Reviewer: Jesse Shapiro

Basic reporting

In this brief article, the authors use published data to investigate whether antibiotic resistance is associated with less genetic diversity among bacteria, as predicted by population genetic theory on selective sweeps. The manuscript is clearly written, well-situated in the literature, and conforms to PeerJ and disciplinary norms. The figures are generally clear (but see detailed comments below).

Thanks!

The raw data and code are available in GitHub. I have a few suggestions to improve the clarity of the paper:

1. The second Introduction paragraph mentions how 'susceptible strains co-exist with resistant strains.' This could be due to resistant strains suffering fitness costs. Such costs of resistance could be mentioned.

Yes, agreed. We have added a sentence and several references.

2. Also in the introduction, line 72, you could specify that sweeps still occur in the scenario of highly diverse resistant strains, just that this is a soft sweep of resistance arising on multiple genetic backgrounds. This would help clarify the concept of soft sweeps.

Yes, we added a short paragraph.

3. Introduction line 79: "lower sequence type diversity." Can you give some information on the time scales of divergence among STs, and whether the time scale is expected to be relevant to the evolution of antibiotic resistance? Even a rough estimate would be useful. There is an implicit assumption throughout the manuscript that resistance generally evolves recently (subsequent to the divergence of STs) but in some cases resistance could be an ancient trait that is ancestral to the entire species, therefore common to all STs (except for occasional loss of resistance). For example, on line 202, could you support the idea that 'resistance is likely a newly evolved trait'? There are many examples of ancient resistance (e.g. https://doi.org/10.1371/journal.pone.0069533)

Yes, we agree that this is our assumption. Added a sentence to this effect.

4. Figure 2A needs a colour legend. Also the asterisks and 'n.s.' text for) is too small defined, and the meaning of these symbols (presumably statistical significance) should be defined in the legend (same for Figure 1).

Done.

Experimental design

The article describes original research with a well-defined and relevant question. It is clearly stated how this work fills a knowledge gap in our understanding of the evolution of antibiotic resistant bacteria.

The methods are generally well defined, but I have two specific suggestions to improve clarity and reproducibility:

1. Were specific search terms used for the literature review? This could be mentioned in the first Methods paragraph.

This is a valid question, but no, we didn't use specific search terms. The datasets we used should be considered a convenience sample.

2. In the linear model (lines 266-269), could you specify whether 'dataset' is included as a fixed or random effect? I think it is fixed, but please clarify.

Yes, it is fixed. We added a sentence to clarify.

Validity of the findings

In general, the conclusions are well-supported by the data and analyses. The consistent results across studies provides good support for the generality of the association between resistance and lower ST diversity, and certain caveats are appropriately mentioned in the discussion. I would like to raise a few additional caveats that could be addressed or mentioned to further support the conclusions.

1. One important caveat to consider is that not all STs are equally related: some could represent ancient evolutionary divergences and others more recent ones. In some cases, it should be possible to cluster STs (e.g. using the BURST algorithm or other tools described on pubmlst.org) which could provide additional resolution to the study – e.g. does one ST likely descend from another, such that resistance in both STs constitutes one, not two, independent acquisitions of resistance? Even if it is not feasible to conduct such analyses, it would be useful to comment on these considerations in the discussion.

Yes, we added some language to acknowledge this issue and suggest future studies.

2. A related point is that there are likely biases in how STs and sensitive/resistance isolates are sampled. For example, sampling might be focused on a hospital outbreak of a resistance clone, which would bias the dataset toward a large number of isolates from the same resistant ST. I appreciate the permutation tests used to control for different sample sizes of resistant vs. sensitive isolates, but I believe the permutations still do not account for this particular type of sampling bias. I

am not sure if there is an easy way to control for this – although see my point directly above: could the relatedness among STs be incorporated into the permutation test? If not, or if this proves too difficult, it could still be mentioned in the discussion if and how such sampling bias might impact the results. Given the relatively small number of studies included in the analysis, the individual publications might give clues about whether or not a hospital outbreak (for example) or other likely clonal expansion of a resistant clone was targeted for sampling.

We added the following to the methods section: The E. coli studies were explicitly from surveillance data: the two Kallonen datasets, the Adams-Sapper dataset and the Yamaji datasets. The other datasets consisted of convenience samples or it was unclear how sampling was done. Importantly, none of them were focused on a specific outbreak situation.

3. It is interesting that 'the cases where the diversity was very high in the resistant population were the cases where resistance was very common' (lines 368-370). To further support the idea that resistance increases over time (in both the total proportion of resistance bacteria and the diversity of resistant STs), would it be possible to use information about the dates when each antibiotic came into clinical use? One would expect the first antibiotics to enter clinical use (e.g .1940s and 1950s) would select for more resistant isolates on more STs compared to more recent antibiotics (1960s…). Just an idea if this information is available.

This is an interesting idea and we may consider this for a future study.

4. In the Discussion, it is rather a big jump on line 398 that "we can conclude that [...] resistance typically occurs by multi-origin soft selective sweeps." In cases where resistance is common, is it possible that resistance is an ancestral trait of the species (see my comment above)? Sweeps imply that there is an increase in resistance-conferring alleles over time, driven by positive selection. You actually have some interesting evidence for this, which is mentioned later in the Discussion (lines 420-428) but perhaps would be more useful to bring in earlier to support the evidence for sweeps.

There is a lot of evidence that resistance has increased over the last decades. We added some of this to the introduction.

Reviewer: Ben Dickins

Basic reporting

This paper is very well written, and care is taken to explain the context and methods in detail. Figures are very clear and well presented. Data and scripts are made available via GitHub.

Thanks!

In lines 132-161 the authors summarise the input studies. I would like to request that the authors include in this (or add after) a statement or statements clarifying the nature of sampling. I understand the input studies to be census-style samples over a period of time at a particular (usually hospital)

location. This is important to reassure the reader that there is not a filter that is being applied that might be associated with resistance/susceptibility status.

Thanks for bringing this up. We added more information. The E coli studies are based on surveillance data. The other datasets consisted of convenience samples, or it was unclear how sampling was done. Importantly, none of them were focused on a specific outbreak situation.

(Elsewhere the authors are clear on discarding ambiguous sequence types, although it might also be helpful to review input studies for handling of intermediate susceptibility).

Good point. We added this information.

It is possible that ascertainment biases (affecting proportions of resistant versus susceptible strains) may operate if clinical entailments of resistance influence the likelihood of hospital follow-up from primary care. This reviewer cannot think of a negative impact of this on the study conclusions, however.

We agree that there may be ascertainment bias – and the fraction of samples that is resistant may be higher in the studies than in reality. However, we don't think this will affect estimates of diversity, which is what we care about in this study.

I have reviewed the code – in particular Diversity_Indices_WithFunctions.R. Please note that the drugRfiles object is not defined but appears to be the same as ResDocs object – please could authors amend this in the code.

OK, Thanks for catching that. We fixed it.

Referencing is thorough, but I have two minor points related to Jonas et al. (2017). First would the authors mind updating the World Bank URL for stability? Second, I note that the 24 million estimate of people added to the extreme poverty category is an outcome of the report's "high-AMR impact" category. Since they also offer a "low-AMR impact" simulation scenario/simulation it would be good to add this rider to reflect uncertainty in findings.

OK, agreed. We added the low AMR estimate too. The url is added too.

In line 289 – may I ask whether additional plots should be placed in a supplementary section in addition to the GitHub repository? Based on my review of GitHub content, the plots that were chosen for the principal manuscript are representative.

We think that that would be too many plots for a supplement.

Some typographical corrections follow:

Full stop missing in line 46. FIXED

Sentence beginning line 60 is incomplete – suggest "that a soft sweep [occurred] with multiple…"

FIXED

Revise bracket use in line 71 FIXED
Line 95 – “the paper include” > “a paper includes” – and subsequent verbs FIXED
Line 243 – stray hyphen to delete FIXED
Line 262 – bracket formatting FIXED
Line 294 – reference made to labelling in figure 1 – does this mean shaded? Individual STs do not appear to be labelled in this plot. FIXED
Line 304 – typo: “resistant strains were significantly less diverse than [susceptible] strains” FIXED

Very minor: it may be helpful to clarify at the outset that the article is referring to incomplete selective sweeps (although this is mentioned in the body of the manuscript). It is possible that some readers may expect this to refer to variation after fixation.

We now included more information about sweeps in the introduction.

Experimental design
A key estimand in this study is diversity (contrasting antibiotic resistant and susceptible strains). First it is proposed that events in the world may lead to reduced real diversity among resistant strains (= the estimand). These events are the emergence or acquisition (by HGT) of resistance in a subset of the population (in a time-dependent manner). Because this occurs in a subset of the population it is essentially a sampling process. Second, the authors also note that the smaller proportion of resistant strains in a sample/dataset can result in reduced measured diversity (= the estimate). This is also a sampling process, but it is artefactual.

The manuscript shows the differences in measured/estimated diversity from datasets illustrating (mostly) lower diversity among resistant strains (figure 2) and then examine the relationship between the proportion of samples that are resistant and diversity normalized by the maximum possible measured diversity (figure 3). This normalization process is an effort to correct for the second process and arrive at a reliable estimate of diversity.

Since, for most antibiotics, the resistant proportion is < 0.5 (figure 2B), one might suppose that the second (artefactual) sampling process would affect interpretation of figure 2A. One might consider an additional panel with normalized GSI values, but I am satisfied with the stated bootstrapping approach (described in lines 231-243 and instantiated in Diversity_Indices_WithFunctions.R). I note that the simulations 1. pick with replacement and 2. pick resistant and susceptible data without dependency. In other words, they are essentially identically conditioned multinomial picks varied only by size of sample. The authors may wish to comment briefly on this (low priority).
Added this:

The sample sizes for these simulated populations were the same as the actual observed resistant and susceptible populations. For example, there were originally 33 Ampicillin-resistant samples in the Yamaji_1999 dataset and 92 Ampicillin-susceptible samples; the bootstrapping procedure created populations of these sizes by sampling with replacement from the entire dataset consisting of 92+33 = 125 strains.

I do have a request regarding the relational analysis with normalized data (figure 3). In lines 262-264 simulated data is described for testing the linear model. Please would the authors share this simulated data (or code and summaries thereof)? May I ask for the text here to be expanded to report more detail: can you describe a distribution of p values as well as a distribution of effects (which may include negatives)?

Will do. Code is now on Github, and the p-values for a linear model on simulated data with and without normalization.

Let me justify this starting with the intuition. Taking the example described in lines 254-256 of the submission the authors report that the maximum possible value of the Gini-Simpson Index (GSI) is reduced in a sample of n=2 to 0.5 (and this is then used for normalization). However, what is also reduced is the likelihood of observing 2 different types if the population from which the sample is taken contains some fixed number of types. For example, if a population contained 50% type A and 50% type B, a sample of size 2 would have a 50% chance of estimating a GSI of 0, while successively larger samples would be more likely to provide the correct estimate of 0.5 or one that is close to it. This is relevant to the testing of significance as well as to the “low power” comment on line 359.

To follow up on this intuition I have simulated the expected effect of sampling from a population with a fixed composition (and therefore a “real” diversity; see attachments). This supports my intuition that, while normalization corrects the expected value of the diversity measure (and brings it in line with the real value plotted as a red dashed line), it also increases the spread of estimates in smaller samples. I would like the authors to comment on this in relation to the observations in figure 3.

In the form of an argument: a concern could be that a left tail + higher variability could lead to some deflation in GSI estimates from resistant, smaller samples, especially if these estimates are not robust to extreme values (as is the median in the box plots). Seeing figure 3 from his perspective may lead to some concern that the regression analysis may be negatively influenced by this. The counter is to provide more detail on the findings from simulated data (as requested above and re lines 262-264) and to describe why the estimated diversity across the datasets is robust.

We agree that indeed, the variance will be higher for smaller samples. Indeed in a sample of 2, if the population frequencies of two sequence types are 50% each, we could observe a normalized GSI or 0 or 1 equally likely.

However, we don't understand how this would create a positive relationship between the fraction resistance and the normalized GSI values.

Validity of the findings

Despite my responses above I do believe that the main findings are robust and that the evidence most likely supports a reduced real diversity among resistant strains. This is because the detailed procedures related in the first analysis (including the empirical p value-based assessment for figure 2B) support this contention. I would be happy to see a little more detail on the simulation supporting the analysis in figure 3 (as outlined in the experimental design section) to strengthen this conclusion.

Additional code and figures are added to the github repo.

Additional comments

I commend the authors for their very clear exposition throughout the manuscript (also extending to the code base which is highly readable). I have enjoyed reviewing this manuscript and consider it to be a valuable contribution.

Thank you!

Reviewer 3

Basic reporting

The article is well written and easy understand. The data are publicly available and script are shared. However I see major problems in the logic and the analyses. The authors hypothesized that bacterial strains that present resistance to antibiotics are less genetically diverse than the one susceptible. They assume that this is due to select sweep. First, the literature lacks important studies that looked at how adaptation works in bacteria (see Richard Lenski's lab experiment for instance).

We love Dr Lenski's work but we are not sure how it is relevant here. Most of the resistance we study is due to Horizontal Gene Transfer from related strains, which is unlikely to happen in the Lenski experiments. In addition, we are trying to determine whether, in real life, resistance is present on many different sequence types and we don't believe that the Lenski experiments contained that kind of genetic diversity.

Second, author are referring to selective sweep all along the manuscript but they actually should refer to clonal interference.

We beg to differ. We now explain why we call the spread of resistant strains (that is, strains that carry resistance, not the spread of a resistance gene from bacterium to bacterium) a selective sweep.

Selective sweep is positive selection acting on a particular allele of a gene that will sweep neutral variation associated with it to get to fixation. However, the in the case of this study, the authors just look at MLST genes which are not the genes associated with antibiotic resistance.

True. In principle, it could have happened that only one sequence type had acquired a resistance gene and thus all resistance in a species would be due to that one sequence type. Instead we find that many different sequence types have acquired resistance.

So lack of diversity could be due to the bottleneck endured after applying antibiotics and not related to the spread of the resistance.

We don't understand this sentence. This bottleneck would have happened inside a host? Or in a country?

So the question that this study tried to answer is more : is the diversity of strains harboring resistance lower than the one of susceptible strains.

Yes, agreed. That is our question.

Experimental design

In my opinion, there are problems with the methods and how they answered the question. First, they do not show a measure of genetic diversity for within species comparison but a measure of diversity used normally for between species comparison. A classical statistics to measure within species genetic diversity is pi or theta. To do so it is necessary to recover the sequences, align them and estimate those quantity.

We don't agree that diversity within species should always be measured using SNPs.

Many others have used Simpson's index for MLST data. We did a quick Google scholar search and found for example: (Isaksson et al. 2016; Hua et al. 2020; Ramadan et al. 2020)

Hua, Xiaoting, Linyue Zhang, Jintao He, Sebastian Leptihn, and Yunsong Yu. 2020. "Population Biology and Epidemiological Studies of Acinetobacter Baumannii in the Era of Whole Genome Sequencing: Is the Oxford Scheme Still Appropriate?" *Frontiers in Microbiology* 11 (April). https://doi.org/10.3389/fmicb.2020.00775.

Isaksson, Jenny, Lucía Gallo Vaulet, Linus Christerson, Anke Ruettger, Konrad Sachse, Carolina Entrocassi, Érica Castro, Marcelo Rodríguez Fermepin, and Björn Herrmann. 2016. "Comparison of Multilocus Sequence Typing and Multilocus Typing Microarray of *Chlamydia Trachomatis* Strains from Argentina and Chile." *Journal of Microbiological Methods* 127 (August):214–18. https://doi.org/10.1016/j.mimet.2016.06.005.

Ramadan, Hazem, Charlene R. Jackson, Jonathan G. Frye, Lari M. Hiott, Mohamed Samir, Amal Awad, and Tiffanie A. Woodley. 2020. "Antimicrobial Resistance, Genetic Diversity and Multilocus Sequence Typing of Escherichia Coli from Humans, Retail Chicken and Ground Beef in Egypt." *Pathogens* 9 (5): 357.

https://doi.org/10.3390/pathogens9050357.

This shall be done as you might expect differences in result because some alleles are shared between types. Second, it is essential to know if the resistance is encoded by a plasmid or by the chromosome. You expect radical differences in how the resistance will spread and how it will affect nuclear genes diversity. Then separate test should be performed accordingly.

This is an interesting question. For some of the species / drug combinations here we don't know what causes resistance. But for *E. coli*, we know that ciprofloxacin resistance is usually caused by a mutation, while resistance to the other drugs is usually caused by resistance genes. If you look at the results in figure 2, ciprofloxacin doesn't stand out. We believe that the rate of mutation and the rate of HGT may be similar. But to test this, a different type of study would need to be designed.

Third, the statistics should be redone and more elaborate. You can fit a linear mixed model that will have diversity as a dependent variable, resistance status as independent variable and Dataset as random variable. I think bootstrap approach is unnecessary in your case and that your glm (it is actually a linear model you do not generalized ie your dependent variable is supposed to be normal) is incorrect as your residuals are not randomly distributed (figure 4). Linear model are anyway incorrect on proportion data.

Thanks for catching the glm / lm issue. We now changed this in the code and use the lm() function instead of glm. We believe that the bootstrap approach is the way to go for a somewhat messy / convenience dataset. Bootstrapping is common in population genetic studies.

Validity of the findings

I do not think the results support the findings. The authors should redo the analysis and clearly assess what they want to show. If they want to look at selective sweep, they should analyse sequence data and not just strain. They should also be careful that plasmid encoded resistance will not lead to selective sweep but just the spread of the plasmid within population without drastic reduction of diversity.

We don't agree here. If a plasmid had only been acquired one time in one *S. aureus* strain, then all *S. aureus* with that plasmid would have been the same sequence type. We would have seen a drastic difference between diversity of resistant strains (very low) and diversity of susceptible strains (high). Our results show that many different sequence types have acquired plasmids / genes or mutations that cause resistance.

Finally, there is also a cost to resistance. It has been shown that resistance are lost fast as soon as no more selective pressure is applied through the exposition to antibiotics. So basically your susceptible strains can be one that have lost resistance and spreading in the population.

Yes, true. We have now added information about costs in the introduction.

So you mix ancestral populations, selected populations for antibiotics and populations under relaxed selection because some antibiotics are no more in use. That is problematic in the way you interpret your data. In conclusion, the authors have to reassess the way they analyzed the data, the objective of their study and what they conclude from it.

We thank you for your review. We have made some changes based on your suggestions, but do not agree that our approach of studying MLST diversity is fundamentally wrong. There are many ways to quantify diversity. We study here whether fast adaptive evolution, as has been observed in bacteria in the last few decades, led to a reduction in MLST diversity. We agree that it would be of interest to repeat this type of study using genomic sequencing data.