, please correct.

Thank you for pointing this out. We moved the images to the source data files and now only refer to Table 1.

10. Line 718, legend for Fig. 2, “a,f Ye.1” should be “a,e Ye.1”, please correct.

Thank you for pointing this out, it has been corrected.

11. Figure 3b, how were the strains selected? Were they chosen based on the shortest genetic distance to the four Ye strains? Additionally, were any temporal signals evident before estimating the date, possibly using tools like TempEst?

The strains included in this analysis (**Figure 3b** of the **manuscript**) were chosen based on the shortest genetic distance to the four Ye strains based on the phylogenetic tree (**Figure 3a** of the **manuscript**).

Treetime was used to evaluate the presence of a temporal signal in the 8 strains. Due to the accumulation of many SNPs (reflected by the long branch leading to Ye.3 and Ye.4), Ye.3 and Ye.4 were excluded from the analysis because they represent outliers (in red). As reflected by the root-to-tip regression (**Figure 5 below**), we can assume the presence of a temporal signal with a  $R^2=0.96$ . Methods and results section were modified according to this new analysis (**lines 166-171** and **new Supplementary Fig. 3b**).

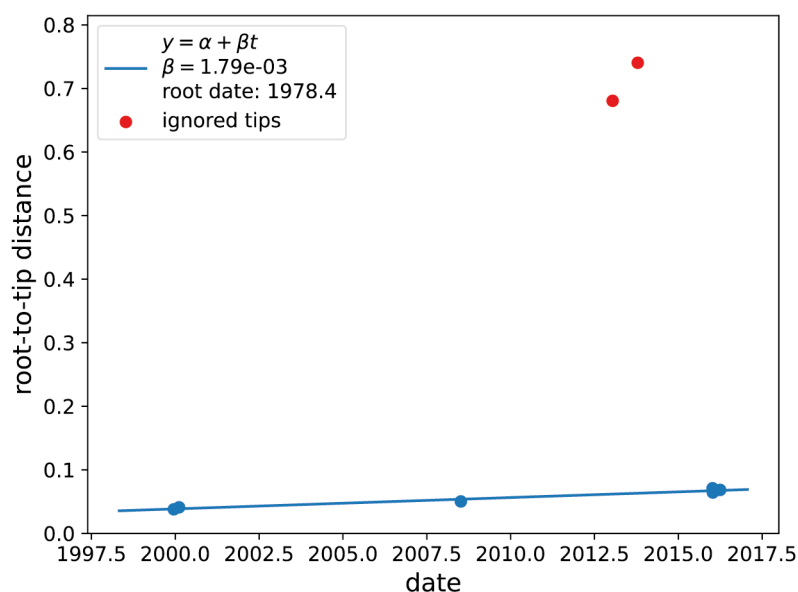


**Figure 5.** Root-to-tip regression generated with Treetime. Detected outliers (Ye.3 and Ye.4) are marked in red.  $R^2=0.96$

12. The title of Table 3, Log2(foldchange) of the proteome pairwise comparisons for selected proteins. I suggest to provide the selecting criteria for the proteins shown in this table.

Thank you for this suggestion. We now show these data in the heatmap **Figure 5** and moved the table to the source data. We added in the heatmap legend “Proteins are grouped in categories based on Y11 RefSeq and Uniprot annotations, and bibliography”.

13. In the tables and supplementary tables, decimal points are consistently shown as commas.

Thank you for pointing this out, we corrected all the tables.

Reviewer #3 (Remarks to the Author):

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

Reviewer #4 (Remarks to the Author):

This is an interesting paper on the evolution of a commensal/pathogen *Yersinia enterocolitica* during 14 years within a patient with bacteremia. This research is important from bio-medical (antibiotic resistance, development of antibiotic therapies, host-pathogen interactions, pathogen persistence evolution) and fundamental prospective (e. g. pangenome evolution). This manuscript is interesting for publication, however, requires some clarifications and more context. Below my comments.

We thank the reviewer for this positive comment.

Comments:

1. Reported pangenome consists of 3,824 genes (line 464), however genomes of *Yersinia enterocolitica* appear to have around 3,900-4,000 genes or more, so pangenome of 4 strains has to be at least 3900 genes. Would it be possible to clarify following questions:

- Provide assembly statistics values like N50 for strains sequenced in this study, genome completeness and contamination obtained for example by checkM.

The statistics of the genome assemblies of the 263 isolates were determined using CheckM version 1.1.3. A new **Supplementary Data 8** was added to the manuscript and the corresponding Method section (**lines 528-529**) completed this new analysis.

- What are the thresholds for gene clustering in pangenome building?

The pangenome was built using PPanGGOLiN using 80% coverage and 80% amino acid sequence identity for gene clustering. The method (**lines 540-541**) section was updated with these thresholds.

Perhaps *Yersinia* has a lot of paralogs which were combined together, or a lot of genes were lost during evolution during bacteremia and pangenome became smaller than average genome of *Yersinia*. Perhaps there are other scenarios/explanations?

Thank you for your precision as there was an inaccuracy in the Methods section regarding the size of the pangenome. The pangenome was identified on a subset of 100 genomes and consists of 5,332 genes, rather than the previously mentioned 3,824 genes—which was indeed incorrect, as the pangenome size cannot be smaller than the number of genes within a single genome. From these genes, a core genome of the 4 isolates of interest was built and consists of 3,829 genes (not 3,824). We updated the manuscript in the Methods section (**lines 539-550**). We believe that this explanation and correction will bring some lights in the understanding in the number of genes of the core genome.

- Did core genome size decline during the 14 years of evolution?

Except for the 3 deletions identified in the late isolates (45.1 kb), we did not identify any other reduction in the core genome of the late isolates compared to the early isolates.

2. It may be somewhat unexpected that strains from this study are on much longer branches in the tree (Figure 3). How pure was the culture, could there be within species divergence in the cultures which were sequenced? Is mutation rate higher than expected? Does the tree have an outgroup from a different species? How was the tree rooted? (Is it rooted?)

We respectfully disagree with the reviewer. When we place the persistent isolates into the phylogeny, we compare their evolution to that of related *Y. enterocolitica* strain within diverse, long-term circulating populations. We argue that the observed longer branches reflect the accumulation of private/exclusive variants within the persistent strain because of the strong genetic bottleneck in the clone that persisted in the chronic infection, which is a typical signature of in-host evolution. The accumulation of variants in chronically infecting strain is expected, as bottlenecks exacerbate the effect of genetic drift so that variants present in the persisting clone are quickly fixed (**Didelot et al. 2016**). This may be a factor in the seemingly higher rate of short-term evolution observed in within-host populations, as some mutations that persist over short time scales would be purged over longer time scales (i.e.: purged in the strains within the general population of the pathogen).

Regarding isolates from our collection, purity of the culture was assessed before phenotypic characterization (API20E and 50CH strips, lipase and pyrazinamidase activity, and seroagglutination). Then, DNA extraction and genome sequencing were performed. Cultures were pure before genome sequencing. Isolates were subcultured only twice before sequencing (one for the characterization and conservation; and the other one from the conservation to DNA extraction). We do not believe that species divergence occurred with only 2 subcultures.

The tree was rooted at midpoint without an outgroup from another species or biotype. Otherwise, the branches of the biotype 4 isolates would have been artificially compressed as we would have a long branch for the outgroup and grouped leaves corresponding to the biotype 4 isolates.

3. If the genome of these strains is significantly reduced what are the origins of the closest public strains? Are there other bacteremia/antibiotic treatment strains or extra-host environment? I think it would be beneficial to plot metadata from Table S3 and bootstrap support values on the tree.

Only the genomes of the 2 latest isolates (Ye.3 and Ye.4) are slightly reduced (3 large deletions – total of 45.1 kb).

Thank you for your suggestion to improve the **Figure 3a** of the article. We plot the metadata of the **Supplementary Data 1** in the **Figure 3a** of the article together with bootstrap support values.

4. To which fraction of the pangenome do genes belong that change their expressions in proteomics (and/or public transcriptomic data)? Are there more core genes which change expression or gain deleterious mutations?

We selected all the genes for which protein abundances were significantly different in one of the isolate comparisons, resulting in 1,456 genes. Among them, 1,386 genes (95%) are present in 95 strains out of the 100 used for the pangenome calculation (that is the 95% pangenome).

More precisely, we defined the following category:

Present in  $\geq 95\%$  of the strains : core

Present in  $\geq 15\%$  and  $< 95\%$  of the strains: shell

Present in  $< 15\%$  of the strains: cloud

When looking at individual comparisons, differentially abundant proteins and corresponding genes are subdivided as followed:

| condition | core-genome | shell-genome | cloud-genome | total |
|-----------|-------------|--------------|--------------|-------|
| Ye.1>Ye.2 | 136         | 36           | 1            | 173   |
| Ye.1<Ye.2 | 99          | 4            | 1            | 104   |
| Ye.1>Ye.3 | 452         | 47           | 1            | 500   |
| Ye.1<Ye.3 | 338         | 6            | 1            | 345   |
| Ye.1>Ye.4 | 620         | 51           | 2            | 673   |
| Ye.1<Ye.4 | 389         | 8            | 0            | 397   |
| Ye.2>Ye.3 | 453         | 20           | 2            | 475   |
| Ye.2<Ye.3 | 382         | 9            | 1            | 392   |
| Ye.2>Ye.4 | 597         | 20           | 2            | 619   |
| Ye.2<Ye.4 | 424         | 8            | 0            | 432   |
| Ye.3>Ye.4 | 215         | 8            | 1            | 224   |
| Ye.3<Ye.4 | 98          | 1            | 0            | 99    |

We conclude that most of the differentially expressed proteins and their associated genes belong to the core genome.

When looking at the ratio of differentially abundant proteins in the pan-genome fractions compared to total differentially abundant proteins in the comparison, we obtained the following results:

| condition | core-genome | shell-genome | cloud-genome | total  |
|-----------|-------------|--------------|--------------|--------|
| Ye.1>Ye.2 | 78.6%       | 20.8%        | 0.6%         | 100.0% |
| Ye.1<Ye.2 | 95.2%       | 3.8%         | 1.0%         | 100.0% |
| Ye.1>Ye.3 | 90.4%       | 9.4%         | 0.2%         | 100.0% |
| Ye.1<Ye.3 | 98.0%       | 1.7%         | 0.3%         | 100.0% |
| Ye.1>Ye.4 | 92.1%       | 7.6%         | 0.3%         | 100.0% |
| Ye.1<Ye.4 | 98.0%       | 2.0%         | 0.0%         | 100.0% |
| Ye.2>Ye.3 | 95.4%       | 4.2%         | 0.4%         | 100.0% |
| Ye.2<Ye.3 | 97.4%       | 2.3%         | 0.3%         | 100.0% |
| Ye.2>Ye.4 | 96.4%       | 3.2%         | 0.3%         | 100.0% |
| Ye.2<Ye.4 | 98.1%       | 1.9%         | 0.0%         | 100.0% |
| Ye.3>Ye.4 | 96.0%       | 3.6%         | 0.4%         | 100.0% |
| Ye.3<Ye.4 | 99.0%       | 1.0%         | 0.0%         | 100.0% |

When calculating the ratio of differentially abundant proteins in the pan-genome fractions compared to total identified proteins in these pan-genome fractions, we obtained the following results:

| condition | core-genome | shell-genome | cloud-genome | total |
|-----------|-------------|--------------|--------------|-------|
| Ye.1>Ye.2 | 5.9%        | 44.4%        | 33.3%        | 7.2%  |
| Ye.1<Ye.2 | 4.3%        | 4.9%         | 33.3%        | 4.3%  |
| Ye.1>Ye.3 | 19.4%       | 58.0%        | 33.3%        | 20.8% |
| Ye.1<Ye.3 | 14.5%       | 7.4%         | 33.3%        | 14.3% |

|           |       |       |       |       |
|-----------|-------|-------|-------|-------|
| Ye.1>Ye.4 | 26.7% | 63.0% | 66.7% | 27.9% |
| Ye.1<Ye.4 | 16.7% | 9.9%  | 0.0%  | 16.5% |
| Ye.2>Ye.3 | 19.5% | 24.7% | 66.7% | 19.7% |
| Ye.2<Ye.3 | 16.4% | 11.1% | 33.3% | 16.3% |
| Ye.2>Ye.4 | 25.7% | 24.7% | 66.7% | 25.7% |
| Ye.2<Ye.4 | 18.2% | 9.9%  | 0.0%  | 17.9% |
| Ye.3>Ye.4 | 9.3%  | 9.9%  | 33.3% | 9.3%  |
| Ye.3<Ye.4 | 4.2%  | 1.2%  | 0.0%  | 4.1%  |

For the accessory genome (shell- and cloud-genome), we observed that the proportion of less abundant proteins was greater in the latest isolates than in the earliest isolate in each comparison. This suggests a global downregulation of the accessory genomes over time during chronic infection. We added these results to our manuscript (**lines 279-283, lines 469-470 and Supplementary Data 5**)

Regarding the 74 truncated genes and genes in deleted regions, 64 belonged to Ye core-genome while 10 belonged to the remaining accessory genome (6 in shell-genome and 1 in cloud-genome) which correspond approximately to the repartition of Ye.1 to Ye.4 pan-genome fractions. We added these results to the manuscript (**lines 184-186 and Supplementary Data 4**).

5. Are there any new genes that were acquired in later strains (e.g Ye4) but were absent in Ye1? We did not identify any gene acquisition common to Ye.3/Ye.4 versus Ye.1/Ye.2.

6. Genomes are deposited to ENA but associated metadata about those genomes (e.g. timepoints of sampling) does not seem to appear on the site.

Associated metadata were updated on ENA (see below for instance)

Sample: ERS1580350

```

1  <SAMPLE_SET>
2  <SAMPLE accession="ERS1580350" alias="Ye.1" center_name="IP">
3  <IDENTIFIERS>
4  <PRIMARY_ID>ERS1580350</PRIMARY_ID>
5  <SUBMITTER_ID namespace="IP">Ye.1</SUBMITTER_ID>
6  </IDENTIFIERS>
7  <TITLE>Within-host evolution of Y. enterocolitica during a chronic human infection</TITLE>
8  <SAMPLE_NAME>
9  <TAXON_ID>630</TAXON_ID>
10 <SCIENTIFIC_NAME>Yersinia enterocolitica</SCIENTIFIC_NAME>
11 </SAMPLE_NAME>
12 <SAMPLE_ATTRIBUTES>
13 <SAMPLE_ATTRIBUTE>
14 <TAG>isolation_source</TAG>
15 <VALUE>blood</VALUE>
16 </SAMPLE_ATTRIBUTE>
17 <SAMPLE_ATTRIBUTE>
18 <TAG>lat_lon</TAG>
19 </SAMPLE_ATTRIBUTE>
20 <SAMPLE_ATTRIBUTE>
21 <TAG>collection_date</TAG>
22 <VALUE>1999-12-15</VALUE>
23 </SAMPLE_ATTRIBUTE>
24 <SAMPLE_ATTRIBUTE>
25 <TAG>geographic location (country and/or sea)</TAG>
26 <VALUE>France</VALUE>

```

## References

Brzostek, K. & Hrebenda, J. Outer-membrane permeability to  $\beta$ -lactam antibiotics in *Yersinia enterocolitica*. J Gen Microbiol 134, 1535–1540 (1988).

Capilla, S. et al. Characterization of the molecular mechanisms of quinolone resistance in *Yersinia enterocolitica* O:3 clinical isolates. J Antimicrob Chemother 53, 1068–1071 (2004).

Ching, C. & Zaman, M. H. Identification of multiple low-level resistance determinants and coselection of motility impairment upon sub-MIC ceftriaxone exposure in *Escherichia coli*. *mSphere* 6, (2021).

Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, et al. (2014) Whole Genome Sequencing of *Mycobacterium tuberculosis* Reveals Slow Growth and Low Mutation Rates during Latent Infections in Humans. *PLOS ONE* 9(3): e91024

Cornelis GR, Boland A, Boyd AP, Geuijen C, Iriarte M, Neyt C, Sory M, Stainier I. 1998. The Virulence Plasmid of *Yersinia*, an Antihost Genome. *Microbiol Mol Biol Rev* 62: .<https://doi.org/10.1128/mmb.62.4.1315-1352.1998>

Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol*. 2016 Mar;14(3):150-62. doi: 10.1038/nrmicro.2015.13. Epub 2016 Jan 19. PMID: 26806595; PMCID: PMC5053366.

Duchêne S, Holt KE, Weill FX, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. Genome-scale rates of evolutionary change in bacteria. *Microb Genom*. 2016 Nov 30;2(11):e000094. doi: 10.1099/mgen.0.000094. PMID: 28348834; PMCID: PMC5320706.

Keita, A.K., Koundouno, F.R., Faye, M. et al. Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. *Nature* 597, 539–543 (2021).

To TH, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. *Syst Biol*. 2016 Jan;65(1):82-97. doi: 10.1093/sysbio/syv068. Epub 2015 Sep 30. PMID: 26424727; PMCID: PMC4678253.

B.C. Young, T. Golubchik, E.M. Batty, R. Fung, H. Larner-Svensson, A.A. Votintseva, R.R. Miller, H. Godwin, K. Knox, R.G. Everitt, Z. Iqbal, A.J. Rimmer, M. Cule, C.L.C. Ip, X. Didelot, R.M. Harding, P. Donnelly, T.E. Peto, D.W. Crook, R. Bowden, & D.J. Wilson, Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease, *Proc. Natl. Acad. Sci. U.S.A.* 109 (12) 4550-4555, <https://doi.org/10.1073/pnas.1113219109> (2012).

Zhao, X. et al. *Pseudomonas aeruginosa* phosphate transporter PitA (PA4292) controls susceptibility to aminoglycoside antibiotics by regulating the proton motive force. *Antimicrob Agents Chemother* 66, (2022).

\*\* See Nature Portfolio's author and referees' website at [www.nature.com/authors](https://www.nature.com/authors) for information about policies, services and author benefits.

## REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

The authors supplemented additional experiments and conducted detailed structural analyses of protein mutants, which effectively addressed the concerns I raised. The revised manuscript now exhibits a more rigorous presentation, with data that better substantiate the research conclusions.

We thank the reviewer for this positive comment.

However, a few minor issues remain and should be addressed as follows:

Line 169, "Genomic analysis supports a chronic Ye infection" is not an appropriate title for the content that followed.

Indeed, this title was not appropriate, and we have changed it to "Genomic evolution during the chronic Ye infection" to better reflect the following content.

Line 190, "..... indicating that evolution had already started within the first months of infection". Evolution is going on all the time. Please consider to revise this sentence.

Indeed, evolution is going all the time. We have changed the sentence by "indicating that detectable genetic divergence occurred within the first months of infection"

Line 278-279, "affecting two cytoplasmic, one 278 non cytoplasmic and three transmembrane domains". This description is not appropriate, please correct. They are not three domains, but different regions in one domain.

We thank the reviewer for this accurate precision, we replace "domains" by "regions" in the manuscript and associated figure 4a.

Line 306-309, "....., we assessed cell invasivity which is independent of the type 3 secretion system encoded by pYV genes" is not correct. Several Yop effectors, like YopE, YopT and YpkA, show enzymatic activity to proteins that are involved in cytoskeletal remodeling, which will influence the invasion capability of the bacteria.

We respectfully disagree with the reviewer. The T3SS is indeed known to impact cytoskeletal remodeling and to prevent phagocytosis primarily by macrophages and neutrophils (Marketon et al. 2005, Demeure et al. 2019). However, in epithelial cells, invasion has been shown to be independent of the T3SS, using pYV- strains (Moreau et al. 2010, Ligeon et al. 2014). Indeed, invasion of epithelial cells is dependent on the surface molecule invasion (Isberg & Falkow 1985, Isberg et al. 1987). Since our Ye.2 and Ye.3 strains do not harbor the pYV- (and therefore do not produce a T3SS), we cured the Ye.1 strain from the pYV plasmid to perform experiments in a similar T3SS- background (which reflects the invasion activity dependent on invasin)

Demeure *et al.* 2019. *Genes Immun.* 20(5):357-370  
Isberg *et al.*, 1985 *Nature* 317(6034):262-4  
Isberg *et al.*, 1987 *Cell* 50(5):769-78  
Marketon *et al.*, 2005 *Science* 309(5741):1739-41  
Moreau K *et al.* 2010. *Cell Microbiol* 12: 1108-23  
Ligeon LA *et al.* 2014. *Autophagy* 10: 1588-602

Supplementary Fig. 11, Cell invasion assay, according to the description of “Gentamycin protection assay” in the Supplementary methods, it appears to be determining the living bacterial inside the cells but not the bacteria that invade the host cells. Ye.1 and Ye.2 exhibit much higher ability to replicate inside macrophages. Please consider to correct the relevant descriptions.

Our gentamicin invasion assay was performed in epithelial cells (HeLa) and not in macrophages. We assessed intracellular bacterial viability after 1 hour of contact and 1 hour of gentamicin treatment (2 hours total). As we observe a 3 log difference in CFU counts between Ye.1/Ye.2 and Ye.4 at an early time point (only 2 hours), we can reasonable propose that this difference reflects rather a difference in invasion (and not in replication), as the doubling time of the bacteria is around 50 min for Ye.1/Ye.2, allowing roughly 2.5x doubling time in 2 hours.

However, comparison between 2, 6 and 24 hours post-infection indeed show an increase in CFU counts for Ye.1/Ye.2 and a decrease in CFU counts for Ye.4, showing a difference in the bacterial ability to replicate within epithelial cells. We have added this information to the manuscript (lines 311-312 of the revised manuscript) »

Reviewer #3 (Remarks to the Author):

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

We thank the reviewer for taking the time to evaluate our manuscript.

Reviewer #4 (Remarks to the Author):

Authors have addressed my questions and suggestions. I would like to recommend the manuscript for publication.

We thank the reviewer for this very positive evaluation of our manuscript.