

Lineage-dependent variations in single-cell antibiotic susceptibility reveal the selective inheritance of phenotypic resistance in bacteria

Corresponding Author: Dr Minsu Kim

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

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Please see attachment.

(Remarks on code availability)
The code was not provided, but I think it should be.

Reviewer #2

(Remarks to the Author)

The paper titled "Lineage-dependent variations in single-cell antibiotic susceptibility reveal the selective inheritance of phenotypic resistance within isogenic bacterial populations" by Stine et al, presents an experimental study of antibiotic effect on the survival of cells along lineages. By tracking cells genealogical trees and carrying out statistical analysis of the cells' survival patterns along those trees, the authors were able to show that the probability of a cell to survive increases as more of its predecessors survive. In other words, cells that survive the exposure to antibiotic pass on this ability to their offspring, which then increases the chances of these offspring cells to survive that exposure as well. The authors argue that this is contradictory to previous theories that proposed that survival of bacteria upon exposure to antibiotics is due to random process that allow some cells to adapt to the selective pressure independent of the cells' history. The authors propose a model to describe their findings, in which cells can switch between healthy and vulnerable states, and that the healthy state can be inherited by the offspring cells and therefore increasing the chances of cells, whose ancestors survived, to survive as well. On the other hand, the vulnerable state can switch to a third state (death), from which there is no recovery. The rates of transitions between the different states are determined from fitting the model predictions to the measured correlations between the different types of cell pairs with various degrees of relationships.

The study is very interesting and can be suitable for this journal, but I have a few important issues that I think should be addressed before I can recommend the manuscript for publication.

Major concerns:

- 1) How is the age of the cell determined? The authors state in "cell tracking" that the labeling of cells is based on the old pole age, but they have no way of knowing the age of the oldest poles in the sequence, and the two cells labeled "0" and "00" in supplementary figure 6 is not the same. Is that important? And if not why? If that is not what the authors mean by age, then they might consider using a different term or explain it better to avoid confusion.
- 2) How was the breakpoint concentration for the antibiotic determined? Was it determined in liquid culture or in agar? And if in one and not the other, isn't the effect of antibiotics different for the growth in different conditions?
- 3) In the last paragraph of the results section (lines 264 – 274) the authors discuss the effect of tol-C expression on the correlations between siblings. I would note here that, a previous study (cited here as well, Bergmiller et al., Science 356,311–315 (2017)) has argued that since the accumulation of these pumps is at the poles, it results in asymmetry between siblings and therefore phenotypic heterogeneity in the response to antibiotics. This could be why delta tol-C strain exhibits

higher correlation between siblings. The question then is, since the ToIC accumulation in the cell is correlated with the cell's age, could its effect be compared between pairs of cells whose oldest pole is of the same age vs. pairs of cells in which the oldest poles are of different age?

4) In lines 285 – 287, the authors state that “This robustness further strengthened across the generations, indicating phenotypic adaptation to antibiotics (Fig. 3). This positive feedback was due to the inheritance of resistance factors whereas antibiotic-induced damage remained constant.” Isn't this statement contradictory to fact that Delta toIC exhibits higher correlations between generations? Or is the correlation in the delta toIC case is higher due to the inheritance of the vulnerable state?

5) In the supplementary movie, cells that survive and continue to proliferate appear to grow over cells that perished. Do the authors think that this can have an effect on the survival chances of cells? For example, the areas where cells have grown and died before could be depleted of antibiotics, or the debris of dead cells could provide resources to the surviving cells? Is there a way to assess this effect or explain why it should not be considered?

6) Can the authors add a supplementary figure showing how the data was fitted by the model to extract the transition rates in Fig. 4b?

Minor concerns:

- 1) Explain breakpoint concentration of antibiotic better and describe how it was determined experimentally.
- 2) Describe the mode of action of the antibiotic used.
- 3) Describe how exactly the frequency of consecutive generations exhibiting robustness expected by random chance is calculated and what is the expected value.
- 4) Define the relative difference in frequency. Is it $(\Delta F)/F_0$?

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Please see attachment.

(Remarks on code availability)

The authors state that the code is uploaded but I do not see where.

Reviewer #2

(Remarks to the Author)

The authors have addressed all my previous concerns satisfactorily.

I only have few minor comments for the authors considerations before publishing.

1) The breakpoint is still not clear to the wide readership from the explanation provided in the manuscript. I think it would be beneficial if the authors could add a concise description of the physiological state of the bacteria at the breakpoint concentration. Similar to the MIC, which is the minimal concentration with no observed culture growth.

2) In line 49 of the introduction, the authors review mechanisms of antibiotic resistance, but do not mention epigenetic adaptation, which is different from what they study in this manuscript. I think this should be added here. See for example: Adam et al, BMC evolutionary biology 2008 (DOI: 10.1186/1471-2148-8-52).

3) In the paragraph from line 175 to 181, I was a bit at loss. With $P(D=2/R_n=1)$, do the authors mean the probability of two daughter cells exhibiting robust survival conditioned upon one prior generation of robustness, or do they mean regular survival? From the sentence, line 177 – 179, I understood that it's regular survival, which does not make sense because it is the requirement for the previous generation to be considered robust, but by the end of the paragraph I gathered that they mean robust survival. I think it should be made sharper.

4) Typo in line 242, “... deltatoIC strain that WT...” should be “...deltatoIC strain than WT...”.

5) In line 284, the authors state that the expression of ToIC is stochastic. What does that mean? Aren't all genes expressed stochastically? Do they mean that it's amount vary among cells, which is the case for all proteins? Or do they mean something else? If that's what they mean, I don't think it should be stated as if it is a unique property of this gene.

6) Related to the previous point, in line 313 – 316, the authors state that experiments using a promoter that would enable precise control of efflux pump expression pattern, would be useful in the future. Is that even possible? don't all genes exhibit similar variability patterns but different average levels? See for example: Taniguchi et al Science 2010 (PMID: 20671182).

7) In line 300 – 301, the authors claim that the robustness is strengthened across the generations is an indication of phenotypic adaptation to antibiotics. To me it seems that this could be random but inherited, and over time it is manifested more. In the model they present also, there is no adaptation parameter, so I don't understand why they attribute this to adaptation. If I'm missing something, this point should then be better explained, or rephrased.

(Remarks on code availability)

There was no change in the code itself in the revision, but I did review the description of the code in the supplementary text and found it to be good and sufficient.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Please see attachment.

(Remarks on code availability)

I have confirmed that the authors provided their code and a file called 'Code Index.rtf' describing the functions included. However, I have not tested their code: it is written in the proprietary software Matlab for which I do not currently have a license.

Reviewer #3

(Remarks to the Author)

I have specifically been asked to comment on the comments of reviewer #1 to suggest which points are necessary for the authors to change.

Data availability:

- I agree with the reviewer that an explanatory file for columns and abbreviations, and red text/shading is needed to meet reproducibility criteria. I do see some definitions in the methods section, but it would help to have these together with the data.
- I do not think that formulae need to be added to source data 1. Most if not all of the calculated quantities are standard or provided in the methods.
- Overall while I think there is room for improvement here, I think this looks to be up to the usual standards of data availability, if a description of the headers/fields is added.
- I would encourage the users to consider whether there is some raw/underlying data from their experiments they could share. The correlations are useful, but computed from other data. To maximise reusability, the data from which these were computed could be made available.

Code:

Available, in MATLAB which requires a license, but this is acceptable.

Mathematical model:

The reviewer's suggestions are helpful and seek to improve the clarity of the work with specific suggestions on how to do so. It would be useful for the authors to respond to these suggestions and incorporate them where possible. These all appear to be minor changes.

Statistical analysis:

My overall summary here is that this debate is going too far down the dogmatic frequentist route in a manner which is becoming unhelpful for the paper. In most of this debate we are only comparing three data points as far as I can tell -- a formal statistical test has very little power here. We can make some simpler and more consistent choices to tidy this all up.

- The reporting summary does indeed state that the test statistic is required. Personally I do not feel this is necessary here. This point is up to the editor in my view.

- Specific frequentist statistical approach.

- * Rather than using t-tests on C_B , just use the p-value of the slope of the regression these coefficients are calculated from. This can easily be done using e.g. `lm()` in the R programming language, or many online tools.

- * Figure 3: Firstly, this caption is far too long and should go elsewhere (methods/results).

I think the correct test is either chi-squared with the counts or z (normal) with the proportions, rather than a t-test.

But honestly I don't think this point really matters that much to the paper overall. If the authors just plot the ratios with some error bars we can see better for ourselves. As there are only three replicates, plotting the points for each of these would be ideal!

So I would suggest removing the formal statistical tests, their overly long description (or at least relegating all of this to the supplement), and just plotting the data and its variation clearly.

- * Now going back to figure 2. We once again have a tortured legend with a very complex statistical procedure. I have exactly the same comment as I did with figure 3, just forgo this and give us the plots. You could also add the p-values on the slopes of the regression, but it's probably not necessary. To compare WT and toIC mutants I think a regression model with interaction terms might be possible, but I am again unconvinced it's really needed.

- Multiple testing. I think there is a confusion in language between reviewers and authors which has led to a misunderstanding. Please see https://en.wikipedia.org/wiki/Multiple_comparisons_problem for what the reviewer is referring to. I would suggest adding an adjusted p-value column. In the R programming language, one can use `p.adjust()` with a list of

p-values to adjust them. I would recommend setting method = "fdr" and choosing a suitable false discovery rate. But, as suggested, these should probably be removed.

(as a side note, if we really did want to do all of this properly, I'd suggest going with a simulation and bootstrapping approach for all the derived quantities -- but it's not going to add anything given the amount of data)

Minor comments:

Should all be addressed

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

I think the revisions here have improved the data presentation of the paper and addressed the previous concerns. Showing the data points in figures 2 & 3 really helps.

I have no further concerns.

(Remarks on code availability)

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 1

Comments to the Authors

In this interesting study, Stine et al. build on the growing body of recent work demonstrating cell-to-cell variation in antibiotic susceptibility, by analysing how susceptibility is correlated along lineages within isogenic bacterial populations. To do so, they analyse data from time-lapse microscopy imaging of E. coli under cefsulodin treatment and additionally develop a 3-state Markov chain model of cell states that can capture history-dependence within lineages. I believe this modelling approach is novel in this context and could be applied more broadly to similar data in the future. Overall, the authors' approach made sense, but their results critically need statistical analysis in order to draw solid conclusions. The paper would also benefit from a more detailed explanation of the modelling approach and provision of the associated code. The writing was mostly clear with some minor exceptions noted within my detailed comments below.

We are very grateful for the constructive comments. These comments were incorporated into the revised manuscript, which further strengthened our work.

Major comments:

1. The results currently lack statistical analysis. Adding statistical testing is critical to back up any statements comparing outcomes with a baseline expectation or comparing strains, generations, etc. These statements were particularly apparent in the Results section "Enrichment of robust lineages", e.g. that "this frequency was much higher than the frequency expected by random chance" (L175-6) or "We observed a decreasing probability of the termination with each additional generation of robustness" (L191-2). Essentially, all the results shown in Figures 2-3 and Suppl. Figures 2-5 need statistical testing to support the conclusions.

We agree that statistical testing is needed to further support our conclusions. We performed statistical tests for data described in Fig. 2-3 and Supplementary Figures 3-5, specified the p-values, and discussed their meanings in Supplementary Text 2. We then added pointers to this Supplementary Text in the corresponding main text and figure captions (e.g., L130, L157, L440, and L463) as well as Supplementary Figure captions. We did not perform statistical tests for Supplementary Fig. 2 because they were just another representation of the data described in Fig. 3a, for which we conducted the statistical tests.

2. I'm not following the reasoning behind the verbal model outlined on L200-208 re: "two potential mechanisms for the observed propagation of robust lineages". The definitions of each mechanism were vague; are there possible molecular mechanisms in mind and are these presumably heterogeneous across lineages? What exactly are the predictions arising from each proposed mechanism? (The conclusion that the data supported one but not the other mechanism was not obvious to me.) Perhaps more importantly, however, this verbal model does not seem to line up with the mathematical model presented later (Fig. 4a). The transitions between 'Healthy' and 'Vulnerable' states in the latter seem to align more intuitively with accumulation/decrement of cellular damage, the mechanism that was rejected in the verbal reasoning. I'm not convinced that the earlier verbal model is needed at all, but if it's kept, it needs a clearer explanation and consistent interpretation compared to the later mathematical model.

Thanks for this insightful comment. As the reviewer accurately pointed out, the earlier verbal argument is confusing and irrelevant to the major conclusion of the section (i.e., selective enrichment of robust lineages). Following the reviewer's suggestion, we removed this verbal argument.

3. The Markov chain modelling approach is nice, but would really benefit from some additional explanation, especially for less-specialized readers. I think each of the following points can be easily addressed, but would collectively make a major improvement.

A. Previous modelling approaches have alternatively been called “branching processes” (Intro, L75) or “two-state Markov chain model[s]” (Results, L232). Please clarify the connection, i.e. that these are the same thing; a branching process tracks live cells, while the two states in the Markov chain are (I think?) live and dead cells. It would also be useful to explicitly state the key assumption of a branching process – i.e. that the fate of each individual is independent and identically distributed – which in turn implies the lack of kinship correlation (L234) in previous models.

This is a good point. We modified the result section. Now, the text reads (L246) “Previously, we and others described the dynamics of populations exposed to antibiotic using a branching process²²⁻²⁴. This process was mathematically modeled as a two-state Markov chain, i.e., where a cell can be in either an alive or a dead state, with the state of each cell is randomly and independently determined.”

B. When a mother cell produces two daughters, do both daughters take on the same state or does each daughter draw its state independently according to the transition probabilities? (Clarify around L243 and again in Supplementary.) C. The state change is upon production of daughter cells, not during the mother’s lifetime. This was stated in the Supplementary but should be clarified in the main text (around L245 – it sounded like recovery of the mother cell then birth of daughters were two steps).

Cells’ fate is independent, is maintained during the lifespan, but can change at the time of division. We modified the main text to clarify this point (L256) “In our model, a cell maintains its state but can change it at the time of division. When a cell divides, a daughter cell’s state is determined independently of the other. A daughter cell can either retain the same state as its mother or transition to an adjacent state. For example,.....” We also modified the first paragraph in Supplementary Text 1 to reflect this change.

D. Add more detail on the fitting method – what algorithm, in what software? It would also be helpful to number the equations in the Supplementary Text and refer to exactly which equation is being fit to the data and hence plotted in Fig. 2b-c.

The necessary information is now added in the Supplementary Text 1 (last paragraph). Briefly, we used MATLAB, with a grid-based fitting function that repeatedly zooms in on the best part of the target space. Also, the equations are numbered now.

E. Please deposit the code used for the model fitting.

It is deposited.

F. For additional clarifications on the mathematical presentation, please see “Minor comments, supplementary text” below.

Minor comments, main text:

1. The experiments were performed with only one antibiotic, cefsulodin. This should already be mentioned in the Abstract (e.g. L19).

Modified as suggested.

2. L57: MIC is a population-scale measure, but not technically a population average. Rather, it tends towards the upper end of the distribution of antibiotic tolerance within the population, since growth from only a few cells will be sufficient to see population growth.

The reviewer is correct. We changed it from “population-averaging” to “population-scale”.

3. L74-5: Ref.25 does not actually use a branching process to model cell growth and death – only the ultimate outcome of ‘establishment’ is considered as a random event (agnostic of underlying dynamics).

The reviewer is correct. We removed this reference in the following sentence “Stochastic branching processes were used to model cell growth and death under antibiotic exposure as random chance events, predicting random variations in population size²²⁻²⁴”

4. L76: I find the wording “rapid population fluctuations” confusing. Stochasticity in cell division/death will introduce variation in population dynamics amongst lineages, but not necessarily sustained up-and-down fluctuations, nor are the changes in population size necessarily ‘rapid’.

We changed it to “random variations in population size” to avoid confusion.

5. Around L84: It would be helpful at this point to mention the experimental set-up in which cells are growing – i.e. as micro-colonies on solid media, as opposed to say microfluidics devices.

We changed the text to “Specifically, we plated an isogenic population of *E. coli* cells on an agarose plate containing a β -lactam antibiotic (cefsulodin)”

*6. Results: I would have found it helpful to introduce the *DtolC* strain earlier on and explain why it is being used, since it already appears from the start in Fig.1.*

This is a good idea. We had discussed the TolC-mediated efflux in the second paragraph of the Introduction. We now altered the last paragraph of the Introduction to remind the reader of this again. We additionally clarified that the experiments were repeated for $\Delta tolC$ strain. See the modified text (L89) “Given the well-known role of TolC-mediated efflux pumps in antibiotic resistance², we repeated the experiments for the *tolC* knockout strain.”

7. L101-2: Please clarify what you are testing (re: “the basis of this [branching process] modelling”) – is it the assumption of independent and identical fates amongst individuals?

That is correct. We wanted to test whether cell fate is truly random. We modified the text (L106) accordingly as follows “To test the basis of this modeling (i.e., each cell’s fate is stochastic), ...”

8. L123: Does the pairwise Pearson correlation analysis account for the fact that not all pairwise relations are available to analyse? For example, the 1st cousin will not exist if the aunt died and therefore we do not have a value of X (survival) for the 1st cousin. Does this bias the analysis towards positive correlations if only descendants of surviving cells are analysed?

Pearson correlation only looks at cell pairs in which both cells are available. This analysis evaluates whether the survival probabilities of the two available cells are independent or not. The

fact that we only analyze the available cells does not introduce the bias towards positive correlations. For example, consider a hypothetical scenario where cell survival is a Markov process. Even if we just analyze the available cells, we still get zero correlation (because the fates are uncorrelated. We had confirmed zero correlation in our computer simulation).

9. L131-2: I would be surprised if “the breakpoint concentration of cefsulodin immediately killed cells” – surely there is some time delay? Did you mean, kills (close to) 100% of cells? Please rephrase.

The reviewer is correct. We rephrase it to L136 “the breakpoint concentration of cefsulodin completely cleared the $\Delta tolC$ populations”

10. L138: “Having analyzed the genealogical tree horizontally” – this wording was confusing because the previous analysis did in fact consider cells in different generations.

The reviewer is correct. We removed the phrase.

11. L141 and elsewhere: $[X]$ is not standard notation for the mean. Not essential, but it might aid the reader to stick to more standard notation of $\langle X \rangle$ or $\bar{X}$.

We changed it to $\langle X \rangle$.

12. L155-165: I found this paragraph hard to follow and I think it would benefit from a few wording tweaks for clarity.

a. L156 – “To further validate this, we characterized cell survival along the lineages” – this statement confused me because I thought you had characterized cell survival along the lineages already in the previous section. Can you be more precise?

We changed it to L162 “To further validate this, we analyzed the various configurations of cell survival along the lineages. Specifically...” We then described what we meant by configurations by going over each example.

b. Please clearly state the definitions of D and R , and also preferably define N before (rather than after) it is used.

We modified two paragraphs (starting at L161 and L175) to define these parameters before they are used.

c. L162-3: “more frequently than expected by random chance” – How is the null expectation derived?

The probability that a cell survives antibiotic treatment is $\langle X \rangle$. Therefore, the probability that two daughter cells survive randomly in the absence of any correlation is $\langle X \rangle^2$. We clarified this by modifying the text (L171) as follows “We compared it to the frequency expected by random chance, $F_0(R_{N=1})$, which we calculated based on the probability for two daughter cell survival being $\langle X \rangle^2$.”

13. L169: I’m not sure what you mean by “left column” in Fig. 3b.

We changed it to “red column”. We believe it should be clearer.

14. L197: *less likely than what?*

We meant less likely than robust lineages. We modified the text (L211) as “these non-robust lineages are much less likely to produce robust daughter cells than robust lineages”

15. L220: *Please define “ancestral survival” (YN).*

Y refers to the survival pattern of ancestors. For example, $Y_{aunt}=1$ or 0 means its aunt cell was alive or dead. To clarify this, we modified the text as follows “Briefly, we used Shannon entropy to quantify the uncertainty²⁶, i.e., missing information in predicting the fates of two daughter cells, $H(D) = -\sum_{D=0,1,2} p_D \ln p_D$, where p_D refers to the probability that D daughter cells survived. Additionally, we can determine these probabilities conditioned upon ancestral survival, whose pattern we denote by Y . For example, we can determine the probability of daughter cells when their aunts were alive ($Y_{aunt}=1$) or dead ($Y_{aunt}=0$).”

16. Around L229 – end of section “Enrichment of robust lineages”: *What about the results for the DtolC strain?*

Good point. We added an additional paragraph to describe the $\Delta tolC$ strain (L238) “We next repeated these analyses for the $\Delta tolC$ strain. Above, we had shown that the $\Delta tolC$ strain exhibited stronger pairwise correlations of cell survival than WT (Fig. 2bc). In agreement with this observation, robust lineages had a higher tendency to produce robust offspring in the $\Delta tolC$ strain (green columns in Fig. 3a-c) than in WT (red columns). As a result, the ancestral history contained more information about the offspring survival in the $\Delta tolC$ strain than WT (Fig. 3d).”

17. Fig. 2b-c: *The connected line suggests an ordering of the relationships on the x-axis. Is there a particular reason behind the ordering of relationships with equal total ‘distance’ (e.g. aunt and niece)? If not, perhaps rethink the data visualisation so that relationships of equal distance are grouped together or otherwise indicate that there is no expectation of decreasing correlation coefficient within the group.*

This is a good point. We inserted brackets for relationships of equal degrees of connection in the figure and described the meaning of these brackets in the figure caption.

18. L273: *Fig. 3c does not clearly support the statement that there is “stronger lineage- dependent dynamics of robustness... in the DtolC compared to WT”, but statistical testing (Major Comment 1) should clarify this point.*

Thank you for catching this. We meant to reference Fig. 3ab, not Fig. 3bc, because Fig. 3ab demonstrates the robust cells whereas Fig. 3c shows the non-robust cells. This mistake is now corrected. We additionally conducted statistical tests to support stronger dependence of robust survival on the number of generations and longer memory in $\Delta tolC$ than WT; please see Supplementary Text 2 (3d).

19. L264-274: *This is a nice explanation of the model estimates and seems plausible. How would you test these hypothesised molecular mechanisms experimentally? I think this would be a useful addition to the Discussion.*

This is a good suggestion. Indeed, the design of promoters that enables precise control of efflux pump expression was proposed (PMID: 29754221). Such precise control of the expression would be useful to test the hypothesized effects of stochastic pump expression on the propagation of phenotypic resistance. We added an explanation of this idea in a new paragraph (the second paragraph in the Discussion section).

20. L291-2: I'm not convinced that a "healthy state" per se is the "minimum ingredient" to explain the data; isn't the fact that there are three states the critical point for producing correlations within lineages?

The review is correct. The Markov chain with three states is what produces the correlations. We modified the sentence as follows L305 "Based on the analysis of robust lineages, we developed a three-state Markov chain model that distinguishes healthy and vulnerable phenotypic states, thus articulating the minimum ingredient to explain our data (Fig. 4). "

21. L306: What is the "resistance state"? Does this correspond to a specific state in the Markov chain model?

Yes. While we consider this question, we realized that the knowledge of not only the transition into and out of this resistance state, but also vulnerable and dead states will be informative for understanding the pharmacodynamics. We therefore changed the phrase to L326 "The dynamics of cell growth and death can be analyzed quantitatively to deduce the rates of transition into and out of the distinct phenotypic states (Fig. 4)".

22. L340: Cells are counted as dead if they are lysed or 'permanently' arrest growth. An unavoidable limitation is that growth arrest is only observed for the limited time span of the experiment. Is there a risk of bias in that later-arising cells are more likely not to have resumed growth (yet) by the end of the experiment, and thus more likely to be counted as 'dead'?

Thank you for pointing this out. Our explanation was not clear. Most growth-arrested cells ultimately lysed. Only a small minority of cells, approximately 1 ~ 5 %, exhibited growth arrest for a few frames (~ 1 hr) before resuming growth. Since we can scroll forward or backward through time when looking at the videos, it is easy to distinguish them. When reviewing the videos, we made sure to mark cells as 'dead' only when cells were lysed. Also, our data cutoff frame was chosen before the end of the experiments. Therefore, for all the cells analyzed, we were able to confirm cell lysis (by scrolling past the cutoff frame). We changed our description in the Method section to reflect this explanation.

23. L362: Please clarify how you calculate "the probability of each microstate" – is it the proportion of pairs of daughter cells exhibiting each outcome?

We now described the procedure without invoking the new term (microstate). In the main text, we previously defined D as the number of daughter cells that survived antibiotic treatment ($D=0$, 1, or 2). We meant the probability that D daughter cells survived antibiotic treatment. We changed the text to reflect this explanation.

24. L364: Define Y ; similarly around L366, remind the reader what X is.

We used Y to denote the fate of the ancestors. For example, when the aunts were alive or dead, $Y_{\text{aunt}} = 1$ or 0 , respectively. X should have been Y (it was a mistake). We rewrote this paragraph to clarify this.

25. Author contributions seems to be missing a mention of who carried out the experiments and who designed/analysed the mathematical model?

This is now corrected.

26. Fig. 2 caption:

a. L401-2 wording clarifications: open/solid circles $\odot/\bullet$ symbols (there are both circles and squares). Columns $\odot/\bullet$ error bars (I think?)

We described the Fig. 2bc and Fig. 2d symbols in one setting, which was confusing. We have split it. We believe it is now clear.

b. L401-2 refers to “raw data from three biological replicates”. Not really ‘raw’ data as you are reporting a summary statistic (mean of X), right? Please mention how many cells were included per biological replicate. This might help explain why the mean for D_{tolC} at ages 1-3 does not appear to fall where it should amongst the 3 replicates if all were equally weighted.

The reviewer is right. This is now corrected. Additionally, we included the number of cells analyzed. “The number of cells with known fates (and thus the amount of usable data) differed between the replicates; the three biological replicates yielded 2529, 5737, and 4246 cells with known fates for WT, and 2535, 2261, and 10070 cells for Δ_{tolC} . “

27. Fig. 3 caption:

a. L412: “ratio” would be clearer than “relative difference” (the latter suggests a subtraction rather than a division)

We now use a term “fold difference” and specify it by “ $F(R_{N=1})/F_0(R_{N=1})$ ” in the main text (L174). We further indicate that it means a ratio in the figure caption (L450).

b. L418: $P(D=2) = X^2$. Should this be $[X]^2$ (mean of X , squared)? Mean over all cells in all generations?

Correct. We changed it to $\langle X \rangle^2$.

28. Fig. 4 caption, L425: “The previous two-state Markov model cannot explain the survival correlation” – this is confusing as the figure does not show the two-state model. This statement rather belongs in the associated main text, not the figure caption.

We agree. We removed this statement in Figure caption (note that we already discussed the shortcoming of the two-state model in the main text: the paragraph starting at L244.)

29. Throughout the figures, wherever a probability or proportion is plotted on a linear scale (Fig. 2d, 3, Suppl. Fig. 3-5): I suggest starting all y-axes at zero. Starting above zero can be unintentionally visually

misleading as it exaggerates the magnitude of differences between bars, and makes it harder to directly compare different plots.

Corrected.

30. Fig. 4a: Perhaps add arrows from Healthy back into Healthy state, and likewise for Vulnerable, indicating that it's also possible for a cell (or rather its daughters) to stay in the same state.

Corrected.

Minor comments, supplementary text (model description):

1. *Around the first paragraph, add a couple of clarifications:*
 1. *Is this a discrete-time Markov chain? (In this case, it would be helpful to mention that 'rates' are transition probabilities in each time step.)*

Correct. We added the term 'discrete-time' and included the following statement "Because this is a discrete-time model, the rates of transition correspond to transition probabilities per generation. "

2. *Does each daughter cell take on its state independently, or do both daughter cells take on the same state?*

Independently. We included the following statement "The two daughter cells in a pair take on their states independently"

2. *Note that D , R refer to different things in the main text (surviving daughter cells and robust cells, respectively) than they do here (damage and recovery rates). D_n is also used below to mean number of dead cells in generation n . It would be helpful to choose distinct notation.*

We agree. We have changed D to P , R to Q , and D_n to DD_n .

3. *Please define the order of states in the rows and columns in the transition matrix T , and specify whether entry T_{ij} corresponds to the transition probability from state i to j or vice versa. Note that T itself does not need to be written within square brackets.*

The state order is now given in the third paragraph.

4. *Where T is defined (in both the 3×3 and 4×4 matrices), both the $[1,2]$ and $[2,1]$ entries are R but I think one of these should be D .*

This is fixed.

5. *Please clarify what happens to dead cells in the 3×3 version of T . The definition of transition probabilities in the matrix ($T_{3,3} = 0$) seems to suggest that dead cells "disappear" after one generation – so would it be more accurate to define D_n as "newly dead" cells in generation n ? I think these entries cannot then strictly be interpreted as "transition probabilities" since from state*

D the cells just disappear rather than transition (or stay) anywhere, i.e. the 3rd column of the matrix does not sum to 1.

This is true – strictly speaking, 3x3 matrix is not a transition matrix. This is why it was necessary to expand this to 4x4 matrix. The 3x3 matrix has the survival rate as its eigenvalue, while the 4x4 matrix is the one that gets multiplied out to find correlations. To clarify this point, we expanded the description in the third paragraph and also in the subsection with heading “Finding Kinship Correlations With a 4-State Transition Matrix”

6. *Bottom of 1st page: What is the solution to this eigenvalue problem? (Or if the solution is not important, why bring up the eigenvalue problem at all?)*

The solution of the eigenvalue problem is now given in Eqn. 4.

7. *Where the 4x4 matrices are defined (2nd page): Please explicitly define the vector [H, V, D, NB] and note that only one of these entries can be 1 (the rest 0) in order to interpret the outputs as probabilities of the i^{th} descendant falling into each state. It would be useful to have distinct notation for the vectors on the left-hand side of the equations to indicate the initial state.*

The vector is now defined in the text above Eqn. 6. We additionally note that for a single cell, it will always consist of a 1 and three 0s.

8. *How are the “elder” and “younger” cells defined if two cells are in the same generation?*

We have removed references to “elder” and “younger” cells from the description of the Markov model since the pole age isn’t relevant here. Now we just talk about two cells called x and y . It doesn’t really matter what order they’re in – x and y could be two cells in a pair of siblings, or an aunt and a niece, or a niece and an aunt, etc.

9. *Page 3: “In order to find the Pearson correlation coefficient” ...of what?*

The text above Eqn. 13 now reads “In order to find the Pearson correlation coefficient for the survival outcomes of cell pairs...”

10. *Please define the notation f_H , f_V , and b_{total} etc. (What does the ‘total’ subscript mean? A population average, weighted by initial frequencies?)*

f_H is the fraction of cells in the steady-state population that are born healthy. f_V is the fraction that are born vulnerable. b_{total} is the probability that both cells in a pair will survive, normalized across the offspring of both healthy and vulnerable parents. As the reviewer pointed out, the subscript total to denote the population average weighted by initial frequencies of healthy and vulnerable cells. There are separate values b_H and b_V for probability of both offspring of a healthy parent surviving, and of both offspring of a vulnerable parent. b_{total} is different in that it is a weighted average that takes into account what fractions of cells are healthy and vulnerable to begin with. We have added this discussion to the text above and below Eqn. 13-16.

11. In the definition of the correlation coefficients:

1. What constitutes a 'datapoint'? A pair of cells?
2. The index i runs over pairs of cells, presumably each with the same relationship?
3. Define $\bar{z}$ and $\bar{w}$
4. If using z and w for the datasets, then also write C_{zw} rather than C_{xy} (though arguably x and y would be better, as this notation was used elsewhere for survival indicator random variables)

12. Page 4: Note that the rewriting now assumes z is the “elder” cell and w is the “younger”

A datapoint is a pair of cells, and i runs over all the cell pairs with a given relationship. We used the $\bar{\cdot}$ to denote their mean. We have changed C_{xy} to C_{zw} . These are now corrected and/or explained.

13. Please add a description of how the model was fitted to the kinship correlation data. Which equation was fitted? How? (Least squares? Some other optimisation algorithm?)

The final paragraph of the Markov Model section now explains this. Optimization was done by the least squares method.

Minor comments, supplementary figures:

1. Suppl. Fig. 1:

1. Perhaps the issue is on my end, but there appears to be a rendering error in the pdf. The axis labels, legends, and panel title on the left panel are missing.

We re-generated the figures. We hope it resolved the issues.

2. “100ul media was transferred...” – “(diluted) culture” rather than “media”?

Corrected.

3. “The experiment was repeated twice” – why not show both replicates?

We now show all three replicates.

2. Suppl. Fig. 2:

1. In the tables, “Observed frequency” is unclear – rather, “Number of robust cells”?
2. In the caption, define the white bars
3. How are the “frequencies expected by random chance” calculated? (Perhaps this belongs in Methods rather than the figure caption.)

We have extensively modified the table legends, figures and captions for additional clarification and also to incorporate other relevant information. For example, “Observed frequency” has been changed to “Number of robust cells observed (m_{robust})”. We have included the details of the calculation. We have described what different symbols meant.

3. *Suppl. Fig. 4 and 5: In the captions, define the dashed line on the plots.*

The dashed lines indicates the value of $P(D = 2 | R_1)$ in the Sup Fig. 4, and the value of $P(D = 0 | R_1)$ in Sup Fig. 5. This information has been added to the captions.

Reviewer 2

Reviewer #2 (Remarks to the Author):

The paper titled “Lineage-dependent variations in single-cell antibiotic susceptibility reveal the selective inheritance of phenotypic resistance within isogenic bacterial populations” by Stine et al, presents an experimental study of antibiotic effect on the survival of cells along lineages. By tracking cells genealogical trees and carrying out statistical analysis of the cells’ survival patterns along those trees, the authors were able to show that the probability of a cell to survive increases as more of its predecessors survive. In other words, cells that survive the exposure to antibiotic pass on this ability to their offspring, which then increases the chances of these offspring cells to survive that exposure as well. The authors argue that this is contradictory to previous theories that proposed that survival of bacteria upon exposure to antibiotics is due to random process that allow some cells to adapt to the selective pressure independent of the cells’ history. The authors propose a model to describe their findings, in which cells can switch between healthy and vulnerable states, and that the healthy state can be inherited by the offspring cells and therefore increasing the chances of cells, whose ancestors survived, to survive as well. On the other hand, the vulnerable state can switch to a third state (death), from which there is no recovery. The rates of transitions between the different states are determined from fitting the model predictions to the measured correlations between the different types of cell pairs with various degrees of relationships.

The study is very interesting and can be suitable for this journal, but I have a few important issues that I think should be addressed before I can recommend the manuscript for publication.

Thank you for the positive evaluation of our work. It was very encouraging. We incorporated the reviewer’s comments into the manuscript. Please see below for the details.

Major concerns:

1) How is the age of the cell determined? The authors state in “cell tracking” that the labeling of cells is based on the old pole age, but they have no way of knowing the age of the oldest poles in the sequence, and the two cells labeled “0” and “00” in supplementary figure 6 is not the same. Is that important? And if not why? If that is not what the authors mean by age, then they might consider using a different term or explain it better to avoid confusion.

We adopted a widely used definition of age for bacteria undergoing binary fission: i.e., the age of the cell’s old pole in generations (e.g., PMID: 15685293 or PMID: 30213942). Our experiment starts with single cells, whose pole ages are not known. Therefore, the initial inoculum was not included in our analysis. The initial inoculum grows and generates new cells. And these new cells age over time as they divide. Our tracking applies to those cells generated during our observation. Watching the video, we pinpoint when these cells are born and follow how many times they divide as the video plays, thereby identifying the age of each cell. Supplementary Fig. 6 shows our bookkeeping protocol, which is similar to the protocol used in PMID: 30213942. We modified the Method section (L368) to clarify this procedure.

2) How was the breakpoint concentration for the antibiotic determined? Was it determined in liquid culture or in agar? And if in one and not the other, isn’t the effect of antibiotics different for the growth in different conditions?

Breakpoint is set based on a combination of laboratory testing, pharmacokinetics, pharmacodynamics, clinical data, and expert consensus (see Ref 6 for details). We did not measure the breakpoint value in our experiments. Instead, we cited the value available in the literature (Refs 28 and 29).

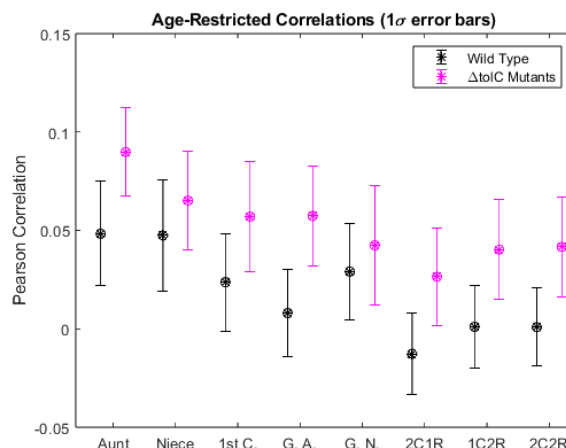
We did measure the MIC for cefsulodin in our experiments. We believe this is what the reviewer was asking. In a liquid culture, we observed that the WT population exhibits growth at 40 $\mu\text{g/ml}$ cefsulodin but no growth at 50 $\mu\text{g/ml}$ (Fig. 1 and Sup Fig 1), thereby determining the MIC to be 50 $\mu\text{g/ml}$. For the ΔtolC strain, we observed growth at 20 $\mu\text{g/ml}$, but not 30 $\mu\text{g/ml}$, thereby determining its MIC to be 30 $\mu\text{g/ml}$.

Additionally, we have included data from MIC measurements conducted on agar plates (Sup Fig. 1c). The MICs were 32 $\mu\text{g/ml}$ for WT and 20 $\mu\text{g/ml}$ the ΔtolC strain, which are slightly lower than the MICs measured in liquid culture. However, we conducted the MIC measurements with fine increments of antibiotic concentrations to identify the concentration suitable for our single-cell-level studies. In typical measurements in the field, the concentration is increased by twofold. Therefore, it is common to see a twofold difference in MIC determination. In this regard, the MIC difference we observed between in liquid and on agar is minimal and generally conserved within a measurement error.

We changed the paragraph starting in L57, Figure 1 caption, Supplementary Figure 1 and its caption to reflect this discussion.

3) In the last paragraph of the results section (lines 264 – 274) the authors discuss the effect of tol-C expression on the correlations between siblings. I would note here that, a previous study (cited here as well, Bergmiller et al., Science 356,311–315 (2017)) has argued that since the accumulation of these pumps is at the poles, it results in asymmetry between siblings and therefore phenotypic heterogeneity in the response to antibiotics. This could be why delta tol-C strain exhibits higher correlation between siblings. The question then is, since the TolC accumulation in the cell is correlated with the cell's age, could its effect be compared between pairs of cells whose oldest pole is of the same age vs. pairs of cells in which the oldest poles are of different age? (can we analyze?)

This is an interesting question. We are also intrigued by this possibility; whether the correlation would be higher when only poles of the same age are compared. However, if we limit kinship correlation to same-age cells, we can no longer analyze sibling pairs because they always have different ages (this is because a cell always has one old pole and one new pole. When this cell divides, one cell inherits the old pole while the other cell inherits the new pole. So, sibling pairs always have different pole ages). But, we can still calculate the correlation for other relationships. Please see the graph below.



The results show a similar pattern: the correlation decreases at distant relationships. The Δ tolC strain exhibits stronger and longer correlations than WT. However, please note that the error bars are significantly larger (compared to Fig. 2bc). This is due to the fact that we had to deal with much smaller sample sizes (more than two thirds of cell pairs have different ages and had to be discarded in the analysis). Because the error bars overlap for almost all relationships, it is difficult to draw meaningful conclusions.

In summary, while we agree the reviewer's hypothesis, our sample size is not large enough to rigorously test it.

4) In lines 285 – 287, the authors state that “This robustness further strengthened across the generations, indicating phenotypic adaptation to antibiotics (Fig. 3). This positive feedback was due to the inheritance of resistance factors whereas antibiotic-induced damage remained constant.” Isn't this statement contradictory to fact that Delta tolC exhibits higher correlations between generations? Or is the correlation in the delta tolC case is higher due to the inheritance of the vulnerable state?

Our model supports the second explanation provided by the reviewer: the Δ tolC strain shows a higher correlation due to the inheritance of the vulnerable state. This is illustrated in Fig. 4b right panel, which shows that the Δ tolC strain has a low rate of transition out of the vulnerable state. Our interpretation, which we believe the reviewer is getting at, is that due to the low rate of transition, the vulnerable state is maintained through generations, which is manifested as higher correlations.

5) In the supplementary movie, cells that survive and continue to proliferate appear to grow over cells that perished. Do the authors think that this can have an effect on the survival chances of cells? For example, the areas where cells have grown and died before could be depleted of antibiotics, or the debris of dead cells could provide resources to the surviving cells? Is there a way to assess this effect or explain why it should not be considered?

This is an interesting question whether dead cells help deplete local antibiotic molecules in the experiments using an agar pad. While this is possible, we believe that its effect is rather minimal. This comes from our prior study utilizing the same experimental protocol (PMID: 29508699). In this prior work, we had tracked the number of live cells during cefsulodin exposure (the present work tracked not only the number of live cells but their relationships). During this prior work, we had repeated the experiment using a microfluidic device, in which antibiotic molecules were constantly replenished. We had found that using this device did not alter the population dynamics, which led us to the conclusion that the depletion of antibiotic molecules on an agar pad was not severe enough to affect the experiments.

6) Can the authors add a supplementary figure showing how the data was fitted by the model to extract the transition rates in Fig. 4b?

Briefly, we derived the mathematical dependence of survival correlations on the transition rates (P : damage rate, Q : recovery rate, and K : kill rate) and the mean survival rate $\langle X \rangle$ through a series of equations (Eqs 1 – 25). K is related to P , Q , and $\langle X \rangle$ as described by Eq. 4, while the $\langle X \rangle$ is experimentally measured (main text). Thus, we have two fitting parameters (P and Q). We find the values of P and Q that best fit our data (Fig. 2bc) using a least squares method. The

identified P , Q , and K values are plotted in Fig. 4b. We agonized over how to illustrate this procedure. Unfortunately, we could not come up with informative figures. Instead, we improved our description throughout Supplementary Text 1 and included additional paragraphs (the last three paragraphs in Supplementary Text 1) to clarify our procedure. We hope this addresses the reviewer's concern.

Minor concerns:

1) Explain breakpoint concentration of antibiotic better and describe how it was determined experimentally.

Please see our response to Q2 for details. Briefly, we improved our explanation of the breakpoint in the paragraph starting at L57. We additionally described how we experimentally determined MICs in Figure 1 caption.

2) Describe the mode of action of the antibiotic used.

Cefsulodin is a β -lactam antibiotic, which inhibits the synthesis of bacterial cell walls. This mode of action is now provided (L85).

3) Describe how exactly the frequency of consecutive generations exhibiting robustness expected by random chance is calculated and what is the expected value.

The expected number of robust lineages based on random chance, m_{expected} , is calculated as follows. Because the probability that a cell survives is $\langle X \rangle$, the probability that two daughter cells survive is $\langle X \rangle^2$. Therefore, the expected number of robust cells based on random chance, m_{expected} , is the total number of cells (m_{total}) multiplied by $\langle X \rangle^2$: $m_{\text{total}} \times \langle X \rangle^2$ for $N = 1$. For two consecutive robust generations, we additionally need to consider the survival of the next generation (four additional cells): thus $m_{\text{total}} \times \langle X \rangle^{2+4}$ for $N = 2$. Three consecutive robust generations require consideration of another generation (eight additional cells): $m_{\text{total}} \times \langle X \rangle^{2+4+8}$ for $N = 3$.

We included this information in Supplementary Fig. 2 caption. The values are provided in Supplementary Fig 2 table.

4) Define the relative difference in frequency. Is it $(\Delta F)/F_0$?

We realized the term, relative difference, was confusing. We rephrased it as “We plotted their fold difference, i.e., $F(R_{N=1})/F_0(R_{N=1})$, in Fig. 3a.” (L174). We additionally inserted the following phrase “Their fold difference (i.e., ratio) is plotted here” in Figure caption (L450).

Reviewer 1

After reading the revised version of the manuscript, I maintain my view that this study is interesting and contains worthwhile ideas, data, and approaches. However, now that the authors have provided some more details of their methods, unfortunately I see a couple of major issues with the quantitative approaches:

1. The mathematical model is not well-defined and the analysis appears to contain a few errors. I think this could be fixed, but would require re-doing the fit to data (with possibly different results). The calculations also make a strong assumption about the cell population starting in steady state.
2. The statistical analysis, which has now been added, is not sufficient. Not all of the key results have been tested statistically, and where tests have been carried out, not enough information has been provided (e.g. sample numbers) and in at least one case the choice of test does not appear suitable. Moreover, the results of statistical testing generally are not mentioned in the main text, and do not uniformly support the headline conclusions. Negative results should not inhibit publication, but need to be clearly stated and discussed. Overall, I believe the statistical analysis and associated presentation of results would require substantial revisions. Please see my detailed comments below for further explanation of these points. Furthermore, there is still not enough information provided to ensure reproducibility and re-usability of the authors' analyses.

We sincerely thank the reviewer for the constructive criticism. The reviewer's feedback has greatly strengthened our work. We apologize for the extended time taken to revise the manuscript. Unfortunately, significant funding shortages required the PI to downsize the lab, resulting in the departure of the first author and the second author being preoccupied with job searches. It has been very challenging to complete the revision. Despite these obstacles, we did our best to address the reviewer's comments and believe that we have done so successfully.

Briefly, we completely re-wrote the modeling section in the main text and significantly expanded our description of the model in Supplementary Material (as a result, the supplementary text of model description increased from 3.5 pages to 6 pages, with the number of equations increasing from 25 to 39). We performed extensive statistical tests for all the relevant data (as a result, the figure caption increased from 2 pages to 5.5 pages). Of course, we modified relevant main texts. We additionally addressed other minor comments. Details are further discussed below. We really hope the reviewer finds the revisions satisfactory.

- The data file provided only contains the numerical values to be plotted in each figure, not the raw data or calculations. It is not possible to work out sample sizes in each group (e.g. generation or kinship pair), confirm which equations have been applied where, or reproduce the analysis.

In addition to the numeral values used for figure, Source data 1.xlsx, we now include Source data 2 WT.xlsx and Source data 2 tolC.xlsx, which includes the sample sizes and calculations.

- The code availability statement says that “Our custom-built software is uploaded” but does not say where, and it does not appear to be included with the files for review (apologies if I have overlooked something). I understand that some elements of analysis (e.g. cell tracking) may have been developed previously, in which case previously published work can simply be cited; but other analyses here, e.g. fitting the mathematical model to data, are novel and require provision of code.

During the submission process, we were directed to deposit our code in an external repository (Code Ocean), and we assumed that reviewers would be notified and provided access. However, it appears that this was not the case. To avoid further confusion, we have now included the code directly in the submission as a supplementary file.

I’m sorry that I cannot be more positive about the manuscript in its present form, but encourage the authors to revise their work accordingly and would be enthusiastic to see a more rigorous approach to the interesting questions they address.

Detailed comments on mathematical modelling

A) • Just as the 3×3 matrix T is not technically a transition matrix, what the authors describe as a “Three-State Markov Model” (first section of Suppl. Text 1) is not actually a well-defined Markov chain. A collection of possible cell states (phenotypes) does not automatically translate to “states” of a stochastic process in the mathematical sense. Now that the authors have explained that each daughter cell takes on its state independently, I think I can see where the problem lies: there are two daughter cells’ phenotypic states to consider, and the authors have not clearly defined how this translates to “state” of the Markov chain, i.e. what exactly is the random variable they are tracking? Moreover, a Markov chain’s state space should include all possible states of the system, so that the process does not simply terminate (it may land in an absorbing state with transition probability of 1 to remain in that state, but should always be somewhere in the defined state space). Therefore the fact that dead cells simply ‘disappear’ after one generation (reflected by the final column of all zeros in T) disqualifies this model as a Markov chain.

In contrast, the 4-state process introduced in the following section *could*, I believe, be interpreted as a Markov chain if the state of the process is defined as the phenotypic state of an individual cell, as you follow a lineage through generations (independently stepping to one or the other daughter cell in each time step). This process should be more rigorously defined. Under that interpretation, I think the following correlation analysis is okay.

As far as I understand, the point of the first section's analysis was, ultimately, to find the net growth rate of the (live cell) population. The way this is currently done uses the 3x3 matrix T but note that the 3rd row/column (corresponding to dead cells) is not actually needed to obtain Eq. [4]: I can re-derive this result using only the first two equations for H and V . However, because the 3-state process is not well defined, I think the matrix entries are not quite right for obtaining the average population growth rate. An alternative, perhaps more intuitive, way to do this would be to consider the population with a 2-type branching process tracking healthy and vulnerable cells. Then use the 2x2 matrix of expected numbers of offspring in the eigenvalue analysis. Note that the matrix entries need to account for two daughter cells each independently taking on a given cell state. This does not exactly line up with the entries in T , which only reflect the probability of one daughter taking on a given state.

If I'm correct on the above, Eq. [4] should change. Since this equation was used in fitting the experimental data, the fits would then need to be re-done, leading to different estimates of rates P , Q , and K as presented in the main Fig. 4 (and Fig. 2bc). It remains to be seen how much this may change the conclusions.

Sorry for the confusion. We realized that our previous description was confusing. To address this problem, we re-wrote the modeling description in the main text; see the paragraph starting in L279. We also completely re-wrote the model description in Supplementary Text 2.

In summary, as the reviewer correctly described, a Markov model follows for all cells' state. However, this was not an intention of our modeling. Instead, our focus was on how many cells are in different phenotypic states. Therefore, we followed an approach commonly used in population ecology to develop a population model. The matrix T used in our model is not a Markov matrix but a population projection matrix, which describes how the number of cells in different states change from time t to $t+1$. It is equivalent to a Leslie matrix, a common age-structured population model, which tracks the number of individuals in different age classes. The difference between our projection matrix and Leslie matrix is that we have different phenotypic classes instead of age classes. However, mathematical derivation is very similar. We did discuss "Markov" in the main text to mean that the future state depends only on the present state, not on the past. We did not mean that our modeling is a Markov model. However, we understand how this term was confusing. The revised manuscript no longer includes the term "Markov". Instead, we clarified that it is a population model.

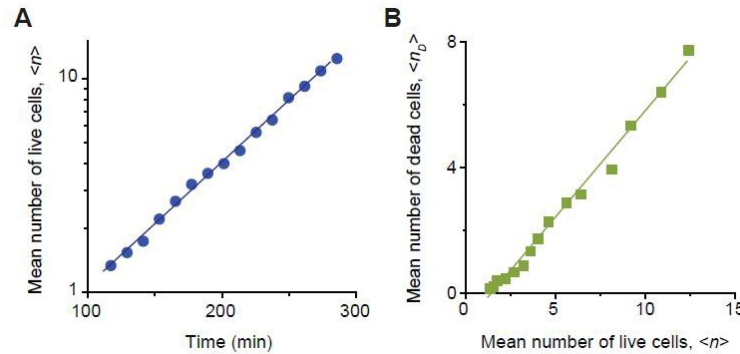
Regarding whether we should use 2x2, 3x3 matrix, or 4x4 matrix, the choice would depend on what we model. Indeed, in the current model description, we begin with 2x2 matrix (Eq. 4) and expand it to 3x3 (Eq. 11) and then to 4x4 (Eq. 16). In this description, we started from the very beginning, precisely defining each parameter and corresponding equations. I believe this explains clearly where each entry in T comes from. As described, 3x3 matrix is sufficient to model population growth. But, it is not sufficient to calculate the Pearson correlation coefficient of cell survival (as explained in the paragraph above Eq. 16). For this calculation, we need to maintain the genealogical tree structure. Our approach to address this problem was to introduce a ‘dummy’ state, i.e., “cells” born from dead cells. This state is merely a placeholder to maintain the tree structure. It does not enter into actual calculation of Pearson correlation coefficient. During revision, we carefully went over all our mathematical steps and found no errors.

We believe that this extensive revision should address the reviewer’s concerns.

B) • In describing the model fitting procedure (end of Suppl. Text 1), the authors now explain that they use the eigenvector from Eq. [3] to determine the fraction of progenitor cells that are healthy/vulnerable. However, the eigenvector would give the steady-state proportions expected to be approached in the long-term in a growing population. I would not expect progenitor cells necessarily to be near steady state. Are there features of the experimental design that could justify this assumption? How sensitive are the results to this assumption?

In our experimental design, cells were pre-treated with cefsulodin in liquid culture for 2~3 generations before being transferred to an agar pad (containing the same concentration of cefsulodin) and observed under a microscope. We did so to achieve steady state growth under antibiotic exposure during the microscope measurement. This approach was motivated by our previously published work (PMID: 29508699), where the same protocol was used to track the total number of live and dead cells under cefsulodin exposure using a microscope. Please see the

figure below (re-plotted from Figure Supplement 5 in PMID: 29508699). The number of live cells increases exponentially over time (panel A), albeit at a reduced rate compared to an antibiotic-free condition because some cells die. The numbers of dead cells and live cells show a linear relationship (panel B). These observations support the steady state growth.



We added this discussion in our Supplementary text when we mentioned the eigenvalues and eigenvectors (below Eq. 14).

C) The added description of model fitting still is not entirely clear. For instance, Eq. 9-12 and 13-16 cannot directly be used to “find” the stated probabilities; rather, they express certain probabilities in terms of P , Q , and K . Then the paragraph stating “Fitting was performed using MATLAB software...” doesn’t specify exactly what was fit. What I *think* the authors are doing (but I’m guessing here) is to use Eq. 17 or 18 to compute the observed correlation coefficients from the raw experimental data (survival status, $X=0$ or 1, for each individual cell) and then equate this to Eq. 25, the theoretical expression derived from the model (which would be clearer if Eq. 25 were written out in terms of P , Q , and K). Please clarify. Also, please make the relevant code available.

Sorry for the confusion. The fitting was performed as follows. First, the algorithm arranged P and Q from 0 to 1 (increasing by 0.02, thereby 2500 different combinations of P and Q values). K was fixed by P , Q , and the measured μ via Eq. 14. We calculated Pearson correlation coefficient, C , for each parameter combination and compared it to the experimentally measured C , with the nine measured kin correlations given equal weight. χ^2 value was calculated to choose the P , Q , K combination with the best fit. We then zoomed in the chosen combination and varied

P and Q near this combination. We repeated the iteration until we find the best fit to seven decimal places.

The original description was modified to reflect this procedure (Supplementary Text page 6 middle part).

All codes used are now added to the submission as supplementary material.

Detailed comments on statistical analysis

- Throughout: please provide more details of the data and tests being used. Make sure it is clear which subset of data is being analysed, with reference to the deposited data file(s): as mentioned earlier, I believe the data deposit should be extended to raw data. What is the sample size in each group (e.g. generation or kinship pair), in each experimental replicate, considered in each analysis? Some of this information is in the table attached to Suppl. Fig. 2 but this is not clearly referenced in the statistical analyses and does not contain the complete information needed. How were data from three experimental replicates pooled? If using a t-test, is it one-sided or two-sided?

We agree with the reviewer. In addition to figure data (Source data 1.xlsx), we now include the raw dataset (Source data 2 WT.xlsx and Source data 2 tolC.xlsx), which provides the complete information. In each statistical analysis (as detailed in Figure captions), we clarified where the data can be found (between Source data 1 or Source data 2). We also clarified whether a t-test is one-sided or two-sided. The detailed explanation about the statistical analysis, including p-values, is now incorporated into each figure caption. When there are too many p-values involved, they are provided in Supplementary tables. The key result of statistical tests was described in the main texts, and a pointer to relevant figure captions was added for details.

D) Stat. test 1: There should probably be a correction for multiple comparisons here.

We were not entirely sure what this comment meant, but we took it as adding more explanations in the main text based on the statistical tests for Fig. 2bc. The key point of Fig. 2bc data is that the correlation coefficient of X was different from zero. We detailed our procedure for the statistical tests in Fig. 2bc caption. We then added “Statistical tests reveal that the correlation is

statistically different from zero for sibling cells but not other relationships (Fig. 2bc caption)“ for WT in L135, and “Statistical tests reveal that the correlation is statistically different from zero for almost all relationships (Fig. 2bc caption)” for the tolC mutant in L146.

E) Stat. test 2: Clarify wording re: what’s being tested: “increase in survival *with pole age*”... “are statistically different *from each other*”.

Stat. test 2 was about Fig. 2d dataset. Please note that in the revised manuscript, a statistical test of Fig. 2d dataset is incorporated directly into the figure caption. We believe that putting the data description and statistical test together in one section makes the explanation clearer. We additionally modified the main text (L158-L164) for further clarification.

F) Stat. test 3a: This does not appear to be a valid application of the Kendall rank correlation coefficient. It appears that you are simply stating that the mean estimate does increase with N (as is visually obvious) but not accounting for sample size or variability among individual cells underlying this trend.

We agree. This comment is related to another comment below (Question K). Our answer to Question K should address both comments. Briefly, to statistically test whether the observed fold-difference ($F(R_N)/F_0(R_N)$) is significantly greater than 1, we performed a one-sided t-test between $F(R_N)/F_0(R_N)$ and 1 for each N using the weighted means and standard errors (as plotted in Fig. 3a, whose values can be found in Source data 1). To statistically evaluate the increase of $F(R_N)/F_0(R_N)$ with N , instead of Kendall test, we performed bivariate correlation between $F(R_N)/F_0(R_N)$ and N using their values from three replicates (whose values can be found in Source data 2). Please see our answer to Question K for more detail.

G) Stat. test 3b: Why only test generation 1 versus 3, and ignore 2?

We agree that the analysis was incomplete. This comment is related to two other comments below (Questions K and L), where we extensively described our new analysis. Briefly, our major point was that the values increase (Fig. 3ab) or decrease (Fig. 3c) with increasing N . We

therefore performed bivariate correlation and calculated Pearson's r for all the data in Fig. 3a-c, which shows that the increase (Fig. 3ab) and the decrease (Fig. 3c) are statistically significant.

Please see our answers to Questions K and L for details.

H) Stat. test 3c: Which procedure is repeated? What exactly is being tested?

Sorry for the incomplete description. We tested whether $P(D = 2 | R_N)$ is significantly greater than $P(D = 2 | \sim R_N)$ for each N (Supplementary Fig. 4) and whether $P(D = 0 | \sim R_N)$ is significantly greater than $P(D = 0 | R_N)$ for each N (Supplementary Fig. 5) using an one-sided t-test. Detailed explanations and p-values are provided in each caption of these Supplementary Fig 4 and 5. The p-value were less than 0.05, except for two cases. But even in these two cases, the p-values were less than 0.08. Therefore, their differences are generally significant. We did not perform a statistical test for Supplementary Fig. 3 data (where $D = 1$) because no claims were made from this dataset.

I) Stat. test 3d: Unclear what you mean by “the values obtained for the three robust generations” – how exactly are you pooling data? Why has this been applied to Fig. 3a, 3b, and 3d, but not 3c?

We significantly expanded our explanation of statistical analyses in Fig. 3 caption. To briefly summarize the procedure, each figure in Fig. 3a, 3b, 3c and 3d shows the weighted means and standard errors obtained using three biological replicates. We performed a one-sided t-test using these means and standard errors to determine the p-values. This procedure is now applied throughout Fig. 3a, 3b, 3c, and 3d. Please see below our answers to Question J, K, L, and M and also Fig. 3 caption for details.

- There remain some key questions which have not yet been tested statistically. These include:

J) (Fig. 2d) Is the increase in survival with pole age significant for either strain?

To test whether the increase is positive and strong, we performed bivariate correlation between $\langle X \rangle_Z - \langle X \rangle$ and Z using their values from all three replicates ($n=21$, shown as open circles in Fig. 2D, whose values can be found in Source data 1) and calculated Pearson's r . The r was 0.76 (p-value = 0.00005 based on a t-distribution, two-sided) for WT and 0.64 (p-value= 0.0018) for $\Delta tolC$, showing a strong positive increase and statistical significance in both cases.

This analysis is now explained in Fig. 2d caption.

K) (Fig. 3a) Is the fold-difference significantly different from 1 (for each strain, in each generation)?

To statistically test whether the observed fold-difference ($F(R_N)/F_0(R_N)$) is significantly greater than 1, we performed a one-sided t-test between $F(R_N)/F_0(R_N)$ and 1 for each N using the weighted means and standard errors (as plotted in Fig. 3a, whose values can be found in Source data 1). See Supplementary Table 3 for the p-value for each strain and each generation. The p-values were less than 0.05 for $N=1, 2$ and 3 in WT, and $N=2$ and 3 in $\Delta tolC$, indicating that the difference is generally significant.

Next, to statistically evaluate the increase of $F(R_N)/F_0(R_N)$ with N , we performed bivariate correlation between $F(R_N)/F_0(R_N)$ and N using their values from the three replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence $n=7$. Pearson's r was 0.89 (p-value=0.003 based on a t-distribution, one-sided) for WT and 0.82 (p-value=0.011) for $\Delta tolC$, showing a strong positive increase and statistical significance in both strains.

This analysis is now explained in Fig. 3a caption.

L) (Fig. 3b-c) Are the conditional probabilities significantly different from the unconditional (null expectation)?

Fig. 3b: To statistically test whether the conditional probabilities $P(D = 2 | R_N)$ are significantly greater than unconditional probability $P(D = 2)$, we performed a one-sided t-test between them for each N using the weighted means and standard errors (as plotted in Fig. 3b, whose values can be found in Source data 1). See Supplementary Table 4 for the p-value. The p-values were less than 0.05, except for $N=1$ in $\Delta tolC$ and $N=3$ in WT. However, even in these two cases, the p-values were less than 0.08. Therefore, the difference is generally significant.

An equally important claim we made was that $P(D = 2 | R_N)$ increases with N . To statistically evaluate this claim, we performed bivariate correlation between $P(D = 2 | R_N)$ and N using these values from three replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence $n=7$. Pearson's r was 0.84 (p-value = 0.009 based on a t-distribution, two-sided) for WT and 0.74 (p-value = 0.028) for $\Delta tolC$, showing a strong positive increase and statistical significance in both cases.

To statistically test whether $P(D = 0 | R_N)$ is significantly less than $P(D = 0)$ in Fig. 3c, we performed a one-sided t-test between them for each N using the weighted means and standard errors (as plotted in Fig. 3c, whose values can be found in Source data 1). See Supplementary Table 5 for the p-value. The p-values were scattered and it was hard to see a pattern. However, our major claim from this dataset was that $P(D = 0 | R_N)$ decreases with N ; see L215. To statistically evaluate this claim, we performed bivariate correlation between $P(D = 0 | R_N)$ and N using their values from three replicates (whose values can be found in Source data 2). In some

replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence ($n=7$). Pearson's r was -0.91 (p-value=0.002 based on a t-distribution, two-sided) for WT and -0.76 (p-value=0.024) for $\Delta tolC$, showing a strong negative trend and statistical significance in both cases. These analysis are now explained in Fig. 3b-c caption.

M) (Fig. 3d) Is the mutual information significantly different from zero?

To statistically test whether the information content is significantly greater than 0, we performed a one-sided t-test between them for each N using the weighted means and standard errors (as plotted in Fig. 3d, whose values can be found in Source data 1). See Supplementary Table 6 for the p-value. The p-values show that the difference is not significant for WT but mildly significant ($0.05 \leq p < 0.1$) for $\Delta tolC$. However, our claim from this data is that the information content is too low and thus negligible at the population level (L246 and L257).

To statistically evaluate the increase of the information content with N , we performed bivariate correlation between the content and N using their values from three replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence ($n=7$). Pearson's r was 0.36 (p-value=0.78 based on a t-distribution, one-sided) for WT and 0.84 (p-value = 0.008) for $\Delta tolC$, showing that the increase is statistically significant for $\Delta tolC$ but not for WT. Again, despite this increase, the information contents remain low (Line 257).

This analysis is now explained in Fig. 3d caption.

- The statistical results should be mentioned in the main text, and indicated on the plots. Essentially, whenever there is a statement about a quantitative result, it should be backed up with the relevant

statistical test result. Interestingly, many of the comparisons turn out *not* to be significant in the statistical tests presented so far. Therefore, the authors should be careful to ensure this is clear in the main text and conclusions are stated with appropriate caution. This may require some reinterpretation and discussion where results turned out (perhaps against expectations) to be statistically non-significant.

We fully agree with the reviewer's suggestions. In our opinion, it is customary to explain the details of statistical tests and p-values in figure caption and discuss the summary of the test results in the main text where the figure was mentioned. In the revised manuscript, the details of our statistical analyses and p-values are now described in appropriate figure captions. When there are too many p-values involved, the values are provided in Supplementary Table. When we make a statement about the quantitative results in the main text, we mentioned the key result of the relevant statistical tests, with a pointer to the caption for details. Please note that our previous conclusion is supported by the statistical tests. Therefore, we did not make changes in the conclusion.

We are deeply thankful for the reviewer's prior criticisms and suggestions about the statistical tests. We believe incorporating them truly strengthens our conclusion.

Other (mostly minor) comments

- In the writing, the authors have repeatedly referred to a cell's fate as "stochastic" or "random" (e.g. L106, L129, L247, L298) when what they really mean is that it is uncorrelated to previous generations. This is a colloquial use of the term "random", but in a mathematical sense, a random variable need not be independent or uncorrelated with others. For example, in L106, the aim is to test whether a cell's fate is independent of its parent's and identically distributed regardless of its lineage.

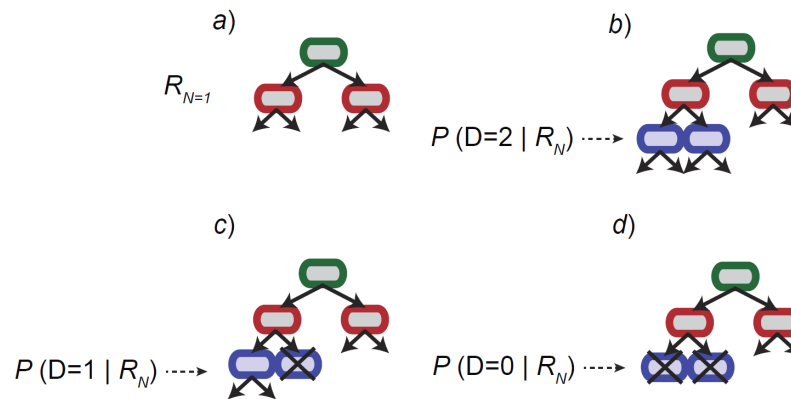
The reviewer is correct. To address this point, throughout the manuscript, when we describe our results, we now use the terms "independent" or "uncorrelated" instead of stochastic and random.

We left these terms in the introduction when we describe the literature (as those terms were used in the cited articles).

- The definition of D with respect to robust cells (L175-181) is confusing. The authors state that they are testing whether daughter cells are robust – meaning that the daughters' daughters should survive (cf. L166) – yet D is defined as "the number of daughter cells that survive" rather than the number that are robust. Perhaps the generation of daughter cells we're talking about is getting confused? Similarly, in

L199, “we analyzed how likely a robust cell gives birth to non-robust cells” but “being non-robust means either one of their daughter cells die...”: “their” is ambiguous; it must be the daughter’s daughters? It may be helpful to label the robust cells, and how many generations of robustness are shown, in the cell lineage sketches in Fig. 3.

We agree with the reviewer that our previous description was confusing. The confusion lies in large part because whether a cell survives or not is determined by whether it divided and produced two daughter cells. So, to confirm whether a cell is alive, we need to look at it until the next generation. To clarify the description, we now include additional figures (Supplementary Fig. 2a-d, re-plotted below). Please see the panel a below first.



A mother cell (marked as green) is alive if it divides (arrows). However, this division itself does not make the mother cell robust. For it to be robust, its two daughter cells (red cells) must divide (arrows). Now, please see the panel b. For a daughter cell (a left red cell) to be robust, both of its own daughter cells (blue cells) should survive (meaning they divide). The D is used to denote the number of blue cells that survive (as denoted by the arrows below them). The panel ‘b’, ‘c’ and ‘d’ respectively denotes the conditional probability $P(D = 2 | R_{N=1})$, $P(D = 1 | R_{N=1})$, and $P(D = 0 | R_{N=1})$. We modified the phrases in the main text (a paragraph starting in L183) and added a pointer to this Supplementary figure and caption.

• L229: in $H(M|Y)$, M has not been defined

Sorry. M was a typo. It should have been H(D|Y). It is now corrected (Line 244).

- L238-242 (newly added text): “the DtolC strain exhibited stronger pairwise correlations of cell survival than WT” – but from Fig. 2bc, this only appears to be true for the more distant kin relationships.

The reviewer is correct. We modified our text. Instead of stating “stronger” pairwise correlations, we now state “Pairwise Pearson correlation analysis again shows positive correlations of survival in the $\Delta tolC$ strain (Fig. 2c). The decrease of the correlations at distant relationships was more gradual in the $\Delta tolC$ strain than in the WT strain (Fig. 2b,c)” in L144.

Also, the reasoning that “robust lineages had a higher tendency to produce robust lineages in DtolC than WT” and “as a result, the ancestral history contained more information about the offspring survival in the DtolC strain”, seems to contradict the argument just made above in L233-7 where it was stated that resistance inheritance within robust lineages does not imply population-wide lineage history dependence.

This is a good point. What we meant to say was that it makes sense that ancestral history contains more information about the offspring survival in $\Delta tolC$ than WT. However, the absolute value is still very low (below 0.02 nat in Fig. 3d) and remains insignificant at the population level, meaning that the information that ancestral history contains is still negligible even for $\Delta tolC$. We clarified this by adding “Yet, the information content is still very low (Fig. 3d) and negligible at the population level “ in L257.

- Minor discrepancy: Cells were said to be exposed to cefsulodin at its breakpoint, given as 32 ug/ml (L97), but Methods states that the dose used was 31 ug/ml for WT cells (L349).

It is corrected to 32 ug/ml.

- L380: Where is the numerical value of C_B reported? What test led to the conclusion that it is not significantly different from zero?

They are reported in the ‘C_B’ column in Source Data 1. The values are very small. For example. C_B was found to be 0.004, 0.013, and 0.0001 for three replicate plates in WT sibling correlation. When we performed a two-sided t-test to see whether they are significantly different from zero, the p-value was 0.27. For aunt correlation, the C_B were 0.0354, 0.006, and 0.0156.

The p-value from a two-sided t-test shows the p-value of 0.16. We consistently observed such high p-values, which indicated that we cannot claim that C_B is different from zero. In any case, we corrected C_B in our analysis (L421). We added this description in L414.

- L387 and L391: indicate what the summation is taken over

The summation in L387 (now L426) runs from $D = 0$ to 2. The equation for the original Shannon entropy is corrected as “ $H(D) \equiv -\sum_{D=0}^{D=2} p_D \ln p_D$ ”. To clarify the summation in L391 (now L430), we added “Here, the summation runs for all different fates of Y .” We then clarified this by using an aunt as an example (L431-437).

- L390-4: I’m not entirely following this. L394 gives an example of Y but not the full equation for what is plotted in Fig. 3d, which gives results for $N=1, 2$, or 3. So is H summed over all combinations of ancestral states in generation N ? Please give a complete description.

When we calculated the Shannon entropy conditioned upon the fates of the ‘1’ generations ($N = 1$), we consider the fate of an aunt (whose relationship is denoted as [2,1] in Fig. 2a); whether the aunt lives or dies, $y_{aunt} = 1$ or 0, respectively. A mother’s state is not considered here because for a cell should be alive, its mother cell should always have lived. Then

$$\begin{aligned} H(D|Y_{aunt}) &= \sum_{y=0}^1 p_y H(D|Y = y) \\ &= p(Y_{aunt} = 1)H(D|Y_{aunt} = 1) + p(Y_{aunt} = 0)H(D|Y_{aunt} = 0). \end{aligned}$$

The difference between the original and conditional Shannon entropy is equal to mutual information, which is plotted in Fig. 3d for $N = 1$. For $N=2$, we consider not only the fate of the aunt [2,1], but also of the great aunt [3, 1] and the first cousin once removed [3, 2]. For $N=3$, we consider the fates of the cells with the relationship [2, 1], [3, 1], [3, 2], [4, 1], [4, 2] and [4, 3].

This information is now added to L431-L440.

- Fig. 2: By eye, it looks like $\langle X \rangle_Z$ is larger than $\langle X \rangle$ for all Z . How is this possible?

Please note that $\langle X \rangle_Z - \langle X \rangle$ is lower than zero at $Z=1$ and 2 ; see Source data 1. There are two factors that contribute to the difference between $\langle X \rangle_Z$ and $\langle X \rangle$. Every newborn cell has $Z=1$, and thus most represented in a population, skewing the population mean toward $\langle X \rangle_{Z=1}$. Also, to calculate $\langle X \rangle_Z$, we search for cells have survived at least up to the Z generation. As Z increases, we are likely to select for robust cells/lineages, which is why $\langle X \rangle_Z$ increases with Z .

- Fig. 3 caption (L457): “the non-conditional probability” of what?

We meant that the non-conditional probability that two daughter cells survived by chance:

$P(D=2)$. It is now corrected (now L531).

- Supp. Fig. 1 caption: It is not clearly stated what is plotted in Fig. 1a (mean and s.d.?) and Fig. 1b is not mentioned at all.

We performed temporal OD600 measurements in three biological replicates. We meant to plot the mean and spread of these OD600 values from three replicates. However, we now believe this mean value is unnecessary for the following reason. In Fig. 1a, we showed photographic images of cells growing up to 40 ug/ml but not above it for WT, and up to 20 ug/ml but not above it for DtolC. The point of the Supplementary Fig. 1 was to show the corresponding OD600 measurements. The individual graphs for all three replicates (now Supplementary Fig. 1a) show OD600 increasing to its saturating density up to 40 ug/ml, but not above this concentration for WT, and up to 20 ug/ml but not above this concentration for DtolC, agreeing with Fig. 1a image. Therefore, we now believe that just showing OD600 for three biological replicates individually is sufficient. We therefore deleted means and spreads.

- Suppl. Fig. 1c: How exactly is MIC determined from plating efficiency?

It was determined by concentration at which plating efficiency becomes zero (which was at 32 ug/ml). We clarified it in the caption.

- Suppl. Fig. 2 table: Does N refer to the generation the cell is in, or the generations of robustness?

We meant the former. We clarified the label in the table.

- Suppl. Fig. 2 plots: What do the error bars indicate?

Suppl. Fig. 2 contains a table of raw data of $F(R_N)$ as well as the plot of this data. The error bars which the reviewer was referring to was the standard deviation of $F(R_N)$ from three biological replications. However, we plotted $F(R_N)/F_0(R_N)$ in Fig. 3a. We felt that the plot of $F(R_N)$ alone in this Suppl. Fig. 2 was redundant and unnecessary. Therefore, the modified manuscript version no longer contains this plot.

Reviewer #1 (Remarks on code availability): The authors state that the code is uploaded but I do not see where.

We do not understand why the reviewer was not notified and provided access to Nature-supported external repository (Code Ocean). To avoid further confusion, we have now included the code directly in the submission as a supplementary file.

Reviewer #2 (Remarks to the Author):

The authors have addressed all my previous concerns satisfactorily.

We are very glad to hear this. Thank you again for the constructive criticism and comments.

I only have few minor comments for the authors considerations before publishing.

1) The breakpoint is still not clear to the wide readership from the explanation provided in the manuscript. I think it would be beneficial if the authors could add a concise description of the physiological state of the bacteria at the breakpoint concentration. Similar to the MIC, which is the minimal concentration with no observed culture growth.

We acknowledge the reviewer's point that a breakpoint is not a well-defined parameter.

Conceptually, a breakpoint is the threshold concentration of an antibiotic used to define whether a bacterial strain is likely to be treatable by a given antibiotic or not in a patient. However, the process for setting breakpoints is often unclear. As described in the reference [29], breakpoint setting requires “the integration of knowledge of the wild-type distribution of MICs, assessment of the pharmacokinetics/pharmacodynamics of the antibacterial, and study of the clinical outcome of infections when the antibacterial is used.” Notably, two major breakpoint-setting organizations, the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical and Laboratory Standards Institute (CLSI) recommend different breakpoint values for some antibiotics, indicating there is not a single straightforward procedure to determine the breakpoint values. The values can change over time as more studies are conducted. The minimum inhibitory concentration (MIC) represents the concentration at which bacterial growth and death are balanced, such that the population size as a whole ceases to change (this was extensively described in our previous paper, PMID: 29508699). In practice, the relationship between the MIC and the breakpoint is that a bacterial strain is categorized as susceptible if $MIC < \text{breakpoint}$ and resistant if $MIC \geq \text{breakpoint}$.

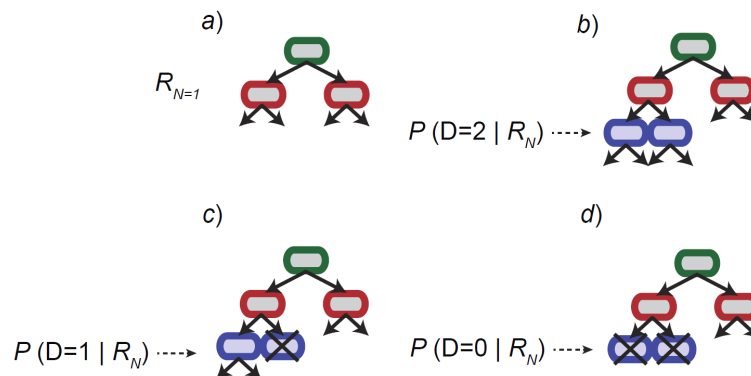
We added this discussion in Supplementary Text 1 and added a pointer to it in the main text (L82).

2) In line 49 of the introduction, the authors review mechanisms of antibiotic resistance, but do not mention epigenetic adaptation, which is different from what they study in this manuscript. I think this should be added here. See for example: Adam et al, BMC evolutionary biology 2008 (DOI: 10.1186/1471-2148-8-52).

This is a good suggestion. We added the following statement “Epigenetic adaptation, potentially mediated by mechanisms akin to DNA methylation observed in mammalian cells, could also contribute to resistance ^{3,4}.” in L53.

3) In the paragraph from line 175 to 181, I was a bit at loss. With $P(D=2/R_N=1)$, do the authors mean the probability of two daughter cells exhibiting robust survival conditioned upon one prior generation of robustness, or do they mean regular survival? From the sentence, line 177 – 179, I understood that it’s regular survival, which does not make sense because it is the requirement for the previous generation to be considered robust, but by the end of the paragraph I gathered that they mean robust survival. I think it should be made sharper.

Sorry for the confusion. We meant the former (the probability of two daughter cells exhibiting robust survival conditioned upon one prior generation of robustness). To make this clearer, we re-phrased the description (a paragraph starting in L183). We also added Supplementary Fig. 2 (re-plotted below), which graphically explains $P(D=2 | R_{N=1})$. With this changes, we believe that its meaning should be clearer.



4) Typo in line 242, “... deltatolC strain that WT...” should be “...deltatolC strain than WT...”.

Corrected

5) In line 284, the authors state that the expression of TolC is stochastic. What does that mean? Aren’t all genes expressed stochastically? Do they mean that it’s amount vary among cells, which is the case for all proteins? Or do they mean something else? If that’s what they mean, I don’t think it should be stated as if it is a unique property of this gene.

We meant that its amounts vary among cells. We agree with the reviewer’s perspective. We

incorporated this comment by modifying the text to “Interestingly, recent studies found that the

heterogeneous expression of TolC, as well as of Acr, results in different response of cells to antibiotics” (L342).

6) Related to the previous point, in line 313 – 316, the authors state that experiments using a promoter that would enable precise control of efflux pump expression pattern, would be useful in the future. Is that even possible? don’t all genes exhibit similar variability patterns but different average levels? See for example: Taniguchi et al Science 2010 (PMID: 20671182).

We agree that precise control of efflux pump expression is an overstatement. As the reviewer pointed out, all genes show some degree of variability in their expression. However, there are elements that can amplify or decrease the variability, e.g., adding negative or positive feedback regulation (as referenced in ref [46]. We changed our text to address this nuance in L347 “the designs of efflux gene promoters can be altered to increase or decrease the heterogeneity in gene expression⁴⁴. Future studies could benefit from using such promoter”

7) In line 300 – 301, the authors claim that the robustness is strengthened across the generations is an indication of phenotypic adaptation to antibiotics. To me it seems that this could be random but inherited, and over time it is manifested more. In the model they present also, there is no adaptation parameter, so I don’t understand why they attribute this to adaptation. If I’m missing something, this point should then be better explained, or rephrased.

The reviewer’s insight is correct. As the reviewer pointed out, our claim is that phenotypic resistance is “random but inherited”. We realized that we used the phrase “phenotypic adaptation to antibiotics” (L322). We deleted this phrase and added the following sentence “This robustness can be inherited from generation to generation”.

Reviewer #2 (Remarks on code availability): There was no change in the code itself in the revision, but I did review the description of the code in the supplementary text and found it to be good and sufficient.

Comments to authors (NCOMMS-24-12302B-R2)

Reviewer #1 (Remarks to the Author):

We were informed that a mitigating referee was invited to evaluate the review process. This referee provided specific guidance on the *Data availability* and *Statistical analysis* sections, and recommended that we address Reviewer 1's comments on *Mathematical model* and *Minor comments*. The editor instructed us to follow the mitigating referee's suggestions, and we have done so in the point-by-point responses below.

I appreciate that the authors have put more work into this manuscript and added greater description of many steps. However, there are still some issues remaining.

Responses to my previous major comments:

- Data availability: The authors have now provided their data, but it is not in a form that could be readily interpreted or re-used by others. At a minimum, there would need to be an accompanying explanatory/meta-data file or text section explaining what the column headers and abbreviations mean and what is indicated by red text and shaded cells in all spreadsheets. Source Data 1 would additionally need explanation of how numerical quantities are calculated, as these are simply entered as numbers without linking to the data, formula or code used to calculate them.
- Code availability: The authors have provided their code files along with an explanatory file 'Code Index'. I have not tested the code but at face value it looks reasonable.

⇒ The editor instructed us to follow the mitigating referee's comments on Data availability.

- Mathematical model: This has been improved and several issues from the previous review addressed, but some problems remain.

o The new description of the approach as a population model rather than a Markov model, and the additional details, are generally helpful. However, there are problems with defining the parameters and a few errors remaining.

§ P, Q, and K have only vaguely been described as 'rates' (above Eq. 8-10). This is ambiguous. They should more clearly be defined as the probability that each daughter cell (independently) takes on a given cell state, or the proportion of daughter cells taking on each state.

⇒ The reviewer is correct. We changed it as suggested.

§ The wording around what it means for a cell to be killed (versus its daughter(s) to be killed) is unclear. It might help to relate this definition back to what is called 'robust' survival elsewhere.

⇒ We agree with the reviewer. Survival of a cell refers to the case where the cell endured antibiotic treatment and divided. "Kill" refers to the case where the cell lysed. The cells that survived the treatment would give birth to two daughter cells. Only when these two daughter cells divide (Sup Fig. 2a), the survival of the original cell would be considered robust survival. We changed the wording above Eq. 2 and below Eq. 7 to reflect these meanings.

§ The notation K has been re-used with two different meanings. In the first part (Eq. 1-7), it corresponds to the fraction of offspring from all live cells that are killed. In the second part (Eq. 8 onwards), it corresponds to the fraction of offspring only from vulnerable cells that are killed; offspring of healthy cells are never killed. This results in inconsistent equations: Eq. 7 yields $K = 1 - \mu$ whereas Eq. 14 gives a more complicated expression.

- ⇒ The reviewer is correct. It is because we are slowly building our model from the simplest case to more complex scenarios. To clarify this, we made a clear distinction in our model development, dividing it into three sections. Furthermore, we added a statement below Eq. 14 “The mathematical expression of K is no longer equal to $1 - \mu$ (see the section 2) because we assume that only cells in one state (V) die.”

§ Below Eq. 14, μ has been described as ‘the steady state growth rate’, which is incorrect; $\lambda = B - \mu$ has previously been defined (below Eq. 6) as the steady state growth rate.

- ⇒ The reviewer is correct. We changed it to “The μ in Eq. 13 denotes a probability of cell division.”

§ It is unclear why the first part of the modelling (two-type population model, Eq. 1-7) is needed at all: where is it subsequently used?

- ⇒ We realized that the original description of our model was difficult to understand for biologists with limited math backgrounds. We therefore began our description from a simple model (two-type population model), which many biologists are familiar with. We then built the complexity around it. In short, the first part is to set the stage for our complex model.

§ There is an error in Eq. 8-10: every term should be multiplied through by B , consistent with Eq. 11.

- ⇒ The reviewer is correct. It is corrected now.

o The added explanation of the experimental protocol and previous reference to support the steady state assumption is satisfactory.

o The added description of the fitting is adequate, except for:

§ the unclear statement “ χ^2 value was calculated to choose... the best fit”: the χ^2 value of what?

- ⇒ χ^2 is defined as $\sum \frac{(C - C_{cal})^2}{C}$, where the sum was performed for all nine relationships shown in Fig. 2bc. This definition is added where χ^2 was introduced in the Supplementary text.

§ “the nine measured kin correlations given equal weight” – this seems problematic as there is a much larger sample size for some kin relationships than for others; this should at least be discussed

- ⇒ Reviewer is correct. We added the following discussion “, all nine measured kin correlations were given equal weight, although the number of samples decreases

for more distant relationships. Yet, we observed a good agreement between the experimental C (horizontal bars) and calculated C_{cal} (line) in Fig. 2bc.”

- Statistical analysis: Unfortunately, this part of the work still falls short of expected standards.

⇒ The Editor instructed us to follow the mitigating referee’s comments on statistical analysis.

- o The authors have now provided the names of tests and the p-values, but are missing other elements. Specifically, referring to the Nature Portfolio Reporting Summary, the authors have not provided the test statistic (e.g. t value in a t-test) or the degrees of freedom in the test. A more coherent way of reporting would be to provide a table where the results to be tested (e.g. correlation coefficients), number of replicates, test statistic, degrees of freedom, and test output (p-value) are provided side-by-side. In addition, the authors have not addressed the issue of multiple comparisons, e.g. in Suppl. Table 2 there are 18 t-tests performed, so on average one would expect ~1 of these tests to reach the $p=0.05$ threshold by chance. The authors stated that they were not sure what I meant by multiple comparisons, but this is a standard statistical issue and approaches can readily be looked up.

- o I still have some questions as to whether statistical tests have been chosen appropriately, but it is beyond my expertise to state definitively whether there is a problem or to suggest better alternatives. I will flag this limitation of my expertise to the editor. My specific concerns include:

- § t-tests (e.g. on calculated correlation coefficients) are based on 3 biological replicates. It seems potentially problematic that the underlying sample sizes (i.e. number of cell pairs) are not factored in, and any errors in the estimation of correlation coefficients are not propagated forwards. Similar issues apply for the statistics reported in Fig. 3 and Suppl. Figs. 4-5.

- § The same cells will be included in multiple relationship pairs, suggesting that the correlation coefficients are not independent of one another. Is this a problem for any of the statistical tests?

- § Fig. 3a-d: I am unfamiliar with the process of discarding replicates where the sample size is too low and uncertain whether this approach is valid.

- o The analysis reported for Fig. 2d is unclear.

- § L496: How was the slope of the increase estimated? What do the +/- errors in the reported slope represent?

- § L498: Why $n=21$? Note that data from different generations cannot be described as ‘replicates’.

- § L504: Why was a heteroscedastic t-test chosen here?

- o It is still unclear how replicates are combined for the statistical tests associated with Fig. 3a-c. For instance, each replicate will have its own value of $P(D=2)$ along with its own $P(D=2|RN)$. So what exactly is tested in the one-sided t-test in Fig. 3b / Suppl. Table 4?

- o In Fig. 3a-d, the authors variably use either a one-sided or two-sided t-test to test for significance of the correlation coefficients (Pearson’s r). Why isn’t this choice consistent?

- o The authors have stated in their response that they did not perform statistical testing on the results in Suppl. Fig. 3 “because no claims were made from this dataset” – which raises the question of why these results were presented? I strongly suggest that the authors present statistical tests of all quantitative results included.

Other minor comments:

- *The abstract should mention which antibiotic was used (around L21).*

⇒ Added

- L78 cites ref. 28, but this reference did not actually use branching processes as stated. (This reference could however be cited more accurately in L73.)

⇒ We removed ref 28 in L78. We then added “which not only complicates antibiotic treatments but also alters the evolution of resistance”, and then cited it (now ref 25)

- L92 it would be helpful to state that *tolC* is an efflux pump.

⇒ We changed the sentence to clarify it “ Given the well-known role of efflux pumps in antibiotic resistance ³¹, we inactivated them by knocking out a major component of efflux pumps, *tolC*”

- L192 typo: $0.05 \leq p\text{-value} < 1$ should presumably read < 0.1

⇒ Following the mitigating referee’s comment, we removed this sentence.

- L242 missing a word? – “*the probability of daughter cells*”... *surviving*?

⇒ The reviewer is correct. We changed it to “we can determine the probability of daughter cells surviving or dying given...”(L232).

- L345 “a high rate of recovery from antibiotic damage” – clarify, this is in the WT

⇒ The reviewer is correct. We added WT in the sentence (now L336)

- L410-420: Why is the background correlation C_B calculated differently for each kin relationship?

⇒ The background correlation C_B is slightly different for each relationship because the samples are different. Cells get included in the calculation when they have relatives of the correct type whose fates are known. For example, we don’t know whether cells that are born close to the end of the movies successfully divide or not. Therefore, each possible kinship type that we can examine will have a somewhat different set of sampleable cells; since background correlation is calculated for these cells, that gives a slightly different C_B . In practice, C_B is too small to really affect the conclusions, but is included anyways out of an abundance of caution.

- Around L438-440: The Shannon entropy calculation is still unclear.

o I’m confused by the choice of cells at each generation N : For $N=3$, why do you go back as far as 4 generations, which isn’t illustrated in Fig.2a or included in Fig.2bc? Where do cells with the relationship [3,3] come in?

o It still isn't clear what the summation will look like for $N=2$ and beyond, where more than one relation Y needs to be considered.

⇒ We apologize for the confusion. To address, we graphically labeled cells considered in conditional Shannon entropy calculation in Supplementary Fig. 7 and provided equations in its caption. We believe this should clarify it.

⇒ [3,3] is not included because it is in the same generation as a self. It is not ancestors. We believe this is clear now with Supplementary Fig. 7.

• Fig. 3c: Please double-check the correlation coefficients reported in the caption: It would appear that the decline is steeper for DtolC than for WT.

⇒ We double-checked and confirmed that they are correct. It might appear so because data for DtolC are more spread.

• Fig. 4b caption: The test/interpretation is unclear; please clarify that you are evaluating differences between WT and DtolC here.

⇒ Corrected.

• Suppl. Fig. 2b-d: the annotations should indicate $RN=1$, similarly to panel a.

⇒ Corrected.

• L513: Still references Suppl. Fig. 2e, which has been deleted.

⇒ Currently, Suppl. Fig. 2e refers to the table. To avoid confusion, we added "See Supplementary Fig. 2e Table"

• Suppl. Fig. 2e caption: Clarify, is $\langle X \rangle$ calculated across all cells in all generations

⇒ Reviewer is correct. $\langle X \rangle$ is calculated by considering the fates of all cells. The caption was modified to clarify this.

It has been over a year since this manuscript was first submitted, and both my student and postdoc are eagerly awaiting its publication as it is central to their upcoming career transitions. We sincerely hope the Reviewer finds this revision satisfactory.

Reviewer #3 (Remarks to the Author):

I have specifically been asked to comment on the comments of reviewer #1 to suggest which points are necessary for the authors to change.

We are thankful for the mitigating reviewer's expertise and guidance.

Data availability:

-I agree with the reviewer that an explanatory file for columns and abbreviations, and red text/shading is needed to meet reproducibility criteria. I do see some definitions in the methods section, but it would help to have these together with the data.

- I do not think that formulae need to be added to source data 1. Most if not all of the calculated quantities are standard or provided in the methods.

- Overall while I think there is room for improvement here, I think this looks to be up to the usual standards of data availability, if a description of the headers/fields is added.

- I would encourage the users to consider whether there is some raw/underlying data from their experiments they could share. The correlations are useful, but computed from other data. To maximise reusability, the data from which these were computed could be made available.

Thank you for the clear instruction. Two key points are 1) it is “up to the usual standards of data availability, if a description of the headers/fields is added,” and 2) raw/underlying data need to be provided. To address these points, we have added a tab entitled “**Data description**” to the *Source Data 2* Excel files, which explains the meaning of each column header, abbreviation, and red text/shading. The descriptions in each cell of this tab apply to the corresponding cells across the other tabs in the file. Red-highlighted values indicate data points used in the main text. Additionally, we have provided raw/underlying experimental data in the file “RawsixPlateData.mat”.

Code:

Available, in MATLAB which requires a license, but this is acceptable.

Mathematical model:

The reviewer's suggestions are helpful and seek to improve the clarity of the work with specific suggestions on how to do so. It would be useful for the authors to respond to these suggestions and incorporate them where possible. These all appear to be minor changes.

We agree. We responded to these suggestions and incorporated them in the revised manuscript.

Statistical analysis:

My overall summary here is that this debate is going too far down the dogmatic frequentist route in a manner which is becoming unhelpful for the paper. In most of this debate we are only comparing three data points as far as I can tell -- a formal statistical test has very little power here. We can make some simpler and more consistent choices to tidy this all up.

- The reporting summary does indeed state that the test statistic is required. Personally I do not feel this is necessary here. This point is up to the editor in my view.

- Specific frequentist statistical approach.

* Rather than using t-tests on C_B , just use the p-value of the slope of the regression these coefficients are calculated from. This can easily be done using e.g. `lm()` in the R programming language, or many online tools.

We appreciate Reviewer's comment. As suggested, we performed linear regression analyses on the survival parameters (X) between related cells and calculated corresponding p-values. They are shown in Supplementary Table 1. A pointer to this table was added in Figure 2 caption.

* Figure 3: Firstly, this caption is far too long and should go elsewhere (methods/results).

I think the correct test is either chi-squared with the counts or z (normal) with the proportions, rather than a t-test.

But honestly I don't think this point really matters that much to the paper overall. If the authors just plot the ratios with some error bars we can see better for ourselves. As there are only three replicates, plotting the points for each of these would be ideal!

So I would suggest removing the formal statistical tests, their overly long description (or at least relegating all of this to the supplement), and just plotting the data and its variation clearly.

Thank you for this clear guideline. Following the Reviewer's suggestion, Fig. 3 now includes not only the means (horizontal bars) and standard deviation (error bars), but also the points for each replicate (small open circles). We additionally moved the long description of statistical tests to Supplementary text 3. Please note that in Fig. 3d, each replicate point is not available because information content depends on sample size; thus, we calculated one information content for the combined dataset (per strain and per relationship). We did include error bars (i.e., maximum uncertainty of information content that was theoretically calculated).

* Now going back to figure 2. We once again have a tortured legend with a very complex statistical procedure. I have exactly the same comment as I did with figure 3, just forgo this and give us the plots. You could also add the p-values on the slopes of the regression, but it's probably not necessary. To compare WT and tolC mutants I think a regression model with interaction terms might be possible, but I am again unconvinced it's really needed.

Following the Reviewer's suggestion, we moved the statistical test description to Supplementary text 3 and plotted the points for each replicate (small open circles indicate data from each replicate. The horizontal bars indicate their means. The error bars indicate standard deviation).

- Multiple testing. I think there is a confusion in language between reviewers and authors which has led to a misunderstanding. Please see https://en.wikipedia.org/wiki/Multiple_comparisons_problem for what the reviewer is referring to. I would suggest adding an adjusted p-value column. In the R programming language, one can use `p.adjust()` with a list of p-values to adjust them. I would recommend setting `method = "fdr"` and choosing a suitable false discovery rate. But, as suggested, these should probably be removed. (as a side note, if we really did want to do all of this properly, I'd suggest going with a simulation and bootstrapping approach for all the derived quantities -- but it's not going to add anything given the amount of data)

Following the Reviewer's suggestion, we removed this part.

Minor comments:
Should all be addressed

They are addressed as advised.

Comments to the Authors

In this interesting study, Stine et al. build on the growing body of recent work demonstrating cell-to-cell variation in antibiotic susceptibility, by analysing how susceptibility is correlated along lineages within isogenic bacterial populations. To do so, they analyse data from time-lapse microscopy imaging of *E. coli* under cefsulodin treatment and additionally develop a 3-state Markov chain model of cell states that can capture history-dependence within lineages. I believe this modelling approach is novel in this context and could be applied more broadly to similar data in the future. Overall, the authors' approach made sense, but their results critically need statistical analysis in order to draw solid conclusions. The paper would also benefit from a more detailed explanation of the modelling approach and provision of the associated code. The writing was mostly clear with some minor exceptions noted within my detailed comments below.

Major comments:

1. The results currently lack statistical analysis. Adding statistical testing is critical to back up any statements comparing outcomes with a baseline expectation or comparing strains, generations, etc. These statements were particularly apparent in the Results section "Enrichment of robust lineages", e.g. that "this frequency was much higher than the frequency expected by random chance" (L175-6) or "We observed a decreasing probability of the termination with each additional generation of robustness" (L191-2). Essentially, all the results shown in Figures 2-3 and Suppl. Figures 2-5 need statistical testing to support the conclusions.
2. I'm not following the reasoning behind the verbal model outlined on L200-208 re: "two potential mechanisms for the observed propagation of robust lineages". The definitions of each mechanism were vague; are there possible molecular mechanisms in mind and are these presumably heterogeneous across lineages? What exactly are the predictions arising from each proposed mechanism? (The conclusion that the data supported one but not the other mechanism was not obvious to me.) Perhaps more importantly, however, this verbal model does not seem to line up with the mathematical model presented later (Fig. 4a). The transitions between 'Healthy' and 'Vulnerable' states in the latter seem to align more intuitively with accumulation/decrement of cellular damage, the mechanism that was rejected in the verbal reasoning. I'm not convinced that the earlier verbal model is needed at all, but if it's kept, it needs a clearer explanation and consistent interpretation compared to the later mathematical model.
3. The Markov chain modelling approach is nice, but would really benefit from some additional explanation, especially for less-specialized readers. I think each of the following points can be easily addressed, but would collectively make a major improvement.
 - a. Previous modelling approaches have alternatively been called "branching processes" (Intro, L75) or "two-state Markov chain model[s]" (Results, L232). Please clarify the connection, i.e. that these are the same thing; a branching process tracks live cells, while the two states in the Markov chain are (I think?) live and dead cells. It would also be useful to explicitly state the key assumption of a branching process – i.e. that the fate of each individual is independent and identically distributed – which in turn implies the lack of kinship correlation (L234) in previous models.
 - b. When a mother cell produces two daughters, do both daughters take on the same state or does each daughter draw its state independently according to the transition probabilities? (Clarify around L243 and again in Supplementary.)

- c. The state change is upon production of daughter cells, not during the mother's lifetime. This was stated in the Supplementary but should be clarified in the main text (around L245 – it sounded like recovery of the mother cell then birth of daughters were two steps).
- d. Add more detail on the fitting method – what algorithm, in what software? It would also be helpful to number the equations in the Supplementary Text and refer to exactly which equation is being fit to the data and hence plotted in Fig. 2b-c.
- e. Please deposit the code used for the model fitting.
- f. For additional clarifications on the mathematical presentation, please see “Minor comments, supplementary text” below.

Minor comments, main text:

1. The experiments were performed with only one antibiotic, cefsulodin. This should already be mentioned in the Abstract (e.g. L19).
2. L57: MIC is a population-scale measure, but not technically a population average. Rather, it tends towards the upper end of the distribution of antibiotic tolerance within the population, since growth from only a few cells will be sufficient to see population growth.
3. L74-5: Ref.25 does not actually use a branching process to model cell growth and death – only the ultimate outcome of ‘establishment’ is considered as a random event (agnostic of underlying dynamics).
4. L76: I find the wording “rapid population fluctuations” confusing. Stochasticity in cell division/death will introduce variation in population dynamics amongst lineages, but not necessarily sustained up-and-down fluctuations, nor are the changes in population size necessarily ‘rapid’.
5. Around L84: It would be helpful at this point to mention the experimental set-up in which cells are growing – i.e. as micro-colonies on solid media, as opposed to say microfluidics devices.
6. Results: I would have found it helpful to introduce the ΔtolC strain earlier on and explain why it is being used, since it already appears from the start in Fig.1.
7. L101-2: Please clarify what you are testing (re: “the basis of this [branching process] modelling”) – is it the assumption of independent and identical fates amongst individuals?
8. L123: Does the pairwise Pearson correlation analysis account for the fact that not all pairwise relations are available to analyse? For example, the 1st cousin will not exist if the aunt died and therefore we do not have a value of X (survival) for the 1st cousin. Does this bias the analysis towards positive correlations if only descendants of surviving cells are analysed?
9. L131-2: I would be surprised if “the breakpoint concentration of cefsulodin immediately killed cells” – surely there is some time delay? Did you mean, kills (close to) 100% of cells? Please rephrase.
10. L138: “Having analyzed the genealogical tree horizontally” – this wording was confusing because the previous analysis did in fact consider cells in different generations.
11. L141 and elsewhere: $[X]$ is not standard notation for the mean. Not essential, but it might aid the reader to stick to more standard notation of $\langle X \rangle$ or $\bar{X}$.
12. L155-165: I found this paragraph hard to follow and I think it would benefit from a few wording tweaks for clarity.
 - a. L156 – “To further validate this, we characterized cell survival along the lineages” – this statement confused me because I thought you had characterized cell

- survival along the lineages already in the previous section. Can you be more precise?
- b. Please clearly state the definitions of D and R , and also preferably define N before (rather than after) it is used.
 - c. L162-3: “more frequently than expected by random chance” – How is the null expectation derived?
13. L169: I’m not sure what you mean by “left column” in Fig. 3b.
 14. L197: less likely than what?
 15. L220: Please define “ancestral survival” (Y_N).
 16. Around L229 – end of section “Enrichment of robust lineages”: What about the results for the ΔtolC strain?
 17. Fig. 2b-c: The connected line suggests an ordering of the relationships on the x-axis. Is there a particular reason behind the ordering of relationships with equal total ‘distance’ (e.g. aunt and niece)? If not, perhaps rethink the data visualisation so that relationships of equal distance are grouped together or otherwise indicate that there is no expectation of decreasing correlation coefficient within the group.
 18. L273: Fig. 3c does not clearly support the statement that there is “stronger lineage-dependent dynamics of robustness... in the ΔtolC compared to WT”, but statistical testing (Major Comment 1) should clarify this point.
 19. L264-274: This is a nice explanation of the model estimates and seems plausible. How would you test these hypothesised molecular mechanisms experimentally? I think this would be a useful addition to the Discussion.
 20. L291-2: I’m not convinced that a “healthy state” *per se* is the “minimum ingredient” to explain the data; isn’t the fact that there are three states the critical point for producing correlations within lineages?
 21. L306: What is the “resistance state”? Does this correspond to a specific state in the Markov chain model?
 22. L340: Cells are counted as dead if they are lysed or ‘permanently’ arrest growth. An unavoidable limitation is that growth arrest is only observed for the limited time span of the experiment. Is there a risk of bias in that later-arising cells are more likely not to have resumed growth (yet) by the end of the experiment, and thus more likely to be counted as ‘dead’?
 23. L362: Please clarify how you calculate “the probability of each microstate” – is it the proportion of pairs of daughter cells exhibiting each outcome?
 24. L364: Define Y ; similarly around L366, remind the reader what X is.
 25. Author contributions seems to be missing a mention of who carried out the experiments and who designed/analysed the mathematical model?
 26. Fig. 2 caption:
 - a. L401-2 wording clarifications: open/solid circles → symbols (there are both circles and squares). Columns → error bars (I think?)
 - b. L401-2 refers to “raw data from three biological replicates”. Not really ‘raw’ data as you are reporting a summary statistic (mean of X), right? Please mention how many cells were included per biological replicate. This might help explain why the mean for ΔtolC at ages 1-3 does not appear to fall where it should amongst the 3 replicates if all were equally weighted.
 27. Fig. 3 caption:
 - a. L412: “ratio” would be clearer than “relative difference” (the latter suggests a subtraction rather than a division)
 - b. L418: $P(D=2) = X^2$. Should this be $[X]^2$ (mean of X , squared)? Mean over all cells in all generations?

28. Fig. 4 caption, L425: “The previous two-state Markov model cannot explain the survival correlation” – this is confusing as the figure does not show the two-state model. This statement rather belongs in the associated main text, not the figure caption.
29. Throughout the figures, wherever a probability or proportion is plotted on a linear scale (Fig. 2d, 3, Suppl. Fig. 3-5): I suggest starting all y-axes at zero. Starting above zero can be unintentionally visually misleading as it exaggerates the magnitude of differences between bars, and makes it harder to directly compare different plots.
30. Fig. 4a: Perhaps add arrows from Healthy back into Healthy state, and likewise for Vulnerable, indicating that it’s also possible for a cell (or rather its daughters) to stay in the same state.

Minor comments, supplementary text (model description):

1. Around the first paragraph, add a couple of clarifications:
 - a. Is this a discrete-time Markov chain? (In this case, it would be helpful to mention that ‘rates’ are transition probabilities in each time step.)
 - b. Does each daughter cell take on its state independently, or do both daughter cells take on the same state?
2. Note that D , R refer to different things in the main text (surviving daughter cells and robust cells, respectively) than they do here (damage and recovery rates). D_n is also used below to mean number of dead cells in generation n . It would be helpful to choose distinct notation.
3. Please define the order of states in the rows and columns in the transition matrix T , and specify whether entry T_{ij} corresponds to the transition probability from state i to j or vice versa. Note that T itself does not need to be written within square brackets.
4. Where T is defined (in both the 3x3 and 4x4 matrices), both the [1,2] and [2,1] entries are R but I think one of these should be D .
5. Please clarify what happens to dead cells in the 3x3 version of T . The definition of transition probabilities in the matrix ($T_{3,3} = 0$) seems to suggest that dead cells “disappear” after one generation – so would it be more accurate to define D_n as “newly dead” cells in generation n ? I think these entries cannot then strictly be interpreted as “transition probabilities” since from state D the cells just disappear rather than transition (or stay) anywhere, i.e. the 3rd column of the matrix does not sum to 1.
6. Bottom of 1st page: What is the solution to this eigenvalue problem? (Or if the solution is not important, why bring up the eigenvalue problem at all?)
7. Where the 4x4 matrices are defined (2nd page): Please explicitly define the vector $[H, V, D, NB]$ and note that only one of these entries can be 1 (the rest 0) in order to interpret the outputs as probabilities of the i^{th} descendant falling into each state. It would be useful to have distinct notation for the vectors on the left-hand side of the equations to indicate the initial state.
8. How are the “elder” and “younger” cells defined if two cells are in the same generation?
9. Page 3: “In order to find the Pearson correlation coefficient” ...of what?
10. Please define the notation f_H , f_V , and b_{total} etc. (What does the ‘total’ subscript mean? A population average, weighted by initial frequencies?)
11. In the definition of the correlation coefficients:
 - a. What constitutes a ‘datapoint’? A pair of cells?
 - b. The index i runs over pairs of cells, presumably each with the same relationship?
 - c. Define $\bar{z}$ and $\bar{w}$
 - d. If using z and w for the datasets, then also write C_{zw} rather than C_{xy} (though arguably x and y would be better, as this notation was used elsewhere for survival indicator random variables)

12. Page 4: Note that the rewriting now assumes z is the “elder” cell and w is the “younger”
13. Please add a description of how the model was fitted to the kinship correlation data.
Which equation was fitted? How? (Least squares? Some other optimisation algorithm?)

Minor comments, supplementary figures:

1. Suppl. Fig. 1:
 - a. Perhaps the issue is on my end, but there appears to be a rendering error in the pdf. The axis labels, legends, and panel title on the left panel are missing.
 - b. “100ul media was transferred...” – “(diluted) culture” rather than “media”?
 - c. “The experiment was repeated twice” – why not show both replicates?
2. Suppl. Fig. 2:
 - a. In the tables, “Observed frequency” is unclear – rather, “Number of robust cells”?
 - b. In the caption, define the white bars
 - c. How are the “frequencies expected by random chance” calculated? (Perhaps this belongs in Methods rather than the figure caption.)
3. Suppl. Fig. 4 and 5: In the captions, define the dashed line on the plots.

After reading the revised version of the manuscript, I maintain my view that this study is interesting and contains worthwhile ideas, data, and approaches. However, now that the authors have provided some more details of their methods, unfortunately I see a couple of major issues with the quantitative approaches:

1. The mathematical model is not well-defined and the analysis appears to contain a few errors. I think this could be fixed, but would require re-doing the fit to data (with possibly different results). The calculations also make a strong assumption about the cell population starting in steady state.
2. The statistical analysis, which has now been added, is not sufficient. Not all of the key results have been tested statistically, and where tests have been carried out, not enough information has been provided (e.g. sample numbers) and in at least one case the choice of test does not appear suitable. Moreover, the results of statistical testing generally are not mentioned in the main text, and do not uniformly support the headline conclusions. Negative results should not inhibit publication, but need to be clearly stated and discussed. Overall, I believe the statistical analysis and associated presentation of results would require substantial revisions.

Please see my detailed comments below for further explanation of these points.

Furthermore, there is still not enough information provided to ensure reproducibility and re-usability of the authors' analyses.

- The data file provided only contains the numerical values to be plotted in each figure, not the raw data or calculations. It is not possible to work out sample sizes in each group (e.g. generation or kinship pair), confirm which equations have been applied where, or reproduce the analysis.
- The code availability statement says that "Our custom-built software is uploaded" but does not say where, and it does not appear to be included with the files for review (apologies if I have overlooked something). I understand that some elements of analysis (e.g. cell tracking) may have been developed previously, in which case previously published work can simply be cited; but other analyses here, e.g. fitting the mathematical model to data, are novel and require provision of code.

I'm sorry that I cannot be more positive about the manuscript in its present form, but encourage the authors to revise their work accordingly and would be enthusiastic to see a more rigorous approach to the interesting questions they address.

Detailed comments on mathematical modelling

- Just as the 3x3 matrix T is not technically a transition matrix, what the authors describe as a "Three-State Markov Model" (first section of Suppl. Text 1) is not actually a well-defined Markov chain. A collection of possible cell states (phenotypes) does not automatically translate to "states" of a stochastic process in the mathematical sense. Now that the authors have explained that each daughter

cell takes on its state independently, I think I can see where the problem lies: there are two daughter cells' phenotypic states to consider, and the authors have not clearly defined how this translates to "state" of the Markov chain, i.e. what exactly is the random variable they are tracking? Moreover, a Markov chain's state space should include all possible states of the system, so that the process does not simply terminate (it may land in an absorbing state with transition probability of 1 to remain in that state, but should always be somewhere in the defined state space). Therefore the fact that dead cells simply 'disappear' after one generation (reflected by the final column of all zeros in T) disqualifies this model as a Markov chain.

- In contrast, the 4-state process introduced in the following section *could*, I believe, be interpreted as a Markov chain if the state of the process is defined as the phenotypic state of an individual cell, as you follow a lineage through generations (independently stepping to one or the other daughter cell in each time step). This process should be more rigorously defined. Under that interpretation, I think the following correlation analysis is okay.
- As far as I understand, the point of the first section's analysis was, ultimately, to find the net growth rate of the (live cell) population. The way this is currently done uses the 3x3 matrix T but note that the 3rd row/column (corresponding to dead cells) is not actually needed to obtain Eq. [4]: I can re-derive this result using only the first two equations for H and V . However, because the 3-state process is not well-defined, I think the matrix entries are not quite right for obtaining the average population growth rate. An alternative, perhaps more intuitive, way to do this would be to consider the population with a 2-type branching process tracking healthy and vulnerable cells. Then use the 2x2 matrix of expected numbers of offspring in the eigenvalue analysis. Note that the matrix entries need to account for two daughter cells each independently taking on a given cell state. This does not exactly line up with the entries in T , which only reflect the probability of one daughter taking on a given state.
- If I'm correct on the above, Eq. [4] should change. Since this equation was used in fitting the experimental data, the fits would then need to be re-done, leading to different estimates of rates P , Q , and K as presented in the main Fig. 4 (and Fig. 2bc). It remains to be seen how much this may change the conclusions.
- In describing the model fitting procedure (end of Suppl. Text 1), the authors now explain that they use the eigenvector from Eq. [3] to determine the fraction of progenitor cells that are healthy/vulnerable. However, the eigenvector would give the steady-state proportions expected to be approached in the long-term in a growing population. I would not expect progenitor cells necessarily to be near steady state. Are there features of the experimental design that could justify this assumption? How sensitive are the results to this assumption?
- The added description of model fitting still is not entirely clear. For instance, Eq. 9-12 and 13-16 cannot directly be used to "find" the stated probabilities; rather, they express certain probabilities in terms of P , Q , and K . Then the paragraph stating "Fitting was performed using MATLAB software..." doesn't specify exactly what was fit. What I *think* the authors are doing (but I'm guessing here) is to use Eq. 17 or 18 to compute the observed correlation coefficients from the raw experimental data

(survival status, $X=0$ or 1 , for each individual cell) and then equate this to Eq. 25, the theoretical expression derived from the model (which would be clearer if Eq. 25 were written out in terms of P , Q , and K). Please clarify. Also, please make the relevant code available.

Detailed comments on statistical analysis

- Throughout: please provide more details of the data and tests being used.
 - Make sure it is clear which subset of data is being analysed, with reference to the deposited data file(s): as mentioned earlier, I believe the data deposit should be extended to raw data.
 - What is the sample size in each group (e.g. generation or kinship pair), in each experimental replicate, considered in each analysis? Some of this information is in the table attached to Suppl. Fig. 2 but this is not clearly referenced in the statistical analyses and does not contain the complete information needed.
 - How were data from three experimental replicates pooled?
 - If using a t-test, is it one-sided or two-sided?
- Stat. test 1: There should probably be a correction for multiple comparisons here.
- Stat. test 2: Clarify wording re: what's being tested: "increase in survival *with pole age*"... "are statistically different *from each other*".
- Stat. test 3a: This does not appear to be a valid application of the Kendall rank correlation coefficient. It appears that you are simply stating that the mean estimate does increase with N (as is visually obvious) but not accounting for sample size or variability among individual cells underlying this trend.
- Stat. test 3b: Why only test generation 1 versus 3, and ignore 2?
- Stat. test 3c: Which procedure is repeated? What exactly is being tested?
- Stat. test 3d: Unclear what you mean by "the values obtained for the three robust generations" – how exactly are you pooling data? Why has this been applied to Fig. 3a, 3b, and 3d, but not 3c?
- There remain some key questions which have not yet been tested statistically. These include:
 - (Fig. 2d) Is the increase in survival with pole age significant for either strain?
 - (Fig. 3a) Is the fold-difference significantly different from 1 (for each strain, in each generation)?
 - (Fig. 3b-c) Are the conditional probabilities significantly different from the unconditional (null expectation)?
 - (Fig. 3d) Is the mutual information significantly different from zero?

- The statistical results should be mentioned in the main text, and indicated on the plots. Essentially, whenever there is a statement about a quantitative result, it should be backed up with the relevant statistical test result. Interestingly, many of the comparisons turn out *not* to be significant in the statistical tests presented so far. Therefore, the authors should be careful to ensure this is clear in the main text and conclusions are stated with appropriate caution. This may require some re-interpretation and discussion where results turned out (perhaps against expectations) to be statistically non-significant.

Other (mostly minor) comments

- In the writing, the authors have repeatedly referred to a cell's fate as "stochastic" or "random" (e.g. L106, L129, L247, L298) when what they really mean is that it is uncorrelated to previous generations. This is a colloquial use of the term "random", but in a mathematical sense, a random variable need not be independent or uncorrelated with others. For example, in L106, the aim is to test whether a cell's fate is independent of its parent's and identically distributed regardless of its lineage.
- The definition of D with respect to robust cells (L175-181) is confusing. The authors state that they are testing whether daughter cells are robust – meaning that the daughters' daughters should survive (cf. L166) – yet D is defined as "the number of daughter cells that survive" rather than the number that are robust. Perhaps the generation of daughter cells we're talking about is getting confused? Similarly, in L199, "we analyzed how likely a robust cell gives birth to non-robust cells" but "being non-robust means either one of their daughter cells die...": "their" is ambiguous; it must be the daughter's daughters? It may be helpful to label the robust cells, and how many generations of robustness are shown, in the cell lineage sketches in Fig. 3.
- L229: in $H(M|Y)$, M has not been defined
- L238-242 (newly added text): "the ΔtolC strain exhibited stronger pairwise correlations of cell survival than WT" – but from Fig. 2bc, this only appears to be true for the more distant kin relationships. Also, the reasoning that "robust lineages had a higher tendency to produce robust lineages in ΔtolC than WT" and "as a result, the ancestral history contained more information about the offspring survival in the ΔtolC strain", seems to contradict the argument just made above in L233-7 where it was stated that resistance inheritance within robust lineages does not imply population-wide lineage history dependence.
- Minor discrepancy: Cells were said to be exposed to cefsulodin at its breakpoint, given as 32 ug/ml (L97), but Methods states that the dose used was 31 ug/ml for WT cells (L349).
- L380: Where is the numerical value of C_B reported? What test led to the conclusion that it is not significantly different from zero?
- L387 and L391: indicate what the summation is taken over

- L390-4: I'm not entirely following this. L394 gives an example of Y but not the full equation for what is plotted in Fig. 3d, which gives results for $N=1, 2$, or 3 . So is H summed over all combinations of ancestral states in generation N ? Please give a complete description.
- Fig. 2: By eye, it looks like $\langle X \rangle_Z$ is larger than $\langle X \rangle$ for all Z . How is this possible?
- Fig. 3 caption (L457): "the non-conditional probability" of what?
- Suppl. Fig. 1 caption: It is not clearly stated what is plotted in Fig. 1a (mean and s.d.?) and Fig. 1b is not mentioned at all.
- Suppl. Fig. 1c: How exactly is MIC determined from plating efficiency?
- Suppl. Fig. 2 table: Does N refer to the generation the cell is in, or the generations of robustness?
- Suppl. Fig. 2 plots: What do the error bars indicate?

Comments to authors (NCOMMS-24-12302B-R2)

I appreciate that the authors have put more work into this manuscript and added greater description of many steps. However, there are still some issues remaining.

Responses to my previous major comments:

- Data availability: The authors have now provided their data, but it is not in a form that could be readily interpreted or re-used by others. At a minimum, there would need to be an accompanying explanatory/meta-data file or text section explaining what the column headers and abbreviations mean and what is indicated by red text and shaded cells in all spreadsheets. Source Data 1 would additionally need explanation of how numerical quantities are calculated, as these are simply entered as numbers without linking to the data, formula or code used to calculate them.
- Code availability: The authors have provided their code files along with an explanatory file 'Code Index'. I have not tested the code but at face value it looks reasonable.
- Mathematical model: This has been improved and several issues from the previous review addressed, but some problems remain.
 - The new description of the approach as a population model rather than a Markov model, and the additional details, are generally helpful. However, there are problems with defining the parameters and a few errors remaining.
 - P, Q, and K have only vaguely been described as 'rates' (above Eq. 8-10). This is ambiguous. They should more clearly be defined as the probability that each daughter cell (independently) takes on a given cell state, or the proportion of daughter cells taking on each state.
 - The wording around what it means for a cell to be killed (versus its daughter(s) to be killed) is unclear. It might help to relate this definition back to what is called 'robust' survival elsewhere.
 - The notation K has been re-used with two different meanings. In the first part (Eq. 1-7), it corresponds to the fraction of offspring from all live cells that are killed. In the second part (Eq. 8 onwards), it corresponds to the fraction of offspring only from vulnerable cells that are killed; offspring of healthy cells are never killed. This results in inconsistent equations: Eq. 7 yields $K = 1 - \mu$ whereas Eq. 14 gives a more complicated expression.
 - Below Eq. 14, μ has been described as 'the steady state growth rate', which is incorrect; $\lambda = B \mu$ has previously been defined (below Eq. 6) as the steady state growth rate.
 - It is unclear why the first part of the modelling (two-type population model, Eq. 1-7) is needed at all: where is it subsequently used?

- There is an error in Eq. 8-10: every term should be multiplied through by B , consistent with Eq.11.
 - The added explanation of the experimental protocol and previous reference to support the steady state assumption is satisfactory.
 - The added description of the fitting is adequate, except for:
 - the unclear statement “ χ^2 value was calculated to choose... the best fit”: the χ^2 value of what?
 - “the nine measured kin correlations given equal weight” – this seems problematic as there is a much larger sample size for some kin relationships than for others; this should at least be discussed
- Statistical analysis: Unfortunately, this part of the work still falls short of expected standards.
 - The authors have now provided the names of tests and the p-values, but are missing other elements. Specifically, referring to the Nature Portfolio Reporting Summary, the authors have *not* provided the test statistic (e.g. t value in a t -test) or the degrees of freedom in the test. A more coherent way of reporting would be to provide a table where the results to be tested (e.g. correlation coefficients), number of replicates, test statistic, degrees of freedom, and test output (p-value) are provided side-by-side. In addition, the authors have not addressed the issue of multiple comparisons, e.g. in Suppl. Table 2 there are 18 t -tests performed, so on average one would expect ~1 of these tests to reach the $p=0.05$ threshold by chance. The authors stated that they were not sure what I meant by multiple comparisons, but this is a standard statistical issue and approaches can readily be looked up.
 - I still have some questions as to whether statistical tests have been chosen appropriately, but it is beyond my expertise to state definitively whether there is a problem or to suggest better alternatives. I will flag this limitation of my expertise to the editor. My specific concerns include:
 - t -tests (e.g. on calculated correlation coefficients) are based on 3 biological replicates. It seems potentially problematic that the underlying sample sizes (i.e. number of cell pairs) are not factored in, and any errors in the estimation of correlation coefficients are not propagated forwards. Similar issues apply for the statistics reported in Fig. 3 and Suppl. Figs. 4-5.
 - The same cells will be included in multiple relationship pairs, suggesting that the correlation coefficients are not independent of one another. Is this a problem for any of the statistical tests?
 - Fig. 3a-d: I am unfamiliar with the process of discarding replicates where the sample size is too low and uncertain whether this approach is valid.
 - The analysis reported for Fig. 2d is unclear.

- L496: How was the slope of the increase estimated? What do the +/- errors in the reported slope represent?
- L498: Why n=21? Note that data from different generations cannot be described as 'replicates'.
- L504: Why was a heteroscedastic t-test chosen here?
- It is still unclear how replicates are combined for the statistical tests associated with Fig. 3a-c. For instance, each replicate will have its own value of $P(D=2)$ along with its own $P(D=2|R_N)$. So what exactly is tested in the one-sided t-test in Fig. 3b / Suppl. Table 4?
- In Fig. 3a-d, the authors variably use either a one-sided or two-sided t-test to test for significance of the correlation coefficients (Pearson's r). Why isn't this choice consistent?
- The authors have stated in their response that they did not perform statistical testing on the results in Suppl. Fig. 3 "because no claims were made from this dataset" – which raises the question of why these results were presented? I strongly suggest that the authors present statistical tests of all quantitative results included.

Other minor comments:

- The abstract should mention which antibiotic was used (around L21).
- L78 cites ref. 28, but this reference did not actually use branching processes as stated. (This reference could however be cited more accurately in L73.)
- L92 it would be helpful to state that tolC is an efflux pump.
- L192 typo: $0.05 \leq p\text{-value} < 1$ should presumably read < 0.1
- L242 missing a word? – "the probability of daughter cells"... surviving?
- L345 "a high rate of recovery from antibiotic damage" – clarify, this is in the WT
- L410-420: Why is the background correlation C_B calculated differently for each kin relationship?
- Around L438-440: The Shannon entropy calculation is still unclear.
 - I'm confused by the choice of cells at each generation N : For $N=3$, why do you go back as far as 4 generations, which isn't illustrated in Fig. 2a or included in Fig. 2bc? Where do cells with the relationship [3,3] come in?
 - It still isn't clear what the summation will look like for $N=2$ and beyond, where more than one relation Y needs to be considered.
- Fig. 3c: Please double-check the correlation coefficients reported in the caption: It would appear that the decline is steeper for ΔtolC than for WT.
- Fig. 4b caption: The test/interpretation is unclear; please clarify that you are evaluating differences between WT and ΔtolC here.

- Suppl. Fig. 2b-d: the annotations should indicate $R_{N=1}$, similarly to panel a.
- L513: Still references Suppl. Fig. 2e, which has been deleted.
- Suppl. Fig. 2e caption: Clarify, is $\langle X \rangle$ calculated across all cells in all generations?