nature research

Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: There is little evidence that spicy food in hot countries is an adaptation to reducing infection
Corresponding author name(s): Lindell Bromham

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:
28th April 2020

Dear Professor Bromham,

Thank you once again for your manuscript, entitled "Why do hot countries have spicy food? A macroevolutionary analysis", and for your patience during the peer review process.

Your Article has now been evaluated by 3 referees. You will see from their comments copied below that, although they find your work of potential interest, they have raised quite substantial concerns. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version if you are willing and able to fully address reviewer and editorial concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

1) As highlighted by Reviewer 3, your data on pathogen prevalence and foodborne illness are country-

level, and represent just 11 countries. We are not persuaded that the data currently included in your analyses provide sufficient strength of evidence to justify your conclusion that patterns of spice use are not consistent with risk of infection. Please address your central hypothesis with more fine-grained data, using robust and well-powered analyses. For all null results, please provide evidence of sufficient power and report Bayes Factors or equivalence tests (please see our attached checklist regarding journal requirements for the reporting of studies with null results).
2) Reviewer 2 points out that the correlation between hot chilli use and temperature is central to the adaptive cuisine hypothesis; please conduct analyses using your data to investigate if this correlation exists and survives statistical correlation.
3) Please systematically explain and justify your choice of variables for analysis (as raised by Reviewer 3) and the rationale behind your choice of models (as raised by Reviewers 1 and 3). Include details of these in your Methods section such that your methods are transparent and fully replicable. It is also very important to address Reviewer 3's concerns regarding the fact that the selected variables constitute proximate measures to explain the use of spices, but do not test the ultimate evolutionary explanation for why spice use evolved in warmer climates.
4) Please ensure that you adjust the text to preclude any misinterpretation of your results (as raised by Reviewer 1) and discuss alternative hypotheses consistent with your findings (as requested by Reviewer 2).

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript we would hope to receive it within 4 months. We understand that the COVID-19 pandemic is causing significant disruptions which may prevent you from carrying out the additional work required for resubmission of your manuscript within this timeframe. If you are unable to submit your revised manuscript within 6 months, please let us know. We will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

• Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,

Charlotte Payne
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Cultural evolution; Macroevolutionary methods

Reviewer #2: Spice & pathogen resistance

Reviewer #3: Spice & pathogen resistance

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
N.B. PDF version of this attached for ease of reading.

Overall, I enjoyed reading the work and think the authors did a very thorough job of testing all possible hypotheses. The data collection and methods were thorough and impressive, and appropriate. I think they draw in general suitably cautious and very reasonable conclusions.

I do have some minor comments and suggestions.
One key, but easily remedied, point is that there is no explanation of how cultural relatedness was assessed in the main text, nor geographical location. I have read the supplement, but please at least summarize these in the methods section and refer people to the supplement.

One other major issue I had was that it seems to be *almost* implied in several places that poverty (and associated data) could be driving food spiciness. For example lines 85–88:

“The answer lies in the covariation of poverty and health. Poor health outcomes are strongly influenced by socioeconomic factors, so the association between spice use and poor health outcomes may reflect a more general relationship between poverty, health and cuisine”

And then lines 114–119:

“We suggest that socioeconomic indicators such as GDP are correlated with poor health outcomes from a range of causes, including infection, malnutrition and accidents, as reflected in reduced life expectancy, and the correlation between infection and spice is an indirect effect of relatively higher spice use in regions with traditionally lower socioeconomic indicators. In other words, spice use is

better predicted by markers of poverty and not by risk of infection."

I *think* the authors are finally drawing the conclusion that poverty and road deaths etc. are correlated with spiciness because of cultural relatedness and proximity (perhaps the authors can clarify?), but this is not at all clear from the wording. Also, it is not clear if these correlations stand when cultural relatedness and geographic location are included in the model. Is this the case, and if they do stand, then what explains this? Maybe the cultural relatedness variable is not helpful? Maybe the government/nation state makes a difference? Maybe colonial history?

To me (and I think the authors agree, based on their sentence "These results do not suggest that poverty causes higher spice use, just as they do not suggest that car accidents increase spice use", lines 103¬–4) a link between poverty and spiciness doesn't – without some kind of causal link – make any sense. Even if the authors don't actually mean to say that there is an effect of poverty on spiciness, I think it would be very easy to misinterpret the paper as it is currently worded.
Possibly the causal link is that geographically proximate and related cultures tend to share a *lot* of traits, including socioeconomic and culinary ones? …But as mentioned, if – even when relatedness and proximity are taken into account – this correlation still stands (whether this was the case is not made clear in the text), then the authors need to robustly try and explain this, and possibly change the drive of some of the paper. Maybe the cultural relatedness is actually not well represented by the language data? Maybe it has more to do with colonisation patterns (e.g. spreading spices, colonising spice-rich areas, and creating extractive colonial institutions) or to do with the nation states (and their governments, both current and historical) which the different cuisines find themselves in?

Handling these aspects of the paper is especially important because it is likely to be the kind of work attracting public and media interest, and the true message must thus be really clear. I can just see someone saying that spicy food is proven to make people poorer or vice versa! (Assuming this is not the case)

If there is indeed negligible effect of poverty on spiciness when other factors are taken into account, then I would thus suggest that the authors reword several of their sentences (such as those cited above) to make clear that there is no causation between poverty and food spiciness. I would also suggest a new sentence/paragraph especially to make this clear towards the end, and a final sentence in the abstract and conclusion summarizing what they think *does* cause spiciness and why it is predicted reasonably well by poverty (presumably proximity and cultural relatedness?).

On another point, although overall plant and crop diversity may not affect spiciness, I wonder if the *spice* crop diversity could affect it. For example maybe if you plotted the potential growth zone of different spices, the place they all overlap might have the spiciest food. I guess this is a whole new project, but maybe worth mentioning as a future avenue?

Yet another point, but related, is that climate doesn't just come down to temperature – there is humidity, seasonality etc. etc. If this were all taken into account, maybe there would be a correlation with climate? This again is obviously outside of the current study's remit, but should be mentioned in the discussion/conclusions.

Related to the point about the poverty <-> spiciness correlation, I would also like to see it clarified whether, in the correlations cited and given in the table in the main text, relatedness and geographical location have been taken into account or not. Also, the table of correlations presented is just

univariate, but sometimes it is mentioned that other factors were added into the model (e.g. "…adding life expectancy to a model with either historical pathogen load…", lines 76–77). I have two major points: (a) were factors included together in models in a systematic way, or were most models univariate? If they were not, I think maybe you should do this for at least a few models. You probably need to check what correlations remain significant when other factors are included, and what drops out.; (b) can you please provide a full table of ALL the analyses, with full results, as an excel table in the supplement? Although the coloured table is nice, it is quite hard to read and does not show all of the analyses done. A table summarizing correlations in the main text might be nice too.

Overall, please included all the RAW data, and detail exactly ALL the analyses. I found this was lacking in the supplement and needs clarifying.

Phew, well, that's quite a lot of stuff in the end I guess, but this is actually all minor. I like the paper a lot, and I think the overall analyses conducted and drive of the paper is very good. I hope my comments and suggestions can help, and make it more robust to criticism and misinterpretation.

Reviewer #2:
Remarks to the Author:
In this manuscript, authors put into test the adaptive cuisine hypothesis, which states that spice use is higher is hotter countries because of its beneficial properties to humans since they could reduce the risk of infection from spoiled meat in hotter climates. This hypothesis has been championed by the now-classic and popular work of Billing & Sherman. This hypothesis is supported by a positive correlation between spice use (in traditional recipes containing meat, before refrigeration) and mean annual temperature, especially of those spices with more antimicrobial properties. However, since similar cultures tend to share similar features, it could introduce an unintended autocorrelation problem that Billing & Sherman acknowledge but lacked suitable means for correcting for geographic clustering or autocorrelations between similar cultures, a common statistical problem.
In this manuscript, authors grew and curated a more extensive database than Billing & Sherman, by including more spices. They found that mean spices per recipe was correlated with mean annual temperature for countries and regions, replicating similar observations from data of Billing and Sherman. However, this significant correlation did not survive when spatial and phylogenetic non-independence was taken into account. Surprisingly, they found that poverty provides a more reliable predictor of spice use than poor health outcomes or risk of infection.
Even the number of road traffic deaths provide a stronger prediction of spice use than infection threat or malnutrition. Thus they demonstrate "that there is currently no compelling evidence to assign a special explanatory role for risk of infection in variation in the spiciness of cuisines because other socioeconomic indicators of poverty and poor health outcomes provide a better explanation of variation in spice use than pathogen load or incidence of foodborne illness"
The take-home message is that correlation in the context of Darwinian gastronomy should be interpreted with caution. This observation minimally supports the idea that in present refrigeration times, the association between spice use in hotter countries as a tool to mitigate infection risk could NOT satisfactorily be explained in terms of its antimicrobial properties. Thus, current spice use is not reflecting an ongoing response to levels of microbial threat rather it seems to follow historical inertia and cultural inheritance of food preferences.

This is a provocative and exciting manuscript that certainly challenges the adaptive cuisine hypothesis.

However, I think the manuscript needs some important improvements and further discussion of the main findings.

Main concerns
First, the title is misleading since it generates the expectation that a satisfactory answer was found for the question of "Why do hot countries have spicy food?" However, and as authors acknowledged, they could not conclude that poverty or car accidents increase spice use!
1) A new title is needed to reflect the main finding that the correlation between spice use and annual temperature is due to geographical and cultural autocorrelations.
2) In this regard, after reading the manuscript, one is left with no alternative explanation for why hot countries have (or consume more) spicy food. I think authors should discuss in more detail that the lack of correlation between spice use and temperature, after correction, gives more support to other alternative hypotheses: Perhaps due to their chemesthetic sensory properties (which applies to all countries and human beings) and/or due to cultural transmission of food preference, etc..
It does also not discuss the possibility that perhaps the initial higher spice use in hotter countries indeed helped mitigate risk infection, but currently, in post refrigeration times, it just follows cultural food preference transmission mechanisms.
3) Also, a central argument of Billing & Sherman's hypothesis was the correlation between a single spice (hot chili peppers use) and annual temperature. This is an important point that needs to be addressed and explored in your database before the paper can be accepted for publication. It does such a correlation exists, and if it does, it did survive statistical correction?

Reviewer #3:
Remarks to the Author:
The aim of this study is to test the "adaptive cuisine hypothesis" using a global dataset consisting of recipes from 45 cuisines. This study found that poverty and poor health outcomes were better able to explain spice use than infection risk. While the research approach is conceptually novel by considering (to my knowledge) socioeconomic and other "cultural" factors in predicting spice use, there are several flaws that I think should prevent this paper from being published in Nature Human Behavior.

One issue I have with this manuscript is the general lack of clarity about the study motivation, data, analyses, and how measures were defined. For example, the main manuscript and methods files uses terms biocultural diversity, biological diversity, and cultural diversity interchangeably. The authors should note that there is considerable difference between these terms. Moreover, the authors appear to apply these terms in reference to linguistic diversity, and sometimes botanical diversity (I think). The authors should avoid using broad categories unless they define them and use them consistently throughout the study. As another example, the third aim of the paper states that they use "statistical techniques" to explicitly compare the explanatory power of "different hypotheses" for higher spice use, considering not only infection risk mitigation, but also history, proximity, socioeconomic factors, cultural diversity, botanical, and agricultural diversity. What are the statistical techniques? What are the hypotheses? Why did they choose these measures? These "hypotheses" are never spelled out – what are the authors actually predicting aside from the adaptive cuisine hypothesis? (Which also suggests that the hypotheses were not pre-registered)… Which leads me to the next point…

I am more concerned about the lack of a priori justification for the variables chosen for this study. There is nothing suggesting why socioeconomic variables and other measures should be included in

the study. The authors went through a lot of work to find these measures, but the study seems exploratory in nature. The way the study is currently presented – with no alternative motivation or literature review on alternative approaches – makes it seem that the authors cherry-picked some variables and threw them into a model. However, it's hard to know for sure because of the opaque presentation of the results and lack of explanation for how analyses were conducted (I don't think it's good practice to point to other studies rather than clearly outlining your analytic approach). Specifically, I am left wondering what statistical tests were conducted? It seems as though a t-test was used for some analyses, and that a multivariate test was used for others. The authors then report AIC results and point to Figure 2, but there are no additional models? Why report an AIC value (model-level) if the emphasis is on GDP(predictor variable)?

Most importantly, the variables that were selected seem to be proximate measures to explain the use of spices, but do not test the ultimate evolutionary explanation for why spice use evolved in warmer climates. The authors do not discuss this obvious discrepancy.

Furthermore, the reference the authors include for using alcohol and vinegar indicates that wine is shown to have antimicrobial activity, but beer does not. The current study appears to collapse all types of alcohol into one variable, but the literature cited by the authors suggests this is not a well-reasoned choice. Also, why is alcohol/vinegar use not pre-specified as a hypothesis? Was this a post-hoc analysis? It's also not clear why they are analyzed as a proportion.

The sample size for this study was small. The majority of these analyses are conducted on half of the dataset. Most of the raw data from Billing and Sherman are not available. This leads to a breakdown of the following: 8/45 cuisines are from Japan, 8/45 are from India, 10/45 are from a "global" data set, and the remaining 19 are Chinese data. The remaining analyses were conducted on this data set and there was no effect for pathogen load. There was likely no effect because the variance for pathogen load in the Chinese data was 0. The same is true for Japan and India.

Another concern that I have is with the time frame that the data were collected from each of the different sources. Billing and Sherman are data from 1998, but the WHO data and other sources are later. For example, life expectancy was based on the most recent data for each location – this is problematic for comparisons to average spice use that occurs at different periods of time in each data set. Also, the life expectancy data are from different sources – the limitations of this must be addressed. The authors got one set from the World Health Organization and the others via a source from Wikipedia.

Finally, this paper lacks a strong conclusion along with a statement about the limitations of the research. Overall, I do not think the way the data are presented provides a clear and convincing answer to the research question. I also do think the findings are over-generalized, as a majority of the data comes from three Asian countries, and does not actually offer an ultimate explanation for the evolution of spice use; at best, it might provide a proximate explanation. This is fine, but should be made transparent.

Minor comments:
Ln 75: use of "that" when authors mean "than"

Data are sharable, which is good.

Author Rebuttal to Initial comments

28th April 2020

Thank you once again for your manuscript, entitled "Why do hot countries have spicy food? A macroevolutionary analysis", and for your patience during the peer review process.

Your Article has now been evaluated by 3 referees. You will see from their comments copied below that, although they find your work of potential interest, they have raised quite substantial concerns. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version if you are willing and able to fully address reviewer and editorial concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

1) As highlighted by Reviewer 3, your data on pathogen prevalence and foodborne illness are country-level, and represent just 11 countries. We are not persuaded that the data currently included in your analyses provide sufficient strength of evidence to justify your conclusion that patterns of spice use are not consistent with risk of infection. Please address your central hypothesis with more fine-grained data, using robust and well-powered analyses. For all null results, please provide evidence of sufficient power and report Bayes Factors or equivalence tests (please see our attached checklist regarding journal requirements for the reporting of studies with null results).

> **E1.** We have now nearly doubled the number of datapoints, and increased the number of recipes to 33750. We also run all analyses on several different datasets, differing in the spatial grain of resolution (subnational or national) and in the source of the data (from Billing & Sherman's original data from original cookbooks or derived from recipe data for this study). We formally test the power of all of these datasets and show that our data has greater power than either an expanded country-level dataset or the original data from Billing and Sherman's 1998 study (Figure 2). We report Bayes factors for all tests, and provide statistical results for every test conducted in Supplementary information (Table S3).
> We have now added a new variable reflecting infection risk that is resolved to country level - incidence of early childhood diarrhea – from a paper published in Nature on 29 April 2020. These data are based on a large dataset of household surveys so provides an excellent alternative measure of incidence of infection at an average national level (though not specific to foodborne infection). We show that the results are similar to the courser-grained foodborne illness data or the national/subnational historical pathogen load. But these data are not available at subnational level, nor for all countries in our analysis. This published study on childhood diarrhea highlights the challenges of obtaining comparable data for global foodborne illness rates – it requires a single household-level reporting framework for all areas under study. While regional values for socioeconomic data or biodiversity measures have the same essential mode of calculation at all scales, this

is not true for sub-national disease prevalence data which relies on the disease reporting infrastructure that varies dramatically between jurisdictions. This means that some diseases are reported at a subnational level in some countries but not others, and that incidence of different diseases recorded in substantially different ways which would bias our analysis (e.g. underreporting of diarrheal diseases in some jurisdictions would lead to artificially lowered pathogen counts in all of those subnational cuisines). We think that an attempt to gather foodborne incidence data at subnational level from disparate sources for the purposes of testing the adaptive cuisine hypothesis could generate false patterns dependent more on disease surveillance and data availability than on disease prevalence.
While we could technically apply an equivalence test, it would require us to prespecify a value that we consider to be a "meaningful" effect. But in the case of our analysis, we are not evaluating whether or not relationships exist or are statistically significant (they do and they are). Instead, our main result is to show that it is relatively easy to generate statistically significant results – for example between road traffic deaths and spice use – but that it would not be sensible to interpret all such significant results as revealing an important mechanism of cultural evolution. This is why the importance of our study stretches beyond Darwinian gastronomy, and is more generally reflective of widespread challenges in studying human cultural evolution from cross-cultural patterns. Given that we can see that some of the significant results are non-informative, our conclusion is that there is no special reason to believe the somewhat weaker (but still significant) association between infection risk and spice use is privileged in explanatory power.

2) Reviewer 2 points out that the correlation between hot chilli use and temperature is central to the adaptive cuisine hypothesis; please conduct analyses using your data to investigate if this correlation exists and survives statistical correlation.

E2: We are not aware of any special explanatory power attributed to chilli in the primary publications of the Darwinian cuisine hypothesis. While Billing and Sherman do show that chilli is correlated to temperature, they show the same for garlic and onions (Figure 5, Billing & Sherman, 1998). They discuss the hypothesis that chilli consumption is cooling in hot climates because it makes people sweat, but they reject this as a general explanation of the correlation between spice use and temperature. Sherman & Hash (2001) consider that some spices have greater antimicrobial activity than others including chilli, garlic, onion, allspice, oregano, thyme, cinnamon, tarragon, cumin, cloves, lemon grass, bay leaf, rosemary, marjoram, and mustard. But they do not single out chilli as have special significance.

However, we now analyse chilli against pathogen prevalence for the 45 cuisines for which proportion of recipes containing chilli is available (Billing and Sherman analysed this variable but do not provide it in their publication, reporting only the results of the tests, so we cannot conduct this test for the 48 countries in their dataset). We show that use of chilli has no association with foodborne illness for any of the datasets analysed (whether using regional or country level cuisines), nor is it associated with finer grained socioeconomic data such as GDP and life expectancy, even those these variables are significantly correlated with average numbers of spices per recipe.

3) Please systematically explain and justify your choice of variables for analysis (as raised by Reviewer 3) and the rationale behind your choice of models (as raised by Reviewers 1 and 3). Include details of these in your Methods section such that your methods are transparent and fully replicable. It is also very important to address Reviewer 3's concerns regarding the fact that the selected variables constitute proximate measures to explain the use of spices, but do not test the ultimate evolutionary explanation for why spice use evolved in warmer climates.

E3: We have now greatly expanded the methods and results section to more clearly set out the rationale for variable choice, and we have included a detailed figure that we used to plan the analyses and select variables. This figure has a detailed legend with evidence either from this study or previous studies for each of the proposed explanatory links. We have also greatly expanded our description of the analysis, and all steps involved in testing variables against each other are now presented in Table S3.

We agree with Reviewer 3 that such an analysis can test only proximate relationships and cannot establish ultimate causes: indeed this is the primary conclusion of our study. A correlation between two variables tells you there is some link between them such that they do not vary randomly with respect to each other, but given the large number of covarying features of cultures and environments, it will rarely be possible to conclusively attribute this relationship to a single causal variable and rule out all possible covarying factors. We suggest that interpreting relationships such as the correlation between spice and temperature as indicative of direct causal relationships is rarely going to be possible in cross-cultural studies, but that we can formally compare the explanatory power of alternative explanations and rule some out as less plausible than others. We now emphasize this conclusion more strongly in the Discussion.

4) Please ensure that you adjust the text to preclude any misinterpretation of your results (as raised by Reviewer 1) and discuss alternative hypotheses consistent with your findings (as requested by Reviewer 2).

E4: We have now rewritten to make the conclusions of the paper clearer – details given below.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript we would hope to receive it within 4 months. We understand that the COVID-19 pandemic is causing significant disruptions which may prevent you from carrying out the additional work required for resubmission of your manuscript within this timeframe. If you are unable to submit your revised manuscript within 6 months, please let us know. We will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

- Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

- Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,
Charlotte Payne
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Cultural evolution; Macroevolutionary methods
Reviewer #2: Spice & pathogen resistance
Reviewer #3: Spice & pathogen resistance

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
N.B. PDF version of this attached for ease of reading.

Overall, I enjoyed reading the work and think the authors did a very thorough job of testing all possible hypotheses. The data collection and methods were thorough and impressive, and appropriate. I think they draw in general suitably cautious and very reasonable conclusions.

I do have some minor comments and suggestions.

One key, but easily remedied, point is that there is no explanation of how cultural relatedness was assessed in the main text, nor geographical location. I have read the supplement, but please at least summarize these in the methods section and refer people to the supplement.

> **R1.1:** This was mentioned in the text and referred to the Supplementary Information (line 275 in the previous ms) but we have now expanded the description in the main text. The word count in the methods precludes us from including the full description in the main text but it is explained in detail in the Supplementary Information.

One other major issue I had was that it seems to be *almost* implied in several places that poverty (and associated data) could be driving food spiciness. For example lines 85–88:
“The answer lies in the covariation of poverty and health. Poor health outcomes are strongly influenced by socioeconomic factors, so the association between spice use and poor health outcomes may reflect a more general relationship between poverty, health and cuisine”

And then lines 114–119:

“We suggest that socioeconomic indicators such as GDP are correlated with poor health outcomes from a range of causes, including infection, malnutrition and accidents, as reflected in reduced life expectancy, and the correlation between infection and spice is an indirect effect of relatively higher spice use in regions with traditionally lower socioeconomic indicators. In other words, spice use is better predicted by markers of poverty and not by risk of infection.”

I *think* the authors are finally drawing the conclusion that poverty and road deaths etc. are correlated with spiciness because of cultural relatedness and proximity (perhaps the authors can clarify?), but this is not at all clear from the wording. Also, it is not clear if these correlations stand when cultural relatedness and geographic location are included in the model. Is this the case, and if they do stand, then what explains this? Maybe the cultural relatedness variable is not helpful? Maybe the government/nation state makes a difference? Maybe colonial history?

> **R1.2:** We have now edited the manuscript to make it clearer that we do not imply poverty as a causal agent of spice use. We have also emphasized that all analyses explicitly model covariation due to relatedness and proximity in all reported statistics.

To me (and I think the authors agree, based on their sentence “These results do not suggest that poverty causes higher spice use, just as they do not suggest that car accidents increase spice use”, lines 103¬–4) a link between poverty and spiciness doesn’t – without some kind of causal link – make any sense. Even if the authors don’t actually mean to say that there is an effect of poverty on spiciness, I think it would be very easy to misinterpret the paper as it is currently worded.

Possibly the causal link is that geographically proximate and related cultures tend to share a *lot* of traits, including socioeconomic and culinary ones? …But as mentioned, if – even when relatedness and proximity are taken into account – this correlation still stands (whether this was the case is not made clear in the text), then the authors need to robustly try and explain this, and possibly change the drive of some of the paper. Maybe the cultural relatedness is actually not well represented by the language data?

> **R1.3:** We have added a statement that clarifies that the results do not necessarily implicate poverty as a causal driving factor in spice use, but indicate that socioeconomic features that all tend to vary together will also correlate with spice and therefore it is difficult to tease out causal factors using cross-cultural analyses alone. This point is now emphasized more clearly in both the Results and the Discussion.

Maybe it has more to do with colonisation patterns (e.g. spreading spices, colonising spice-rich areas, and creating extractive colonial institutions) or to do with the nation states (and their governments, both current and historical) which the different cuisines find themselves in?

> **R1.4:** We agree that global trade and conquest may have had an important impact on spice use over the world, but it is difficult to frame a clear hypothesis to test here. We have now added an analysis that asks whether cuisines in areas where more spices grow have greater spice use (see below), but this does not directly address the issue of whether colonizers that have access to spice-rich areas will use more spice. It is not clear how we would ask if countries that had colonial power over spice growing areas have greater spice use, since colonizing powers over these areas have changed dramatically over time. For example, the supply of nutmeg and mace came under European control after Portuguese traders reached the Banda islands in the 19thth century, but since that time the Banda islands have been at various times under Portuguese, Dutch, English and Indonesian control. More generally, the global movement of spices has a very long timeframe and different spices will have different relevant colonial periods. For example, while pepper is native to Asia, it was traded far across the Roman empire from the last century BC onward. Chilli spread globally from Mexico in the 15thth and 16thth centuries. So while colonial history (whether as colonizers or colonized) is likely to be an important factor in shaping the socioeconomic variables of each country in our dataset, it is difficult to generalise the predicted effects or to test these predictions in a variable that we could use in our analysis.
>
> If the point here is that patterns of global acquisition were built upon the colonial spice trade and this has shaped current socioeconomic standing, it is also difficult to operationalise this for our dataset. For example, many spices originate from the area now recognized as Indonesia, but islands included this modern country have had multiple different colonial powers at different times. Thailand and China both have high spice use but neither was ever subject to a colonial power. From the Billing and Sherman study, Ethiopia was one of the countries with the highest spice use, but Ethiopian cuisine largely uses spices from Asia, rather than indigenous spices.
>
> We are happy to take further advice from the reviewer on this point, if they are able to clarify what aspect of colonial history we should be testing against GDP and spice use.

Handling these aspects of the paper is especially important because it is likely to be the kind of work attracting public and media interest, and the true message must thus be really clear. I can just see someone saying that spicy food is proven to make people poorer or vice versa! (Assuming this is not the case).

If there is indeed negligible effect of poverty on spiciness when other factors are taken into account, then I would thus suggest that the authors reword several of their sentences (such as those cited above) to make clear that there is no causation between poverty and food spiciness. I would also suggest a new sentence/paragraph especially to make this clear towards the end, and a final sentence in the abstract and conclusion summarizing what they think *does* cause spiciness and why it is predicted reasonably well by poverty (presumably proximity and cultural relatedness?).

R1.5: We have now carefully rephrased to make it clear that the main conclusion of the paper is that, because so many factors have a significant association with spice, and all of those factors covary with each other, it is very difficult to untangle the causal associations based on cross-cultural correlations alone. However, we use additional tests to show that the predictions of the infection risk hypothesis are not strongly supported by these data, and that other antimicrobial ingredients do not seem to be associated with temperature or infection risk.

On another point, although overall plant and crop diversity may not affect spiciness, I wonder if the *spice* crop diversity could affect it. For example maybe if you plotted the potential growth zone of different spices, the place they all overlap might have the spiciest food. I guess this is a whole new project, but maybe worth mentioning as a future avenue?

R1.6: We have added a new analysis that tests spice use against the diversity of spices growing in the cuisine area. However this analysis has the same difficulties as many of the points raised above: access to spice growing areas has changed over time, spices have been grown outside their native distribution for economic reasons, and spices are such “tradeable” commodities that they are used far from their natural distribution.

Yet another point, but related, is that climate doesn’t just come down to temperature – there is humidity, seasonality etc. etc. If this were all taken into account, maybe there would be a correlation with climate? This again is obviously outside of the current study’s remit, but should be mentioned in the discussion/conclusions.

R1.7: We did not include additional climate variables because the measures of biodiversity (plant diversity, crop diversity) were not related to spice use. Climatic variables such as seasonality are primarily expected to influence spice use via conditions for plant growth, so since there was no relationship with plant growth additional explanatory variables were deemed extraneous. Temperature, on the other hand, is expected to influence not only plant growth but, according to the central hypothesis, risk of food poisoning. However we have now added three key parameters reflecting plant growth – mean growing season, temperature seasonality and precipitation seasonality – which have previously been shown to be correlated with cultural diversity (Hua et al 2019 Nature Communications).

Related to the point about the poverty <-> spiciness correlation, I would also like to see it clarified whether, in the correlations cited and given in the table in the main text, relatedness and geographical location have been the correlations cited and given in the table in the main text, relatedness and geographical location have been

R1.8: To clarify the results and make our manuscript simpler to interpret, we have now removed any correlation statistics that do not include correction for autocorrelation (including both spatial autocorrelation and phylogenetic independence). This means we have also now removed the correlation table, which gave statistics for uncorrected correlations.

Also, the table of correlations presented is just univariate, but sometimes it is mentioned that other factors were added into the model (e.g. “…adding life expectancy to a model with either historical pathogen load…”, lines 76–77). I have two major points: (a) were factors included together in models in a systematic way, or were most models univariate? If they were not, I think maybe you should do this for at least a few models. You probably need to check what correlations remain significant when other factors are included, and what drops out.; (b) can you please provide a full table of ALL the analyses, with full results, as an excel table in the supplement? Although the coloured table is nice, it is quite hard to read and does not show all of the analyses done. A table summarizing correlations in the main text might be nice too.

R1.9: Figure S2 was intended as simply a descriptive overview of the data rather than a report of the main analysis, serving only to illustrate the degree of covariation in the data. However we agree with the reviewer that the inclusion of these correlations confused rather than clarified the results, and we have now removed it.

We have now added a full description of all models run and all results of models in the Supplementary Information (Table S3) so that the analysis can be followed step by step.

Overall, please included all the RAW data, and detail exactly ALL the analyses. I found this was lacking in the supplement and needs clarifying.

R1.10: All data analysed in this study have been previously published and we provided details these sources of recipe data in Table 1, and all sources of environmental and socioeconomic data in the Methods, so that any reader can access all of the data we have analysed (e.g. reviewer 3 notes as a positive point that the data are all "shareable").

We have now added a more detailed description of all analyses run to the Extended Data, as described above.

Phew, well, that's quite a lot of stuff in the end I guess, but this is actually all minor. I like the paper a lot, and I think the overall analyses conducted and drive of the paper is very good. I hope my comments and suggestions can help, and make it more robust to criticism and misinterpretation.

Reviewer #2:
Remarks to the Author:
In this manuscript, authors put into test the adaptive cuisine hypothesis, which states that spice use is higher is hotter countries because of its beneficial properties to humans since they could reduce the risk of infection from spoiled meat in hotter climates. This hypothesis has been championed by the now-classic and popular work of Billing & Sherman. This hypothesis is supported by a positive correlation between spice use (in traditional recipes containing meat, before refrigeration) and mean annual temperature, especially of those spices with more antimicrobial properties. However, since similar cultures tend to share similar features, it could introduce an unintended autocorrelation problem that Billing & Sherman acknowledge but lacked suitable means for correcting for geographic clustering or autocorrelations between similar cultures, a common statistical problem.

In this manuscript, authors grew and curated a more extensive database than Billing & Sherman, by including more spices. They found that mean spices per recipe was correlated with mean annual temperature for countries and regions, replicating similar observations from data of Billing and Sherman. However, this significant correlation did not survive when spatial and phylogenetic non-independence was taken into account. Surprisingly, they found that poverty provides a more reliable predictor of spice use than poor health outcomes or risk of infection.

Even the number of road traffic deaths provide a stronger prediction of spice use than infection threat or malnutrition. Thus they demonstrate "that there is currently no compelling evidence to assign a special explanatory role for risk of infection in variation in the spiciness of cuisines because other socioeconomic indicators of poverty and poor health outcomes provide a better explanation of variation in spice use than pathogen load or incidence of foodborne illness"

The take-home message is that correlation in the context of Darwinian gastronomy should be interpreted with caution. This observation minimally supports the idea that in present refrigeration times, the association between spice use in hotter countries as a tool to mitigate infection risk could NOT satisfactorily be explained in terms of its

antimicrobial properties. Thus, current spice use is not reflecting an ongoing response to levels of microbial threat rather it seems to follow historical inertia and cultural inheritance of food preferences.

This is a provocative and exciting manuscript that certainly challenges the adaptive cuisine hypothesis. However, I think the manuscript needs some important improvements and further discussion of the main findings

Main concerns
First, the title is misleading since it generates the expectation that a satisfactory answer was found for the question of "Why do hot countries have spicy food?" However, and as authors acknowledged, they could not conclude that poverty or car accidents increase spice use!
1) A new title is needed to reflect the main finding that the correlation between spice use and annual temperature is due to geographical and cultural autocorrelations.

> **R2.1:** We have changed the title to "Why do hot countries have spicy food? The challenges of interpreting cross-cultural correlations."

2) In this regard, after reading the manuscript, one is left with no alternative explanation for why hot countries have (or consume more) spicy food. I think authors should discuss in more detail that the lack of correlation between spice use and temperature, after correction, gives more support to other alternative hypotheses:
Perhaps due to their chemesthetic sensory properties (which applies to all countries and human beings) and/or due to cultural transmission of food preference, etc..

> **R2.2:** An important conclusion of this paper is that it's very difficult to prove any particular hypotheses for cross-cultural data when so many variables covary with each other – including the inevitably large number of unmeasured variables. Instead, the best we can do is show that some potential explanations have more or less support in the data than others. In this, the infection risk hypothesis has less support in the data, but this does not mean that we can say that this increases support for poverty or any other variables as indicating a causal factor – the true causal factor may vary with GDP and other socioeconomic indicators. We now make this point more clearly in the paper, in both the Results and Discussion.

It does also not discuss the possibility that perhaps the initial higher spice use in hotter countries indeed helped mitigate risk infection, but currently, in post refrigeration times, it just follows cultural food preference transmission mechanisms.

> **R2.3:** Actually our results do shed some light on this possibility. If this hypothesis was true then we would expect spice use, shaped by historical cultural preference, to vary with historical pathogen load more than current socioeconomic indicators, but our analysis provides no support than historical pathogen load has stronger explanatory power of spice use than any other variable. In fact the recent socioeconomic data has higher explanatory power for spice use than historical pathogen load. Note that the Billing and Sherman and Ohtsubo data specifically targeted traditional recipes, but other databases are more general and may include both modern and traditional recipes, and yet we find the same patterns across all datasets analysed. We also point out that if it was true that high pathogen load in past times shapes current recipe tradition, then other ingredients that act against foodborne pathogens such as vinegar and alcohol should also correlate with historical pathogen load, but we find no evidence that this is the case. We discuss the lability of cultural responses in the Discussion.

3) Also, a central argument of Billing & Sherman's hypothesis was the correlation between a single spice (hot chili peppers use) and annual temperature. This is an important point that needs to be addressed and explored in your database before the paper can be accepted for publication. It does such a correlation exists, and if it does, it did survive statistical correction?

R2.4: As described in point E2 above, it is not clear from our reading of the original Darwinian Gastronomy studies that chilli does play a central role in Billing & Sherman's hypothesis – they perform a single variable correlation on chilli also for other spices including garlic, onion, anise, cinnamon, coriander cumin, ginger, lemongrass and turmeric (Billing & Sherman 1998, p 14, also figure 5). However we now test chilli against foodborne illness, GDP and life expectancy (which are the best predictors of spice use).

Reviewer #3:
Remarks to the Author:
The aim of this study is to test the "adaptive cuisine hypothesis" using a global dataset consisting of recipes from 45 cuisines. This study found that poverty and poor health outcomes were better able to explain spice use than infection risk. While the research approach is conceptually novel by considering (to my knowledge) socioeconomic and other "cultural" factors in predicting spice use, there are several flaws that I think should prevent this paper from being published in Nature Human Behavior.

One issue I have with this manuscript is the general lack of clarity about the study motivation, data, analyses, and how measures were defined. For example, the main manuscript and methods files uses terms biocultural diversity, biological diversity, and cultural diversity interchangeably. The authors should note that there is considerable difference between these terms. Moreover, the authors appear to apply these terms in reference to linguistic diversity, and sometimes botanical diversity (I think). The authors should avoid using broad categories unless they define them and use them consistently throughout the study.

R3.1: All of these aspects of diversity – cultural diversity, biodiversity and crop diversity - were represented by separate variables. We have now made it clear what role each of these variables plays in the analysis and emphasized that we are not conflating biological and cultural diversity. In our diagram (Figure 3), we assign cultural diversity to a different category than biodiversity.

As another example, the third aim of the paper states that they use "statistical techniques" to explicitly compare the explanatory power of "different hypotheses" for higher spice use, considering not only infection risk mitigation, but also history, proximity, socioeconomic factors, cultural diversity, botanical, and agricultural diversity. What are the statistical techniques?

R3.2: Given the restrictions on word count, we had referred to previous publications where the methods were fully explained. We have now expanded our description of methods substantially both in the main text and in the supplementary methods, and we provide a full description of every statistical test conducted in the Supplementary Information (Table S3)

What are the hypotheses? Why did they choose these measures?
These "hypotheses" are never spelled out – what are the authors actually predicting aside from the adaptive cuisine hypothesis? (Which also suggests that the hypotheses were not pre-registered)… Which leads me to the next point…

R3.4: The central aim of our study is to ask whether the significant correlation between spice use and temperature (and with historical pathogen prevalence) provides convincing evidence for the adaptive cuisine hypothesis. Because there is a known significant correlation between these variables, but we wish to examine the underlying cause of that correlation, our hypothesis testing does not take the simple, traditional form of asking whether there is a relationship between two key variables, but instead must take a more exploratory approach. Therefore our task was to consider all possible alternative explanations for the relationship between spice and temperature for which we could find appropriate representative variables.

We have now substantially expanded our description of this hypothesis testing rationale in the Results and Methods section, and we provide details of every test conducted in Table S3. We also provide the flow chart we used to plan the analysis, which indicates the possible explanatory links that we explore between spice and infection risk (Figure 3).

I am more concerned about the lack of a priori justification for the variables chosen for this study. There is nothing suggesting why socioeconomic variables and other measures should be included in the study. The authors went through a lot of work to find these measures, but the study seems exploratory in nature. The way the study is currently presented – with no alternative motivation or literature review on alternative approaches – makes it seem that the authors cherry-picked some variables and threw them into a model.

R3.5: We apologise that the brevity of the initial submission did not allow us to fully explain the rationale behind variable choice. We now include a longer description in the methods and supplementary methods and have included the flow chart that we used to plan the analysis strategy in order to collect the appropriate variables. There are few published studies that directly test these links that we can refer to as there have been relatively few attempts to test these hypotheses until now – we have referred to all of the published tests of this hypothesis that we are aware of, and we have included all of the variables that were considered in those publications (spice, temperature, parasites, vinegar, area). We have also increased our references to the literature where these potential links are discussed, both in the text and in the legend to Figure 3, so that we have now nearly doubled the number of references. No other study has considered such a broad range of interacting factors.

However, it's hard to know for sure because of the opaque presentation of the results and lack of explanation for how analyses were conducted (I don't think it's good practice to point to other studies rather than clearly outlining your analytic approach). Specifically, I am left wondering what statistical tests were conducted? It seems as though a t-test was used for some analyses, and that a multivariate test was used for others. The authors then report AIC results and point to Figure 2, but there are no additional models? Why report an AIC value (model-level) if the emphasis is on GDP(predictor variable)?

R3.6: Again we apologise for the brevity of the initial submission, we have now included a full description of all tests conducted, with the rationale and results of every step of the analysis. Table S3 contains all steps of the analysis with all descriptive statistics, repeated for four alternative datasets.

Most importantly, the variables that were selected seem to be proximate measures to explain the use of spices, but do not test the ultimate evolutionary explanation for why spice use evolved in warmer climates. The authors do not discuss this obvious discrepancy.

R3.7: We agree with the reviewer that these variables are likely to represent proximate measures not ultimate causes, and we emphasize this point in both the Results and Discussion. As for points R1.5 and R2.2 above, we now put additional emphasis on the challenges inherent in trying to make clear causal statements based on significant cross-cultural correlations, given the vary strong covariation between factors associated with socioeconomic, biodiversity, climate and cultural diversity. Therefore it is not our aim to identify the ultimate evolutionary explanation, which is rarely accessible to direct investigation, but instead to weigh the relative explanatory power of different variables associated with spice use.

We would be happy to add any additional variables to the analysis if the reviewer feels that any particular line of investigation has been missed.

Furthermore, the reference the authors include for using alcohol and vinegar indicates that wine is shown to have antimicrobial activity, but beer does not. The current study appears to collapse all types of alcohol into one variable, but the literature cited by the authors suggests this is not a well-reasoned choice.

> **R3.8:** We have now divided the alcohol+vinegar variable into three separate variables - vinegar, alcohol, and non-beer alcohol - and analysed each separately against the potential explanatory variables. None of these variables shows an association with variables associated with infection risk.

Also, why is alcohol/vinegar use not pre-specified as a hypothesis? Was this a post-hoc analysis? It's also not clear why they are analyzed as a proportion.

> **R3.9:** Alcohol and vinegar were indeed prespecified hypothesis, as these predictions have been made in the literature and these studies were referenced in the text as providing the impetus for including these variables in this study. In particular, Ohtsubo states that exploring whether other antimicrobial agents such as vinegar show the same pattern predicted for spices was a primary purpose of their study (see Ohtsubo 2009, p678). In addition to providing a test of a previously published prediction, testing the association between alcohol and risk of foodborne illness is a valuable way of getting around the problem of the covariation between cultural, environmental and socioeconomic factors, because it acts essentially as an independent response variable: if antimicrobial action plays a role in the cultural evolution of cuisine, then it should apply to many ingredients with antimicrobial action such as vinegar, alcohol and salt. Salt, as all previous researchers have mentioned (e.g. Billing & Sherman, Ohtsubo), does not provide a clear test because it has so many purposes and is often either not mentioned in recipes or added in highly variable amounts. But vinegar and alcohol ought to show similar results to spices if the value of antimicrobials in areas of high parasite load is a major driving factor shaping cuisine.
>
> We use proportion of recipes containing alcohol because presence or absence of alcohol per recipe is a binary variable, unlike spices which (following Billing and Sherman) are continuous count data. The average of a binary variable is the proportion of cases in which it is present.

The sample size for this study was small. The majority of these analyses are conducted on half of the dataset. Most of the raw data from Billing and Sherman are not available. This leads to a breakdown of the following: 8/45 cuisines are from Japan, 8/45 are from India, 10/45 are from a "global" data set, and the remaining 19 are Chinese data. The remaining analyses were conducted on this data set and there was no effect for pathogen load. There was likely no effect because the variance for pathogen load in the Chinese data was 0. The same is true for Japan and India.

> **R3.10:** We have now nearly doubled the number of datapoints for the core analyses (from 45 to 70 cuisines). We repeat all analyses on four datasets – all cuisines, Billing and Sherman's original data, only country-level data (no subnational cuisines) and only recipe-derived data (to ensure consistency of spice counts). All datasets show essentially the same patterns, except that the expanded dataset (now with 70 cuisines) does not show significant support for an association between spice use and foodborne illness. We also conduct a formal power analysis showing that the recipe-derived data (45 cuisines) has almost the same power as the larger dataset (70 cuisines) and higher power than the country-level data.

Another concern that I have is with the time frame that the data were collected from each of the different sources. Billing and Sherman are data from 1998, but the WHO data and other sources are later. For example, life expectancy was based on the most recent data for each location – this is problematic for comparisons to average spice use that occurs at different periods of time in each data set. Also, the life expectancy data are from different sources – the limitations of this must be addressed. The authors got one set from the World Health Organization and the others via a source from Wikipedia.

> **R3.11:** Billing and Sherman's study was published in 1998 but the recipes they use are from

"traditional cookbooks", so it is difficult to assign a particular age to them. Subsequent studies of the adaptive cuisine hypothesis have either used recipes from internet databases (Zhu et al), cookbooks or from interviews with "housewives" (Ohtsubo). Our additional databases use a mix of books and internet recipes (Jain, Bagler). Our analyses show no difference in patterns of spice use between the traditional recipes of Billing and Sherman with any of these other datasets.

World Health Organisation life expectancy data goes back to the year 2000, and World Bank data on life expectancy goes back to 1960, so we could analyse these data for the country level cuisine data. However, we are not convinced that would give a clearer picture of the relationship between spice use and general health outcomes. The adaptive cuisine hypothesis expects that there is an ongoing association between a cultural tradition of spice use and infection threat, but clearly spice use is not static over time. As explained in the manuscript, spices used change over time with availability and fashion, and continue to do so.

Our regional data is from National Bureau of Statistics of the People's Republic of China, United Nations Development Programme, the Ministry of Health, Labour and Welfare of Japan, and the US Institute for Health Metrics and Evaluation. For some of these sources, the government databases are not directly searchable, but the data has been republished on websites such as StatisticsTimes.com, stats-japan.com and Wikipedia. We only use these republished data when the variables where the sources are clearly identified from a single agency. One of the valuable aspects of sites such as StatisticsTimes and Wikipedia is that they act as an access point for many sources of government data that are otherwise not readily accessible online (for example, which are made available by print but not in open access online databases). It would not be possible for most researchers to access these government data without open access statistics databases such as these.

Finally, this paper lacks a strong conclusion along with a statement about the limitations of the research. Overall, I do not think the way the data are presented provides a clear and convincing answer to the research question. I also do think the findings are over-generalized, as a majority of the data comes from three Asian countries, and does not actually offer an ultimate explanation for the evolution of spice use; at best, it might provide a proximate explanation. This is fine, but should be made transparent.

R3.12: We have now rewritten much of the Discussion and reworded our conclusion.

Our dataset is not unbiased with respect to geographic location of cuisines, but it is less so than previous analyses. We present results based on the original Darwinian cuisine dataset (Billing and Sherman), on country-level data, and on regional data. Figure 1 clearly shows the geographic spread of the cuisines included in this study. Less than half the points are from Asian countries. By contrast, 60% of the world's population live in Asia, so it could be argued that Asian cuisines are under-represented. For example, European cuisines are over-represented (20% of the points are European cuisines, when only 10% of the world's population lives in Europe) and India is under-represented (10% of the points are from India when 20% of the world's population live in India). But we have sampled global cuisines as comprehensively as we could, and our dataset is much less geographically biased than previous studies that focussed on single countries (China, Japan) or global datasets biased toward European cuisine (Billing & Sherman's data was one third European cuisines).

The only generalisation we can draw from our study is the very great challenges of inferring causal mechanisms from cross-cultural correlations when so many variables strongly covary. However, we do believe it is possible and fruitful to examine the relative strength of support for different causal explanations. Our primary conclusion in this paper is that there is no compelling evidence to privilege the adaptive cuisine hypothesis – that spice use is an adaptive response to risk of infection – because other variables provide as good or better correlation with spice use, and corollary predictions of the hypotheses (such as vinegar use) are not supported from our analyses. We agree that it is difficult to be certain of identifying ultimate causes

from cross-cultural correlations, and we hope we have now made this point even stronger in the revised manuscript.

Minor comments:
Ln 75: use of “that” when authors mean “than”

Data are sharable, which is good.

Decision Letter, first revision:

14th August 2020

Dear Professor Bromham,

Thank you once again for your revised manuscript, entitled "Why do hot countries have spicy food? The challenges of testing hypotheses with cross-cultural data," and for your patience during the peer review process.

Your manuscript has once again been evaluated by our 3 reviewers, whose comments are included at the end of this letter. Although two of our reviewers are satisfied with the revisions, Reviewer 1 raises some important outstanding concerns. We are interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised manuscript before we make a decision on publication.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritised set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

1) Please thoroughly address Reviewer 1’s concern that you have not adequately discussed the GDP-spice correlation in terms of whether it may be driven by colonial history (notably the extractive colonisation that may have been in part driven by spice production in some regions) and/or geographical bias in the data.
2) Please ensure that data paucity is explicitly highlighted as a major limitation in your study and make it clear this constitutes a very clear and strong caveat in the discussion and conclusions, as requested by Reviewer 1.
3) Please include the table suggested by Reviewer 1, with key findings from all models corrected for autocorrelation.
4) Finally, please address all other queries raised by Reviewer 1, including their points about data transparency and availability, and the validity of your language-based trees.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its

consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within six weeks. We understand that the COVID-19 pandemic is causing significant disruption for many of our authors and reviewers. If you cannot send your revised manuscript within this time, please let us know - we will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

• Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,

Charlotte Payne
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Cultural evolution; Macroevolutionary methods

Reviewer #2: Spice and pathogen resistance

Reviewer #3: Spice and pathogen resistance

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
N.B. Fully formatted version attached.

Overall, I am pretty happy with the work the authors did in addressing my comments and those of the other reviewers, especially the number of novel analyses conducted to rigorously test potential hypotheses. I am also particularly glad to see the extended methods and data in the supplement. Below are my replies to their response, and a few additional points which I consider still need addressing.

Replies:

-I take all the points made re colonial history. I was not expecting that the author do further analyses, and indeed they would be very complicated, as the authors make clear. I was rather hoping that they would discuss the potential impact of colonial history, especially in relation to the GDP-spice correlation (see below). For example, much of South and South East Asia was colonized in part *because* of its spice production, and so some of the correlation between low GDP and spice use may relate to this (extractive institutions). As I said, a discussion is fine. If the authors were very keen to test it, then I would suggest something simple, like “was/was not an extractive [i.e. non-settler] colony”, or “number of years as extractive [non-settler] colony” as a variable, and then to test the correlation with GDP after any correlation with this variable is removed.

-I appreciate the testing of the number of spices growing within a country as a variable.

-I am broadly pretty happy with the increased nuance of the discussion, and the clear emphasis that the study shows little impact of antimicrobial effect on spice use and that other associations are complex, but see below.

-While I am glad the authors took on board the point about the confusion between results corrected and not corrected for autocorrelation (R1.9), I think some kind of table would be useful in the main text because otherwise one is searching for the “final” results in the text itself. Is it possible to somehow summarize the key findings from the excel sheet with the different models, with variables starred for significance or something? Only summarize the models corrected for autocorrelation. Also can you state at the start of the results that all results are with autocorrelation taken into account?

-I am very happy to see a much extended supplement (R1.10). However, in the interests of a fully open approach to science, I would urge the authors to make public their full raw data (and if there are steps in the calculation, then the data at each step), as this allows others to easily check their results

and take them further. If they are sensitive for some reason (e.g. forthcoming work), I would urge the authors to set a timeframe for publishing the data at a predefined DOI.
Further points to address:

1. It is still not, as far as I can see, explained in words in either the main text or supplement the approach taken to model building. In one of the supplemental excel sheets there are questions linked to model constructions, but they are not explained verbally. Given this is a key part of the hypothesis construction and testing, please spell this out clearly, ideally in the main text (in at least abridged form) and certainly (and in full detail) in the supplement.

2. The authors also did a valiant job of trying to address the "issue" of their rather robust correlation between GDP and spiciness, and I very much like the "network" diagram showing potential causal relationships. I understand that the correlation does not readily lend itself to explanation, and they are right in their nuanced and cautious discussion! However, I am still not completely satisfied they have fully addressed this issue – there must be a reason for this correlation, even if it is historical contigency or data/analysis issues. Looking at the raw data, one would certainly expect this correlation because of "bland" western European origin cuisines and Japanese cuisines (and these are money-rich places) and then a lot of spicier S/SE-Asian cuisines (and these are money-poorer places). What is more surprising is that this correlation remains after removal of spatial and cultural autocorrelation…

One simple point: the GDP for "Jain" seems one order of magnitude too high, which may affect things. Another: was the correlation test done on logged data? (I know the figure is, but please make clear somewhere exactly which variables were logged) Furthermore, there are issues with the language tree (see below), which might affect this, and I would encourage the authors to consider running trial analyses with an adjusted tree. Furthermore, could the authors provide the final latitude and longitude data used as the geographic location for each cuisine area, so that it can be checked by other workers? It also appears that the authors did not test the association between GDP and spiciness at a country level – although their effect should *in theory* removed by the Hua et al./Freckleton & Jetz approach, I wonder if the repeated sampling in Japan and India may be driving part of this correlation (and, although of course there is a lot of variety, both countries also share a *lot* of cultural traits between cultures).

Still on this point, perhaps it is possible that the "missing factor" seemingly linking GDP and blandness here is in fact just geographic location? What happens if the authors test the *relative* importance of geographic location (not just latitude) and GDP (see, e.g. Sookias et al. DOI: 10.1098/rsos.171411)? Is GDP a better predictor than country location? Also, although again in theory adjusted for by using the language phylogeny, recent colonial history may be influencing the outcome – what happens if former European settler colonies (i.e. where the local people were displaced by whatever culture has brought the cuisine – Australia, Caribbean, North America, South Africa, Brazil and to a lesser extent Mexico) are removed?

Finally, the rather limited global sampling (see below) may also be playing a role here. For example, off the cuff, adding in say Singapore (which I imagine might have spicy food and be rich) and Nicaragua (which has rather bland food but is poor) could make a difference…I would *not* suggest the authors try to expand their sample (unless very straightfoward), but please at least discuss this possibility. As pointed out by another reviewer, even in the expanded data there is a strong sway in the dataset to China, then India and Japan. Indeed, almost 1/3 of the data is from China, even in the

expanded dataset! (and see above and below) Although of course China has varied cuisines and a large geographic area, so does, for example, South America or the Arab world, and these are currently represnted by one and two samples respectively!! Please at least flag up these weaknesses of the dataset clearly (see below)

3. So, my queries about the language-based trees are:

-It is mentioned that the more resolved grouping of Danish and Swedish as “East Scandinavian” is used, but in the trees stated these languages are in a polytomy with Norwegian (this is unlikely to affect results – just needs justifying in text)

-Why was Mughlai cuisine placed with Gujarati+Rajasthani? The language of the Mughal empire was Hindustani (=Urdu/Hindi), which is not especially close to Gujarati/Rajasthani is it? Also in recent analyses (e.g. Grey and Atkinson 2003, DOI: 10.1038/nature02029) Hindi is placed closer to Punjabi (again, I guess this will not make a difference to results)…

-Why is Caribbean cuisine not grouped with Indo-European languages? All of the countries mentioned have either English, Spanish or French as their first language.

-Ireland seems to have been counted as having a Celtic language, despite by far the majority speaking English as a first language. What was the justification here? I mean Mexico I think was counted as Spanish, despite this not being the “native” language. Indeed, indigenous languages are in better shape in Mexico than in Ireland… I would advise trying a tree with Ireland counted as English, and one where Mexico is considered say Nahuatl.

-Why was Ethiopia not grouped with Semitic languages, given Amharic is Semitic?

-While I see that glottolog and ethnologue both do not make this grouping, in recent computational language phylogenies (e.g. Bouckaert et al. 2014 DOI: 10.1126/science.1219669) Germanic+Romance+Slavic+Celtic always form a group to the exclusion of Indo-Iranian, and Romance+Celtic usually form a group. This could make a difference to results, and I would encourage the authors to test this.

4. Finally, and probably the most important point: I think one of the major limiting factors is still the data, as was raised by other reviewers and which you attempted to address. However, I think the global coverage is still very limited, and this must be made clear in the discussion. The exclusion of much (or rather *most*) of Africa, Europe, the Middle East, central Asia, the Caucasus/Anatolia, and South America (even if for the reason of ensuring data quality) means that reaching firm conclusions, especially at a global scale, is extremely hard. For example, I am pretty sure that food gets spicier as one moves north in South America (and maybe Africa), but this is not captured in the study. As mentioned above, N70 dataset is strongly biased towards China, Japan and India, and although you eliminate this bias in the country-level data, these are not what the main discussion seems to be based on, and at country level the global coverage is poor. Regarding the point you made in response to another reviewer’s comment about higher population in Asia, whilst this is true, the Asian spread of the data is also not wide outside India, China and Japan, and the relevance of population is debatable – given that neighbouring and culturally-related cuisines are probably likely to be similar (which is why you took out the possible autocorrelation), a wide geographic and cultural spread is presumably more important than weighting by population?

I am aware this was a limitation of the sources upon which the study relied, and that similar biases were present in previous work, and this does not undermine the study overall because it is a "tenative step" towards better understanding, but please make this a *very* clear and strong caveat in the discussion and conclusions. It may also be leading to the rather odd correlation between GDP and spiciness which I refer to above. Future worthwhile work for new studies could actually be collecting more extensive and comparable cuisine data first hand/from recipe books from across the world, to allow more accurate statistical work to be undertaken – perhaps mention this as a future avenue.

Reviewer #2:
Remarks to the Author:

Authors did a great job revising the manuscript. The new title and the addition of Chilli in the analysis improve the paper. I have no further comments. Congratulations!!!

Reviewer #3:
Remarks to the Author:
Thank you for addressing my comments. The manuscript has greatly improved and I look forward to seeing this published.

Author Rebuttal, first revision:

Why do hot countries have spicy food? The challenges of testing hypotheses with cross-cultural data
Changes made in response to Editor's and Reviewer's comments

---

1) Please thoroughly address Reviewer 1's concern that you have not adequately discussed the GDP-spice correlation in terms of whether it may be driven by colonial history (notably the extractive colonisation that may have been in part driven by spice production in some regions) and/or geographical bias in the data.

E1 We have now added two additional analysis to the supplementary information, and we discuss this point in the Results. We have expanded our discussion of how the analysis deals with geographic bias.

2) Please ensure that data paucity is explicitly highlighted as a major limitation in your study and make it clear this constitutes a very clear and strong caveat in the discussion and conclusions, as requested by Reviewer 1.

E2 We have added a discussion of this point to the main text, with some analyses and a new figure. We are happy to take your advice on whether this figure should be in the main text of the supplementary information, but given the importance of this point we have put it in the main text at this stage.

3) Please include the table suggested by Reviewer 1, with key findings from all models corrected for autocorrelation.

E3 We have added the table as requested (Table 2)

4) Finally, please address all other queries raised by Reviewer 1, including their points about data transparency and availability, and the validity of your language-based trees.

E4 We have added a statement of data availability (see details below), and we have conducted additional analyses with the changes to the language trees suggested by the reviewer.

Reviewer expertise:

Reviewer #1: Cultural evolution; Macroevolutionary methods

Reviewer #2: Spice and pathogen resistance

Reviewer #3: Spice and pathogen resistance

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
N.B. Fully formatted version attached.

Overall, I am pretty happy with the work the authors did in addressing my comments and those of the other reviewers, especially the number of novel analyses conducted to rigorously test potential hypotheses. I am also particularly glad to see the extended methods and data in the supplement. Below are my replies to their response, and a few additional points which I consider still need addressing.

Replies:
-I take all the points made re colonial history. I was not expecting that the author do further analyses, and indeed they would be very complicated, as the authors make clear. I was rather hoping that they would discuss the potential impact of colonial history, especially in relation to the GDP-spice correlation (see below). For example, much of South and South East Asia was colonized in part *because* of its spice production, and so some of the correlation between low GDP and spice use may relate to this (extractive institutions). As I said, a discussion is fine. If the authors were very keen to test it, then I would suggest something simple, like “was/was not an extractive [i.e. non-settler] colony”, or “number of years as extractive [non-settler] colony” as a variable, and then to test the correlation with GDP after any correlation with this variable is removed.

> *R1.1 We have now added two analyses to address this point. Firstly, if the relationship between GDP and spice we observe in our data is driven by the historical spice trade then the relationship should disappear when we remove these datapoints, so we rerun the analysis of spice and GDP without these Spice Trade colonial powers (in our dataset, United Kingdom, Spain, Portugal) and without the Spice Islands they occupied (Indonesia, Malaysia). Secondly, as the reviewer suggests, we add a variable to the analysis that records whether a cuisine area has been an extractive colony (we found the “settler” definition difficult to apply as all such colonies involved some resident families settling in the colony, so it is a matter of degree – we explain our definition on page 8-9 in the supplementary information). We then test if the relationship between GDP and spice exists in cuisines that had not been under colonial control.*
>
> *We summarise our findings on page 5-6 as follows: “Another possible explanation is that places where spices grow naturally were the target of European colonial occupation and exploitation. This pattern of aggressive resource capitalisation in the “Spice Islands” by predominantly Dutch, English, Spanish and Portuguese traders resulted in the increase in wealth in the European countries but may have come at the expense of wealth accumulation in the source countries for spices (represented in our dataset by Indonesia and Malaysia). However, the association between spice use and GDPpc is still significant after removing cuisines associated with the Spice Trade, both the colonizers and the colonized (ß=- 0.39, t41=-4.13,*

p<0.001, BF=11.8), so the pattern of colonial exploitation association with the Spice Trade cannot provide a general explanation for the negative relationship between GDP and spice use. Furthermore, the relationship between spice use and GDPpc holds for a dataset consisting only of cuisines from areas that were never under colonial rule (ß=-0.39, t31=-3.07, p=0.004, BF=5.7), so patterns of colonization seem unlikely to provide a general explanation of the association between spice use and poverty. "

-I appreciate the testing of the number of spices growing within a country as a variable.

-I am broadly pretty happy with the increased nuance of the discussion, and the clear emphasis that the study shows little impact of antimicrobial effect on spice use and that other associations are complex, but see below.

-While I am glad the authors took on board the point about the confusion between results corrected and not corrected for autocorrelation (R1.9), I think some kind of table would be useful in the main text because otherwise one is searching for the "final" results in the text itself. Is it possible to somehow summarize the key findings from the excel sheet with the different models, with variables starred for significance or something? Only summarize the models corrected for autocorrelation. Also can you state at the start of the results that all results are with autocorrelation taken into account?

R1.2 We have added a table as requested (Table 2), and the results are also summarised in the legend to Figure 2.

We stated in the results on 5 "All results reported here explicitly model autocorrelation due to relatedness and proximity" but we have now repeated this on page 4 in the first paragraph of the results and in all tables and figures.

-I am very happy to see a much extended supplement (R1.10). However, in the interests of a fully open approach to science, I would urge the authors to make public their full raw data (and if there are steps in the calculation, then the data at each step), as this allows others to easily check their results and take them further. If they are sensitive for some reason (e.g. forthcoming work), I would urge the authors to set a timeframe for publishing the data at a predefined DOI.

R1.3 We appreciate the reviewer's request for open data: all of the data we analyse has previously been published and in principle we are happy to provide it. We already provide all summary data that would allow anyone to redo our analysis (Tables S4 and S5). However two of the datasets were provided to us by their original authors (Zhu and Ohtsubo), and we do not have explicit permission from those authors to pass these datasets on to a third party. Therefore we have added a statement that the data is available from the authors and if requested we will provide the raw data from the internet databases and will contact Ohtsubo and Zhu for permission to distribute their data.

Further points to address:

1. It is still not, as far as I can see, explained in words in either the main text or supplement the approach taken to model building. In one of the supplemental excel sheets there are questions linked to model constructions, but they are not explained verbally. Given this is a key part of the hypothesis construction and testing, please spell this out clearly, ideally in the main text (in at least abridged form) and certainly (and in full detail) in the supplement.

R1.4 We are not clear what the reviewer means by "model building", but we wonder if they are asking whether we took a multiple regression approach where all variables are added and through a process of winnowing, reduced to a set of informative variables. But this is not the analytical approach we took in this paper. In fact our approach is much simpler and more intuitive – as described on page 2 "Our approach

consists of a series of tests, each designed to probe a possible explanation for the observed relationship between spice use, temperature and infection risk."

We have now added the following statement that emphasises that our analysis consists of a series of tests to page 4:
"Each test asks whether spice use is related to a particular environmental or socioeconomic variable, and, if it is, whether that relationship might be explained by covariation with another relevant variable. In this way, we evaluate the explanatory power of each variable in a step-wise fashion, rather than including all possible variables in a single multivariate analysis."

Each step is also explained verbally in the legend to Figure 2, and with numbers in the new Table 2.

2. The authors also did a valiant job of trying to address the "issue" of their rather robust correlation between GDP and spiciness, and I very much like the "network" diagram showing potential causal relationships. I understand that the correlation does not readily lend itself to explanation, and they are right in their nuanced and cautious discussion! However, I am still not completely satisfied they have fully addressed this issue – there must be a reason for this correlation, even if it is historical contigency or data/analysis issues. Looking at the raw data, one would certainly expect this correlation because of "bland" western European origin cuisines and Japanese cuisines (and these are money-rich places) and then a lot of spicier S/SE-Asian cuisines (and these are money- poorer places). What is more surprising is that this correlation remains after removal of spatial and cultural autocorrelation…

R1.5 Indeed we agree there must be a reason for the correlation, but it would be misleading of us to claim that we know what it is. We expand our discussion of this point on page 7-8 as follows. "indicators of poverty and poor health outcome, including GDP, life expectancy and road traffic deaths, provide stronger explanatory power for variation in spice use than indicators of risk of infection, including foodborne infection, early childhood diarrhea and historical disease prevalence. But the covariation between socioeconomic and cultural parameters makes it difficult to ascertain the causes of the patterns with any great certainty. We can never include all relevant cultural, socioeconomic or environmental variables, so we cannot be certain that observed correlations between variables are not an indirect result of covariation with an unmeasured quantity. Even though road traffic deaths are a better predictor of spice use than infectious disease, we do not expect there is a direct causal relationship between traffic accidents and spice use. Instead, it seems likely that we could find additional socioeconomic or cultural variables that provide even stronger correlation with spice use, but if we attempted to include all relevant cultural and socioeconomic parameters, the number of variables would eventually outnumber the number of cuisines, increasing the risk of spurious correlations."

We hope our study will prompt other researchers to develop and test other hypotheses.

One simple point: the GDP for "Jain" seems one order of magnitude too high, which may affect things.

R1.6 We are very grateful to the reviewer for their attention to detail in detecting this error, which arose from a spreadsheet error. We have corrected this in the data table and checked that all analyses use the correct figure. We have also checked that there are no other errors in the database by checking variables against original data source. We found the Jain Life Expectancy was also incorrect so we have corrected that and rerun all analyses that use either GDP or Life Expectancy. The results have not changed appreciably but we report all new analysis statistics and have corrected Table S4 and Figure 3. Thank you very much for pointing this out so it could be corrected before publication.

Another: was the correlation test done on logged data? (I know the figure is, but please make clear somewhere exactly which variables were logged)

R1.7 We stated in the methods (previous version: page 17 line 554: this submission page 17 line 851): "Some variables are right skewed, so we applied log transformation to each of these variables if transforming the variable increases the maximum likelihood of the model that includes the variable, because increasing likelihood suggesting a better fit of residuals in the response variable to a multivariate normal distribution." Transformed variables are named with suffix _tr in Table S3.

Furthermore, there are issues with the language tree (see below), which might affect this, and I would encourage the authors to consider running trial analyses with an adjusted tree.

R1.8 We run all analyses on an alternative tree as suggested by the reviewer, and show that it does not affect our results. We describe the additional analysis in the methods, and give details and report the results in the supplementary materials and in Table S3, as follows: " Of course, it is not possible to describe an undisputed hierarchy of relationships between cultures or cuisines, and in some cases a cuisine could be represented by an alternative language that might alter its place in the tree. To test the robustness of our conclusions to changes in the position of cuisines in the hierarchy, we tested key analyses on an alternative tree with the following modifications to placement of cuisines: Caribbean associated with the European languages that are recognised as official languages in many countries; Mughlai grouped with Punjabi; Irish with English; Ethiopia with Semitic languages; and with European languages arranged according to Bouckaert et al. (2014)10."

"To test the sensitivity of our analysis to assumptions made about the relationship between cuisines, we rerun analyses on the combined dataset using an alternative tree, as outlined in Section 1.2. We find that using alternative tree does not change the results qualitatively (Table S3), except that the evidence is stronger for link between spice use and foodborne illness (ß=0.27, t68=2.62, p=0.011, Bayes factor=2.1)."

Furthermore, could the authors provide the final latitude and longitude data used as the geographic location for each cuisine area, so that it can be checked by other workers?

R1.9 Lats and longs of centroids now provided in the data table and mentioned on page 16 in the methods

It also appears that the authors did not test the association between GDP and spiciness at a country level – although their effect should *in theory* removed by the Hua et al./Freckleton & Jetz approach, I wonder if the repeated sampling in Japan and India may be driving part of this correlation (and, although of course there is a lot of variety, both countries also share a *lot* of cultural traits between cultures).

R1.10 We have now included test on the association between GDP and spice use at country level, which gives qualitatively the same results as the combined data. The results are no association between GDP and spice use, but road traffic deaths is still a significantly better predictor of spice use than foodborne illness.

Still on this point, perhaps it is possible that the "missing factor" seemingly linking GDP and blandness here is in fact just geographic location? What happens if the authors test the *relative* importance of geographic location (not just latitude) and GDP (see, e.g. Sookias et al. DOI: 10.1098/rsos.171411)? Is GDP a better predictor than country location?

R1.11 This is the core issue that our analysis is explicitly designed to test – could these relationships between cuisines be driven by geographic location? We incorporate spatial autocorrelation into all of our analyses, using a method that is related to the one described in the Sookias et al paper (that is, ours is a modification of the Freckleton & Jetz method used in Sookias – full details of the method are given in the supplementary information). If location was driving the relationship between spice and GDP then accounting for location would cause the relationship to disappear. Another advantage of this analysis is that it provides gives a statistical measure of the relative importance of geographic location in explaining the variation in

these variables – we now give this point more emphasis on page 3. For example, we can explain the association between spice and temperature through the observation that nearby countries tend to share both similar temperature and similar spiciness of cuisine. But this is not the case for GDP – while nearby countries do indeed tend to have similar GDP, there is a relationship between spice and GDP even when we take the relationship due to proximity into account. We now make this point more strongly on page 17, and we cite the Sookias paper as another example of a study that uses the same method we do to make a point about the influence of relatedness and proximity on similarity between cultural evolution variables.

Also, although again in theory adjusted for by using the language phylogeny, recent colonial history may be influencing the outcome – what happens if former European settler colonies (i.e. where the local people were displaced by whatever culture has brought the cuisine – Australia, Caribbean, North America, South Africa, Brazil and to a lesser extent Mexico) are removed?

R1.12 We have now added a new analysis that addresses this issue– see point R.1 above.

Finally, the rather limited global sampling (see below) may also be playing a role here. For example, off the cuff, adding in say Singapore (which I imagine might have spicy food and be rich) and Nicaragua (which has rather bland food but is poor) could make a difference…I would *not* suggest the authors try to expand their sample (unless very straightfoward), but please at least discuss this possibility. As pointed out by another reviewer, even in the expanded data there is a strong sway in the dataset to China, then India and Japan. Indeed, almost 1/3 of the data is from China, even in the expanded dataset! (and see above and below) Although of course China has varied cuisines and a large geographic area, so does, for example, South America or the Arab world, and these are currently represnted by one and two samples respectively!! Please at least flag up these weaknesses of the dataset clearly (see below)

R1.13 We agree that the study has limited and geographically biased sampling, and we make that point more clearly now, as follows: "Like many global studies of cultural evolution, our study is biased by the available data36. Previous tests of the hypothesis that spicy food is a response to higher temperature have either focussed on a single region (such as China35 or Japan8), or used a global dataset that contained an overrepresentation of recipes from European cuisines (e.g 30% of the cuisines from Europe even though it represents only 10% of the world's population2). We have greatly expanded global representation, in particular our dataset is less biased toward European cultures, and the sampling of cuisines from mainland Asia is more in line with global population (Figure 5). However, cuisines from Africa and South America are relatively underrepresented in our combined database. We hope future studies will fill in these large gaps, and provide more detailed sampling in underrepresented regions."

We also provide a new Figure (Figure 5) which makes these geographic biases clear.

3. So, my queries about the language-based trees are:

-It is mentioned that the more resolved grouping of Danish and Swedish as "East Scandinavian" is used, but in the trees stated these languages are in a polytomy with Norwegian (this is unlikely to affect results – just needs justifying in text)

-Why was Mughlai cuisine placed with Gujarati+Rajasthani? The language of the Mughal empire was Hindustani (=Urdu/Hindi), which is not especially close to Gujarati/Rajasthani is it? Also in recent analyses (e.g. Grey and Atkinson 2003, DOI: 10.1038/nature02029) Hindi is placed closer to Punjabi (again, I guess this will not make a difference to results)…

-Why is Caribbean cuisine not grouped with Indo-European languages? All of the countries mentioned have either English, Spanish or French as their first language.

-Ireland seems to have been counted as having a Celtic language, despite by far the majority speaking English as a first language. What was the justification here? I mean Mexico I think was counted as Spanish, despite this not being the “native” language. Indeed, indigenous languages are in better shape in Mexico than in Ireland… I would advise trying a tree with Ireland counted as English, and one where Mexico is considered say Nahuatl.

-Why was Ethiopia not grouped with Semitic languages, given Amharic is Semitic?

-While I see that glottolog and ethnologue both do not make this grouping, in recent computational language phylogenies (e.g. Bouckaert et al. 2014 DOI: 10.1126/science.1219669) Germanic+Romance+Slavic+Celtic always form a group to the exclusion of Indo-Iranian, and Romance+Celtic usually form a group. This could make a difference to results, and I would encourage the authors to test this.

> *R1.14 We have now made an alternative tree with these changes and reanalysed key results using this alternative tree – see point R.18 above*

4. Finally, and probably the most important point: I think one of the major limiting factors is still the data, as was raised by other reviewers and which you attempted to address. However, I think the global coverage is still very limited, and this must be made clear in the discussion. The exclusion of much (or rather *most*) of Africa, Europe, the Middle East, central Asia, the Caucasus/Anatolia, and South America (even if for the reason of ensuring data quality) means that reaching firm conclusions, especially at a global scale, is extremely hard. For example, I am pretty sure that food gets spicier as one moves north in South America (and maybe Africa), but this is not captured in the study. As mentioned above, N70 dataset is strongly biased towards China, Japan and India, and although you eliminate this bias in the country-level data, these are not what the main discussion seems to be based on, and at country level the global coverage is poor. Regarding the point you made in response to another reviewer’s comment about higher population in Asia, whilst this is true, the Asian spread of the data is also not wide outside India, China and Japan, and the relevance of population is debatable – given that neighbouring and culturally-related cuisines are probably likely to be similar (which is why you took out the possible autocorrelation), a wide geographic and cultural spread is presumably more important than weighting by population?

I am aware this was a limitation of the sources upon which the study relied, and that similar biases were present in previous work, and this does not undermine the study overall because it is a “tenative step” towards better understanding, but please make this a *very* clear and strong caveat in the discussion and conclusions. It may also be leading to the rather odd correlation between GDP and spiciness which I refer to above. Future worthwhile work for new studies could actually be collecting more extensive and comparable cuisine data first hand/from recipe books from across the world, to allow more accurate statistical work to be undertaken – perhaps mention this as a future avenue.

> *R1.15. We agree this is an important point and deserving of more emphasis in this manuscript, particularly because our focus is on the difficulties of testing cultural evolution hypotheses using comparative data. Global sampling bias is well recognized as a serious source of bias in such studies, and we now give this more emphasis on page 11: see point R.13 above.*

Reviewer #2:
Remarks to the Author:

Authors did a great job revising the manuscript. The new title and the addition of Chilli in the analysis improve the paper. I have no further comments. Congratulations!!!

Reviewer #3:

Remarks to the Author:

Thank you for addressing my comments. The manuscript has greatly improved and I look forward to seeing this published.

Decision Letter, second revision:

22nd October 2020

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Professor Bromham,

Thank you once again for submitting your revised manuscript, entitled "Why do hot countries have spicy food? The challenges of testing hypotheses with cross-cultural data," and for your patience during the re-review process.

Your revised manuscript has now been evaluated by Reviewer 1, who is satisfied that their points have been adequately addressed. In the light of their advice (and the advice of the other reviewers in response to the previous round of revision) I am delighted to say that we can in principle offer to publish it. First, however, we would like you to revise your paper to ensure that it complies with our Guide to Authors at http://www.nature.com/nathumbehav/info/gta.

One of the main reasons for delays in formal acceptance is failure to fully comply with editorial policies and formatting requirements. To assist you with finalizing your manuscript for publication, I attach a checklist that lists all of our editorial policies and formatting requirements. I also attach a template document, which exemplifies our policies and formatting requirements.

Please attend to *every item* in the checklist and upload a copy of the completed checklist with your submission. I have highlighted in the checklist items that require your attention. I also mention here a few points that are frequently missed and can cause delays:

1) Ensure that all corresponding authors have linked their ORCID to their account on our online manuscript handling system. This is very frequently missed and invariably causes delays in formal acceptance.

2) Ensure that you provide all of the materials requested in the attached checklist and below with your final submission. Please note that the Licence to Publish needs to be hand-signed.

Nature Human Behaviour offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter ‘I wish to participate in transparent peer review’ if you want to opt in, or ‘I do not wish to participate in transparent peer review’ if you don’t. Failure to state your

preference will result in delays in accepting your manuscript for publication.
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

We hope to hear from you within 10 days; please let us know if the revision process is likely to take longer.

To submit your revised manuscript, you will need to provide the following:
• Cover letter
• Point-by-point response to the reviewers (if applicable)
• Manuscript text (not including the figures) in .docx or .tex format
• Individual figure files (one figure per file)
• Extended Data & Supplementary Information, as instructed
• Reporting summary
• Editorial policy checklist
• License to publish
• Third-party rights table (if applicable)
• Suggestions for cover illustrations (if desired)

Consortia authorship:
For papers containing one or more consortia, all members of the consortium who contributed to the paper must be listed in the paper (i.e., print/online PDF). If necessary, individual authors can be listed in both the main author list and as a member of a consortium listed at the end of the paper. When submitting your revised manuscript via the online submission system, the consortium name should be entered as an author, together with the contact details of a nominated consortium representative. See https://www.nature.com/authors/policies/authorship.html for our authorship policy and https://www.nature.com/documents/nr-consortia-formatting.pdf for further consortia formatting guidelines, which should be adhered to prior to acceptance.

Reviewer Recognition:
In recognition of the time and expertise our reviewers provide to Nature Human Behaviour's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Why do hot countries have spicy food? The challenges of testing hypotheses with cross-cultural data". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Please use the following link for uploading these materials:
[REDACTED]

If you have any further questions, please feel free to contact me.

With best regards,

Charlotte Payne
Editor
Nature Human Behaviour

Reviewer #1:
Remarks to the Author:
I would like to thank the authors for so thoroughly considering and addressing my points. I am now happy to see the study published. Although clearly time-consuming, I hope that the additions have made the work more robust to critique.

Final Decision Letter:

Dear Professor Bromham,

We are pleased to inform you that your Article "There is little evidence that spicy food in hot countries is an adaptation to reducing infection", has now been accepted for publication in Nature Human Behaviour.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

The subeditor may send you the edited text for your approval. Once your manuscript is typeset you will receive a link to your electronic proof via email within 20 working days, with a request to make any corrections within 48 hours. If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

The Author's Accepted Manuscript (the accepted version of the manuscript as submitted by the author) may only be posted 6 months after the paper is published, consistent with our self-archiving embargo. Please note that the Author's Accepted Manuscript may not be released under a Creative Commons license. For Nature Research Terms of Reuse of archived manuscripts please see: http://www.nature.com/authors/policies/license.html#terms
If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

We look forward to publishing your paper.

With best regards,

Charlotte Payne
Editor
Nature Human Behaviour