
Supplementary information

There is little evidence that spicy food in hot countries is an adaptation to reducing infection risk

In the format provided by the authors and unedited

There is little evidence that spicy food in hot countries is an adaptation to reducing infection

Lindell Bromham^{*1}, Alexander Skeels¹, Hilde Schneemann^{1,2}, Russell Dinnage¹, Xia Hua^{1,3}

¹ *Macroevolution & Macroecology, Research School of Biology, Australian National University, Canberra ACT 0200 AUSTRALIA.*

² *Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK*

³ *Mathematical Sciences Institute, Australian National University, Canberra ACT 0200 AUSTRALIA.*

* Corresponding author: Lindell Bromham, Lindell.Bromham@anu.edu.au

Supplementary Information

1. Data
 - 1.1. Cuisines
 - 1.1.1. Billing & Sherman data
 - 1.1.2. Expanded global data
 - 1.2. Relatedness between cuisines
 - 1.3. Spatial, environmental and biodiversity data
 - 1.4. Socioeconomic data
2. Details of Analysis
 - 2.1. Autocorrelation due to proximity and relatedness between cultures
 - 2.2. Calculation of marginal likelihood
3. Additional analyses & results
 - 3.1. Median spice
 - 3.2. Alternative datasets

Supplementary Table 1 Selection of literature on antimicrobial properties of spices

Supplementary Table 2 List of spice blends

Supplementary Data 1 Details of all statistical tests conducted on all datasets

Supplementary Data 2 Data analysed for this study (combined dataset)

Supplementary Data 3 Data analysed for this study (country-level dataset)

Supplementary Figure 1 Power of the GLS analysis on different datasets.

1. Data

1.1. Cuisines

1.1.1 *Billing & Sherman data*

Billing & Sherman (1998) reported a significant correlation between average spice per recipe and average temperature. These data were derived from 93 traditional cookbooks from 36 national level cuisines and four sub-national regions (northern and southern United States, northeast and southwest China)¹. For our analysis, these sub-national regions had to be assigned to specific areas in order to correct for proximity and derive environmental and diversity variables. For “Southern United States”, we included the states of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia, California, Arizona, New Mexico. “Northern United States” was all remaining states except Alaska and Hawaii. Similarly, we interpreted “Southwestern China” to include Chongqing, Sichuan, Yunnan, Guizhou and Xizang, and “Northeastern China” to include Liaoning, Jilin, Heilongjiang and Nei Mongol. Billing and Sherman (1998) provide average spice counts for these countries rather than the raw data on ingredients, so we are unable to derive variables such as number of ingredients, median spice, vinegar, alcohol or chilli.

1.1.2 *Expanded global dataset.*

To provide a more wide-ranging test of the adaptive cuisine hypothesis for spice use, we constructed a dataset that has a greater diversity of spices and a wider range of variables to analyse. This dataset is based on recipe data from four published sources²⁻⁵.

The CulinaryDB database² has recipe data grouped into a range of geographic regions. Some regions were too large to be useful for testing this hypothesis (e.g. South America, Africa, Eastern Europe), some regional cuisines included a large number of international recipes (e.g. Australia, Canada, US), and some had too few savoury recipes (e.g. Scandinavia), so we selected only those cuisines representing regions that could be considered comparable to the other regional units in our study. Most of these are national level, but we also include DACH (Germany+Australia+Switzerland) and Caribbean (which, for the purposes of estimating environmental and socioeconomic variables, we take to include Grenada, Dominica, Barbados, Bahamas, Saint Kitts and Nevis, Antigua and Barbuda, Dominican Republic, Jamaica, Cuba, Puerto Rico, Haiti, Trinidad and Tobago, Saint Lucia, St. Vincent and the Grenadines).

The Chinese regional recipe data was provided by Zhu et al (2013), who derived the ingredient lists from online recipe databases, resulting in a collection of 8,498 recipes with a total of 2,911 ingredients⁴. Each recipe was assigned to one of 20 regions. We analysed 19 separate regions: we did not include recipes labeled “Other” as this includes a number of widely separated areas such as Tibet and Taiwan. Following the map provided in Zhu et al. (2013), we assigned the “muslim” cuisine (referred to in the original dataset as qingzhencai) to the Sinkiang province, and refer to this in our database as Xinjiang. We translated the list of ingredients for the recipes using online searches on both Chinese characters and English translations. We then coded each ingredient into categories, including spice, chilli, vinegar, alcohol and meat (we did not use the original spice coding published by Zhu et al. as it was inconsistent and did not separately count the spices in spice blends and sauces). While Zhu et al. (2013) counted many medicinal ingredients as spices, we only include those that have anti-spoilage or anti-microbial effects on food preparation (see Table 2 and Methods for details).

The Japanese regional cuisine data was provided by Ohtsubo (2009) from a study designed to test the adaptive spice hypothesis using recipes from two sources: traditional recipes from elderly Japanese housewives, and cookbooks by experts in traditional cuisine⁵. We could not use the second dataset from cookbooks because these recipes are not identified to a particular regional cuisine. For the Japanese regional cuisines, we used the prefectures listed in the raw data⁵ to assign recipes to eight larger areas in order to have a sufficient sample of recipes per area. Our assignment of prefectures to areas is as follows: Tōhoku (Aomori, Iwate, Miyagi, Akita, Yamagata, Fukushima), Kantō (Ibaraki, Tochigi, Gunma, Saitama, Chiba, Tōkyō, Kanagawa), Chūbu (Niigata, Toyama, Ishikawa, Fukui, Yamanashi, Nagano, Gifu, Shizuoka, Aichi), Kansai (Mie, Shiga, Kyōto, Ōsaka, Hyōgo, Nara, Wakayama), Chūgoku (Tottori, Shimane, Okayama, Hiroshima, Yamaguchi), Shikoku (Tokushima, Kagawa, Ehime, Kōchi), Kyūshū (Fukuoka, Saga, Nagasaki, Kumamoto, Ōita, Miyazaki, Kagoshima). Hokkaido prefecture was counted as a single area. We exclude Okinawa prefecture because it is geographically separate from other prefectures (over 600 km south of the mainland of Japan) so could not be meaningfully combined with other prefectures, and it has too few recipes recorded to be included on its own. The Japanese dataset does record a temperature for each recipe by using thirty years of meteorological records from the closest observation point to the region where each interview was conducted⁵. However, because this detailed temperature data is only available for the eight Japanese cuisines, we

use a consistent procedure to calculate mean average temperature and other environmental variables for each of the countries and regions in our dataset.

The Indian regional cuisine data is derived from a study of food flavour pairing by Jain et al (2015) using data from 156 recipes from an online recipe repository³. To obtain environmental and spatial variables, we assigned cuisines to states most commonly associated with those cultures. Gudjurati, Maharashtrian, Punjabi and Rajasthani cuisines were each assigned to the Indian state of the same name. For Bengali cuisine, we combined the areas of three Indian states: West Bengal, Tripura and southern Assam. For the South Indian cuisine, we included the five major southern states of India: Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Telangana. Mughlai cuisine is associated with northern India so we include the states of Jammu and Kashmir, Haryana, Himachal Pradesh, Uttar Pradesh and Uttarakhand. Jain cuisine is not easily assigned to a modern state but the Jain population is concentrated around Maharashtra, Rajasthan and Gudjurat, so we use those states to derive regional data for the Jain cuisine. Note that spatial overlap between cuisines is permissible, as we are using the location both to correct for spatial distance between cultures but also to characterise the environment that the cuisine is found in.

1.2 Relatedness between cultures

To represent similarity between cuisines due to relatedness, we first identified the principal language of each country or region as given in Ethnologue⁶. We then used the hierarchical relationships between languages in both the Ethnologue⁶ and Glottolog⁷ to construct a hierarchical arrangement of these cultures to use for phylogenetic correction. In some cases, we had to make a decision which of several principle languages to use. In these cases, we were guided by the number of speakers of the language within the country or region, geographic extent of the language within that country or region, and the source of the recipe data. For South Indian regions, all four of the languages with the largest number of speakers (Telugu, Tamil, Kannada, Malayalam) are Dravidian, so can all be represented at the same place in the tree. We assumed that the “muslim” cuisine identified by Zhu et al. (2013) to the Sinkiang province⁴ is associated with the Uyghur language. Since Jain religious writings are commonly associated with Sanskrit and Prakrit, we have used Sanskrit to place the Jain cuisine in the tree. For South Africa in the B&S dataset¹, the cookbooks from which the recipes were derived are written by Afrikaans-speaking authors⁸, so we have used the Afrikaans language to represent this cuisine.

To create the trees, we consulted the hierarchical taxonomy of languages in the 2016 releases of both Ethnologue⁶ and Glottolog⁷. Where the two taxonomies were not identical, we accepted the taxonomy with greater resolution as long as it did not conflict with the other source. For example, Ethnologue placed Danish, Swedish and Norwegian as a polytomy within East Scandinavian, but Glottolog grouped Danish and Swedish together in East Scandinavian and Norwegian in West Scandinavian. Therefore in our tree we grouped Danish and Swedish together as the more resolved grouping that is compatible with both classifications. For the relationships between regional dialects of both Japanese and Chinese Mandarin, we followed Glottolog, which provides a finer level of resolution than Ethnologue. For example, while the Chinese regions have regional dialects, these were all classified by Ethnologue as Mandarin without additional hierarchical structure, but Glottolog grouped them into Northern, Central and Southern dialects. Similarly, the Japanese regions included in this study are all classified as Japanese speaking in Ethnologue, but many dialects are classified into Eastern and Western groups in Glottolog. The Hokkaido prefecture includes both Japanese and the language isolate Ainu. We do not know whether the recipes from the Hokkaido region are Ainu or Japanese, but as they are described as Japanese recipes we have included Hokkaido in the Eastern Japanese group.

For the Billings and Sherman dataset we use the following tree: (((Lebanon, Morocco), Israel), ((Malaysia, Indonesia), Philippines), (((South_Africa, (Germany, Austria)), (Australia, United_Kingdom, North_US, South_US)), (Norway, Denmark, Sweden)), (India, Iran), (((Brazil, Portugal), Mexico), France), Italy), Greece, Poland, Ireland), ((Kenya, Nigeria), Ghana), (Hungary, Finland), (North_China, South_China), Japan, South_Korea, Vietnam, Thailand, Ethiopia)

For the global dataset, including CulinaryDB and the regional cuisines for India, China and Japan (see Table 1), we used this tree: (((DACH_Countries, United_Kingdom),(((Gujarat ,Rajasthan) , Mughlai), Punjab, Bengal, Maharashtra, Jain), (((Mexico, Spain), France), Italy), Greece), (Min, (Xiang, (Hu, Su, Zhe), Jiangxi, ((Chuan, Dongbei, Hubei, Hui, Jing, Lu, Xibei, Yu, Yungui), Shanxi), (Yue, Hong Kong))), ((Tohoku, Kanto, Chubu, Hokkaido), (Kansai, Chugoku, Shikoku), Kyushu), South_Korea, Thailand, Xinjiang, Caribbean, South_India);

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.

We use a previously published method to scale branch lengths, setting a maximum clade height relative to the root and dividing it by the number of language classification levels⁹. We scaled the branch heights according to the method given in Bromham et al 2018⁹, using the relative age of language families to set an arbitrary node height for each clade (C). The purpose of this procedure is to provide an approximate depth of diversification for the phylogenetic analysis, to allow for the fact that distantly-related cultures will have had more time to accumulate differences than more recently diverged cultures. Family depths were derived from a range of published sources¹⁰⁻¹⁷, and used to scale the height of each clade relative to the root. We stress that the absolute ages of language families are immaterial for branch length scaling, only the relative depths of divergence are informative for this analysis.

Then for each clade, we counted the number of levels in the language classification between the root and the tips. We then used this to divide the clade height into arbitrary units representing depth of divergence within the clade. Then we scaled the node height within each language group by the classification level of tips, using the formula

$$(L_c - L_s) * C$$

where L_c is the maximum classification level in the clade, L_s is the classification level containing the sampled languages, and C is the clade height relative to the root. Again, this is not a perfect measure of degree of cultural relatedness, but it is better than either not using any relationship information or using unscaled branch lengths.

1.3. Spatial, environmental and biodiversity data

We retrieved high resolution climatic data from WorldClim¹⁸ and rescaled it to equal area projection with 10x10km grid cells. We calculated the mean value of mean annual temperature (Bio01) of grid cells whose centroid falls within the spatial polygon for each region. We estimate species richness for cuisine areas based on species distribution data mapped to hexagonal polygons of roughly 7800km². The figure reported is the average

number of species for all such hexagons whose centroid falls within the cuisine area, as defined above. Plant diversity and crop diversity was derived from Ellis et al (2012)¹⁹ which is provided in hexagonal polygons of roughly 7800km²; crop diversity is originally from Monfreda et al. (2008)²⁰; plant diversity from Kreft and Jetz (2007)²¹. We also include language diversity as a measure of human cultural diversity, counting the number of L1 spoken languages whose distribution occurs partially or wholly within each cuisine area^{9,22,23}. We use bird and mammal species richness as a proxy to represent variation in biodiversity, including parasite diversity, since human pathogens correlate strongly with bird and mammal species richness²⁴. Bird and mammal species richness data was obtained from [Biodiversitymapping.org](https://www.biodiversitymapping.org) in equal area grid cells of 10x10km resolution, using data based derived from BirdLife International and NatureServe databases. We took the mean richness of grid cell values falling within each regional polygon

For each spice in the dataset (Table 2) we extracted occurrence records for these species from the Global Biodiversity Information Facility (GBIF.org 2020, <https://doi.org/10.15468/dl.dv5hu4>). Occurrence records were then cleaned for common errors and data quality issues using the CoordinateCleaner package in R²⁵. To reduce the effects of outlier points and variation in sampling intensity between spice species we converted the occurrence points into a continuous spatial density on the globe using spherical kernel density estimation. For this we used the spherical_kde Python package (https://github.com/williamjameshandley/spherical_kde), which was accessed from R using the reticulate R package²⁶. Using the spherical KDE fit to the occurrence points, we predicted the density of points for all spice species across a global 1000 by 500 grid. To account for uneven sampling we then normalized the predicted densities to fall between 0 and 1. Then, for each cuisine polygon in our analysis we calculated the average density of grid points falling within its boundary, for every spice species. Finally, we estimated the diversity of spice ingredients growing in the cuisine area using an abundance weighted metric, Simpson's diversity index²⁷, where we weighted by the average density of each spice rather than abundance.

1.4 Socioeconomic data

Population density was obtained from Ellis et al. (2012)¹⁹ and based on the population model of Klein Goldewijk et al. (2010)²⁸. For each cuisine area, mean population density (inhabitants per square kilometre) was extracted from all hexagons whose centroid fell within the polygon of the cuisine area. Life expectancy is based on most recent data for each

region or country. Country-level data is from the World Health Organisation. Chinese regional data on life expectancy is based on National Bureau of Statistics of the People's Republic of China using the map in Zhu et al. (2009) to identify regions associated with cuisines. Data on life expectancy for Indian states is from the United Nations Development Programme²⁹. Life expectancy for Japanese prefectures is from Ministry of Health, Labour and Welfare of Japan (www.mhlw.go.jp). For US states, we used 2017 figures from Institute for Health Metrics and Evaluation. For super-national regions, such as Caribbean or DACH, we average over values for each country. GDP per capita in US dollars for countries is from World Bank figures for 2017-2018 (data.worldbank.org); for Indian states is from Ministry of Statistics and Programme Implementation; for Chinese regions from National Bureau of Statistics of China; for Japanese prefectures from OECD.

To reflect food availability and diet quality, we used data on child malnutrition, represented by the proportion of children aged five and under with stunted growth³⁰. We used the gridded 5x5km data, averaged over the cuisine area. This dataset is for mid- to low-income countries, so for the high-income countries not included in this dataset we assign the lowest recorded value which is for Hong Kong. This is consistent with data from WHO which suggests that child stunting is typically 1% or less for high-income countries (<https://data.worldbank.org/indicator/SH.STA.STNT.ZS>), which is similar to levels recorded in Hong Kong³¹.

To represent non-disease-based causes of death, we recorded road traffic fatalities per 100,000 population from WHO Global Health Observatory data repository, using the 2013 figures (because they had the greatest coverage at country level). For US states, 2017 data from U.S. Department of Transportation's Fatality Analysis Reporting System (via www.iihs.org); Chinese regions from the National Bureau of Statistics 2017; Japanese prefectures from Ministry of Health, Labour and Welfare 2016 data. While state-based estimates of road traffic deaths are available for Indian states from the Ministry of Road Transport and Highways³², these variables are calculated in a different way from the WHO statistics, with the estimates much lower in all states than the national average reported for India in the WHO data (for both the 2008 and 2018 figures). Therefore we deemed that the Indian state data was not directly comparable to the other estimates of road traffic deaths, and only use the country-level estimate for India from the WHO in the country-level dataset.

To reflect the possible effect of colonization on socioeconomic measures such as GDP, we record for each cuisine whether it is in a country that has had a substantial period as a colony. Levels of colonial occupation and impact vary widely, but we recorded a binary variable as 1 for those countries that had been under a foreign colonial power (e.g. excluding neighbouring countries expansion of influence or disputed borders) for a period of decades (e.g. excluding occupation during world wars) within the past four centuries (e.g. excluding ancient empires such as Roman and Phoenician).

For some of these socioeconomic datasets, the government database is not available for open searches, but the data has been made available on public platforms such as Wikipedia or StatisticsTimes.com, with fully referenced sources. We use this republished data for regional values for Life expectancy for China, USA, and Japan, and for regional values of GDP per capita for China, India and Japan. Road traffic deaths for Japan were accessed via stats-japan.com.

2. Details of Analysis

2.1. Autocorrelation due to proximity and relatedness between cultures

We accounted for autocorrelation by constraining the residual correlation in spice use between each pair of cuisines to be a linear function of the spatial proximity and phylogenetic similarity between the two cuisines. Using the GLS analysis developed by Hua et al. (2019), the correlation matrix has the form:

$$(1 - a_1)I + a_1[a_2P + (1 - a_2)D]$$

, where I is an identity matrix, P is the phylogenetic similarity matrix, and D is the spatial proximity matrix, a_1 represents the relative contribution of spatial and phylogenetic versus other residual effects, a_2 represents the relative contribution of spatial versus phylogenetic effects. To calculate the spatial proximity matrix, we modeled the decay in spatial autocorrelation of cuisines with distance as the Gaussian function $e^{-(d/a_3)^2}$, where d is the great-circle distance between the between centroids of two cuisine polygons, calculated using functions *gCentroid* and *earth.dist* from the R packages *rgeos* and *fossil* respectively^{33,34}; a_3 is the coefficient describing how fast similarity decays over the distance between grid cells and is estimated by maximum likelihood approach.

2.2. Calculation of marginal likelihood

Our regression model takes the form $y = bX + \varepsilon$, where ε follows a multivariate normal distribution with mean 0 and covariance $\sigma^2 C$, where C is the best fit correlation structure of residuals in y . We used multivariate normal inverse gamma distribution as the prior for $(b, \sigma^2) \sim MNIG(\mu, \lambda, \alpha, \beta)$, which is conjugate to multivariate normal distribution, so the posterior distribution of b and σ^2 also follows a multivariate normal inverse gamma distribution:

$$(b, \sigma^2 | y, X) \sim MNIG \left(\begin{array}{l} \mu' = (\lambda')^{-1}(\lambda I_p \mu + X^T C^{-1} y) \\ \lambda' = \lambda I_p + X^T C^{-1} X \\ \alpha' = \alpha + \frac{n}{2} \\ \beta' = \beta + \frac{1}{2}(\mu^T \lambda I_p \mu + y^T C^{-1} y - \mu'^T \lambda' \mu') \end{array} \right).$$

The marginal likelihood of the data is then the ratio of the normalizing constant between the prior and the posterior distributions, which equals:

$$\frac{|C|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} |\lambda I_p|^{-\frac{1}{2}} \frac{\beta^\alpha}{\Gamma(\alpha)}}{|\lambda'|^{-\frac{1}{2}} \frac{\beta'^{\alpha'}}{\Gamma(\alpha')}}.$$

In the prior distribution, α, β define an inverse gamma distribution for σ^2 , so we set $\alpha = 0.1$ to give a rather flat distribution, and set β to make the mode of the distribution equal to the MLE of σ^2 ; μ, λ define a multivariate normal distribution for b , with mean μ and variance σ^2/λ , so we set $\mu = 0$, and $\lambda = 1$.

3. Additional analyses and results

3.1. Median spice

In the main text, we use average number of spice ingredients per recipe to measure spice use, following Billing & Sherman (1998)¹. Here we repeat the analyses using median number of spice ingredients per recipe to reduce the influence of cuisines that contain some complex recipes with very large numbers of spices. These analyses can only be conducted for the cuisines for which we have access to the raw data. The results are largely consistent with the results reported in the paper. Cuisines with higher average temperature have higher median number of spices per recipe for our dataset ($t_{43}=3.67, p<0.001$), but this association is nonsignificant after accounting for spatial and phylogenetic autocorrelations ($\beta=0.12, t_{43}=1.02, p=0.315$). The data supported the absence of the relationship with a Bayes Factor of 5.2. But there is less variation in median spice use and so lower power to detect

associations than mean spice use, with sufficient power (>0.8) to detect any effect with $\beta \geq 0.2$ (Supplementary Figure 1).

The major results that are different from mean spice use is that median spice use is significantly associated with historical pathogen load ($\beta=0.43$, $t_{43}=2.52$, $p=0.015$, Bayes factor against the null result (BF) = 1.4), but not significantly associated with incidence of foodborne illness after accounting for spatial and phylogenetic autocorrelations ($\beta=0.24$, $t_{43}=1.78$, $p=0.082$, Bayes factor for the null result (BF_{null}) = 1.6). As indicators of poverty, road traffic deaths (BF=12.2) is a significantly better predictor of median spice than foodborne illness, and GDPpc (BF=1.8) predicts median spice as well as foodborne illness.

The data provide evidence that the association between median spice use and poverty could be explained by low food availability and shared latitude gradients: adding GDPpc to a model with malnutrition does not increase model fit to median spice use (likelihood ratio = 1.58, $p=0.209$; Bayes factor=0.57), neither does adding GDPpc to a model with latitude (likelihood ratio = 2.35, $p=0.125$; Bayes factor=0.4). But we find no evidence that the association between spice use and poverty is influenced by the historical patterns of colonization. In fact, the association between spice use and GDPpc is stronger in cuisines that have not been under colonial control ($\beta=-0.69$, $t_{31}=-4.71$, $p<0.001$, BF=13.3).

Median spice use is weakly correlated with average life expectancy ($\beta=-0.24$, $t_{43}=-2.14$, $p=0.038$, Bayes factor=0.6). Adding foodborne illness does not significantly increase the fit of a model of either GDPpc, or road traffic deaths, or average life expectancy predicting median spice use (Supplementary Data 1), suggesting that foodborne illness has little additional explanatory power for spice use above its association with poverty and poor health outcomes in general.

The data also provides no evidence for other explanatory paths linking spice use and foodborne illness. Median spice use is not significantly linked to any biocultural diversity (Supplementary Data 1). Within all the climatic variables, median spice use is only significantly associated with precipitation seasonality ($\beta=0.43$, $t_{43}=2.52$, $p=0.015$), and the link is weakly supported as suggested by Bayes factor of 1.4 (Supplementary Data 1). Unlike indicators of poverty, adding precipitation seasonality does not significantly increase the fit of a model of foodborne illness predicting median spice use (likelihood ratio = 2.72, $p=0.099$; Bayes factor=0.6).

3.2. Alternative datasets

In the main text we report results for the “Total” dataset (N=45, See Table 1), that includes only cuisines for which we were able to estimate spice counts directly. Here we repeated all analyses on the Combined dataset (N=70). This dataset has the highest statistical power, with sufficient power (>0.8) to detect any effect with $\beta \geq 0.125$ (Figure 3). Despite the higher power, the evidence is still weak for the link between spice use and foodborne illness after accounting for spatial and phylogenetic autocorrelations ($\beta=0.23$, $t_{68}=2.27$, $p=0.027$, Bayes factor=0.9). So is the link between spice use and child diarrhea ($\beta=0.41$, $t_{40}=2.20$, $p=0.034$, Bayes factor=1.4). In this dataset, there is no significant support for a link between spice use and GDPpc ($\beta=0.23$, $t_{68}=-1.82$, $p=0.073$, Bayes factor=0.5), as well as the link between spice use and average life expectancy ($\beta=-0.09$, $t_{68}=-0.92$, $p=0.361$, Bayes factor=0.2). Road traffic deaths predict mean spice as well as foodborne illness (BF=1.0). However, the association between spice use and GDPpc becomes significant after removing Spice Trade cuisines ($\beta=-0.23$, $t_{40}=-2.28$, $p=0.026$, BF=1.6). Spice use is also significantly associated with GDPpc in cuisines that had not been under colonial control ($\beta=-0.40$, $t_{40}=-3.50$, $p=0.001$, BF=24.5).

All other results are qualitatively the same as the results for mean and median spice use from N45 “Total” dataset (Supplementary Data 1). 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 (Supplementary Data 1), except that the evidence is stronger for link between spice use and foodborne illness ($\beta=0.27$, $t_{68}=2.62$, $p=0.011$, Bayes factor=2.1).

We also performed the same tests for the adaptive cuisine hypothesis on the original Billing & Sherman dataset¹ (N=38) and a country-level dataset (N=39). For the country level dataset, we exclude any supranational regions (e.g. Carribean, DACH), and for countries that have sub-national cuisines, we produce one value for each country by a weighted average over all the cuisines they contain. Due to smaller data size, both of these datasets have relatively low statistical power with sufficient power (>0.8) to detect effect with $\beta \geq 0.3$ (Figure 3). Yet, both datasets give qualitatively the same results as other datasets analysed. Mean spice use is significantly associated with temperature and historical pathogen load before accounting for spatial and phylogenetic autocorrelations in Billing & Sherman dataset and in the country-level dataset (Supplementary Data 1). But after accounting for

these autocorrelations, mean spice use is not significantly associated with either temperature ($\beta=0.34$, $t_{36}=1.68$, $p=0.102$, Bayes factor for the null result ($BF_{\text{null}}=1.8$) or historical pathogen load ($\beta=0.07$, $t_{36}=0.31$, $p=0.759$, $BF_{\text{null}}=4.8$) in Billing & Sherman dataset, nor in the country-level dataset (temperature: $\beta=0.29$, $t_{37}=1.37$, $p=0.178$, $BF_{\text{null}}=2.0$; historical pathogen load: $\beta=0.04$, $t_{37}=0.24$, $p=0.814$, $BF_{\text{null}}=5.2$). Both datasets also provide very weak evidence for the link between spice use and risk of foodborne infection after accounting for spatial and phylogenetic autocorrelations. In Billing & Sherman dataset, mean spice use is not significantly associated with foodborne illness ($\beta=0.24$, $t_{36}=1.93$, $p=0.061$, $BF_{\text{null}}=1.7$). In country-level dataset, mean spice use is significantly associated with foodborne illness ($\beta=0.28$, $t_{36}=2.23$, $p=0.032$), but this link is weakly supported by Bayes factor 1.1. Mean spice use is also not related to child diarrhea ($\beta=0.22$, $t_{16}=0.84$, $p=0.413$, $BF_{\text{null}}=3.2$).

Supplementary Table 1 Selection of literature on antimicrobial properties of spices.

Spices in Table 1 include all of those identified as a spice by Billing & Sherman (1998) plus additional spices present in recipes as defined in the Methods. If in doubt as to whether an ingredient should be considered a spice for the purpose of testing the adaptive cuisine hypothesis, we sought published studies showing antimicrobial activity. Examples of published studies are included in this table, but it is not an exhaustive list.

Spice	Example Reference
Anjelica	Fraternale, D., G. Flamini, and D. Ricci. 2014. Essential oil composition and antimicrobial activity of <i>Angelica archangelica</i> L. (Apiaceae) roots. <i>J. Med. Food</i> 17:1043–1047.
Ajwain (Carom)	Hassanshahian, M., Z. Bayat, S. Saeidi, and Y. Shiri. 2014. Antimicrobial activity of <i>Trachyspermum ammi</i> essential oil against human bacterial. <i>Int. J. Adv. Biol. Biomed. Res.</i> 2:18–24.
Asafoetida	Iranshahy, M. & Iranshahi, M. Traditional uses, phytochemistry and pharmacology of asafoetida (<i>Ferula assa-foetida</i> oleo-gum-resin)—A review. <i>Journal of ethnopharmacology</i> 134, 1-10 (2011).
Chamomile	Miraj, S., and S. Alesaeidi. 2016. A systematic review study of therapeutic effects of <i>Matricaria recuitta chamomile</i> (chamomile). <i>Electron. physician</i> 8:3024–3031.
Chervil	Cho, E. J., J. M. Choi, H. M. Kim, K. Choi, J. Ku, K. W. Park, J. Kim, and S. Lee. 2013. Antibacterial activity and protective effect against gastric cancer by <i>Anthriscus sylvestris</i> fractions. <i>Hortic. Environ. Biotechnol.</i> 54:326–330.
Chrysanthemum sp.	Shunying, Z., Y. Yang, Y. Huaidong, Y. Yue, and Z. Guolin. 2005. Chemical composition and antimicrobial activity of the essential oils of <i>Chrysanthemum indicum</i> . <i>J. Ethnopharmacol.</i> 96:151–158.
Curry leaf	Ali, S., H. Al Harbi, and U. M. Irfan. 2016. The Antibacterial Effect Of Curry Leaves (<i>Murraya koenigii</i>) <i>European Journal of Pharmaceutical and Medical Research</i> 3(10): 382-387.
Drumstick leaves	Singh, K., and G. M. Tafida. 2014. Antibacterial activity of <i>Moringa oleifera</i> (Lam) leaves extracts against some selected bacteria. <i>Int. J. Pharm. Pharm. Sci.</i> 6:52–54
Epazote	Ravindran, P. N., G. S. Pillai, and M. Divakaran. 2012. 28 - Other herbs and spices: mango ginger to wasabi. Pp. 557–582 in K. V. B. T.-H. of H. and S. (Second E. Peter, ed. <i>Woodhead Publishing Series in Food Science, Technology and Nutrition</i> . Woodhead Publishing.
Fuki	Kim S-J, Cho AR, Han J (2013) Antioxidant and antimicrobial activities of leafy green vegetable extracts and their applications to meat product preservation. <i>Food Control</i> 29(1):112-120
Garcinia indica	Lakshmi, C., K. A. Kumar, T. J. Dennis, and T. S. S. P. N. S. Sanath Kumar. 2011. Antibacterial activity of polyphenols of garcinia indica. <i>Indian J. Pharm. Sci.</i> 73:470–473.
Gardenia	Chaichana J, Niwatananum W, Vejabhikul S, Somna S, Chansakaow S (2009) Volatile constituents and biological activities of <i>Gardenia jasminoides</i> . <i>J Health Res</i> 23(3):141-145
Hyssop	Kizil, S., N. Haşimi, V. Tolan, E. Kiliç, and H. Karataş. 2010. Chemical composition, antimicrobial and antioxidant activities of hyssop (<i>Hyssopus officinalis</i> L.) essential oil. <i>Not. Bot. Horti Agrobot. Cluj-Napoca</i> 38:99–103.
myoga (Japanese ginger)	Abe M, Ozawa Y, Uda Y, Yamada F, Morimitsu Y, Nakamura Y, Osawa T (2004) Antimicrobial activities of diterpene dialdehydes, constituents from myoga (<i>Zingiber mioga</i> Roscoe), and their quantitative analysis. <i>Bioscience, Biotechnology, and Biochemistry</i> 68(7):1601-1604

Japanese parsley	Lu J, Fu X, Liu T, Zheng Y, Chen J, Luo F (2018) Phenolic composition, antioxidant, antibacterial and anti-inflammatory activities of leaf and stem extracts from <i>Cryptotaenia japonica</i> Hassk. <i>Industrial Crops and Products</i> 122:522-532
Lemon balm	Abdellatif, F., H. Boudjella, A. Zitouni, and A. Hassani. 2014. Chemical composition and antimicrobial activity of the essential oil from leaves of Algerian <i>Melissa officinalis</i> L. <i>EXCLI J.</i> 13:772-781.
Lemon verbena	Koohsari, H., E. A. Ghaemi, M. S. S. Poli, and A. Sadegh. 2013. Evaluation of antibacterial activity of Lemon verbena (<i>Lippia citriodora</i>) leaves. <i>Ann. Biol. Res.</i> 4:52-55.
Lovage	Shafaghat A. Chemical constituents, antimicrobial and antioxidant activity of the hexane extract from root and seed of <i>Levisticum persicum</i> Freyn and Bornm. <i>Journal of Medicinal Plants Research.</i> 2011 Sep 30;5(20):5127-31.
Mastic Gum	Koutsoudaki, C., M. Krsek, and A. Rodger. 2005. Chemical composition and antibacterial activity of the essential oil and the gum of <i>Pistacia lentiscus</i> Var. chia. <i>J. Agric. Food Chem.</i> 53:7681-7685.
Mate	Prado Martin, J. G., E. Porto, S. M. de Alencar, E. M. da Glória, C. B. Corrêa, and I. S. Ribeiro Cabral. 2013. Antimicrobial activity of yerba mate (<i>Ilex paraguariensis</i> St. Hil.) against food pathogens. <i>Rev. Argent. Microbiol.</i> 45:93-98.
Mexican oregano	Arana-Sánchez, A., M. Estarrón-Espinosa, E. N. Obledo-Vázquez, E. Padilla-Camberos, R. Silva-Vázquez, and E. Lugo-Cervantes. 2010. Antimicrobial and antioxidant activities of Mexican oregano essential oils (<i>Lippia graveolens</i> H. B. K.) with different composition when microencapsulated in β -cyclodextrin. <i>Lett. Appl. Microbiol.</i> 50:585-590.
Nigella	Hanafy, M. S. M., and M. E. Hatem. 1991. Studies on the antimicrobial activity of <i>Nigella sativa</i> seed (black cumin). <i>J. Ethnopharmacol.</i> 34:275-278.
Peppermint	Singh, R., M. A. M. Shushni, and A. Belkheir. 2015. Antibacterial and antioxidant activities of <i>Mentha piperita</i> L. <i>Arab. J. Chem.</i> 8:322-328.
Perilla	Kang, R., R. Helms, M. J. Stout, H. Jaber, Z. Chen, and T. Nakatsu. 1992. Antimicrobial activity of the volatile constituents of <i>Perilla frutescens</i> and its synergistic effects with polygodial. <i>J. Agric. Food Chem.</i> 40:2328-2330.
Rose	Hirulkar, N. 2011. Antimicrobial activity of rose petals extract against some pathogenic bacteria. <i>Int. J. Pharm. Biol.</i> 1:478-484.
Saffron	Vahidi, H., M. Kamalinejad, and N. Sedaghati. 2002. Antimicrobial Properties of <i>Crocus sativus</i> L. <i>Iran. J. Pharm. Res.</i> 1:33-35.
Sandalwood	Jirovetz, L., G. Buchbauer, Z. Denkova, A. Stoyanova, I. Murgov, V. Gearon, S. Birkbeck, E. Schmidt, and M. Geissler. 2006. Comparative study on the antimicrobial activities of different sandalwood essential oils of various origin. <i>Flavour Fragr. J.</i> 21:465-468.
Toon	Wu, J.-G., W. Peng, J. Yi, Y.-B. Wu, T.-Q. Chen, K.-H. Wong, and J.-Z. Wu. 2014. Chemical composition, antimicrobial activity against <i>Staphylococcus aureus</i> and a pro-apoptotic effect in SGC-7901 of the essential oil from <i>Toona sinensis</i> (A. Juss.) Roem. leaves. <i>J. Ethnopharmacol.</i> 154:198-205
Vetiver	Vázquez-Sánchez, D., M. L. Cabo, and J. J. Rodríguez-Herrera. 2015. Antimicrobial activity of essential oils against <i>Staphylococcus aureus</i> biofilms. <i>Food Sci. Technol. Int.</i> 21:559-570.

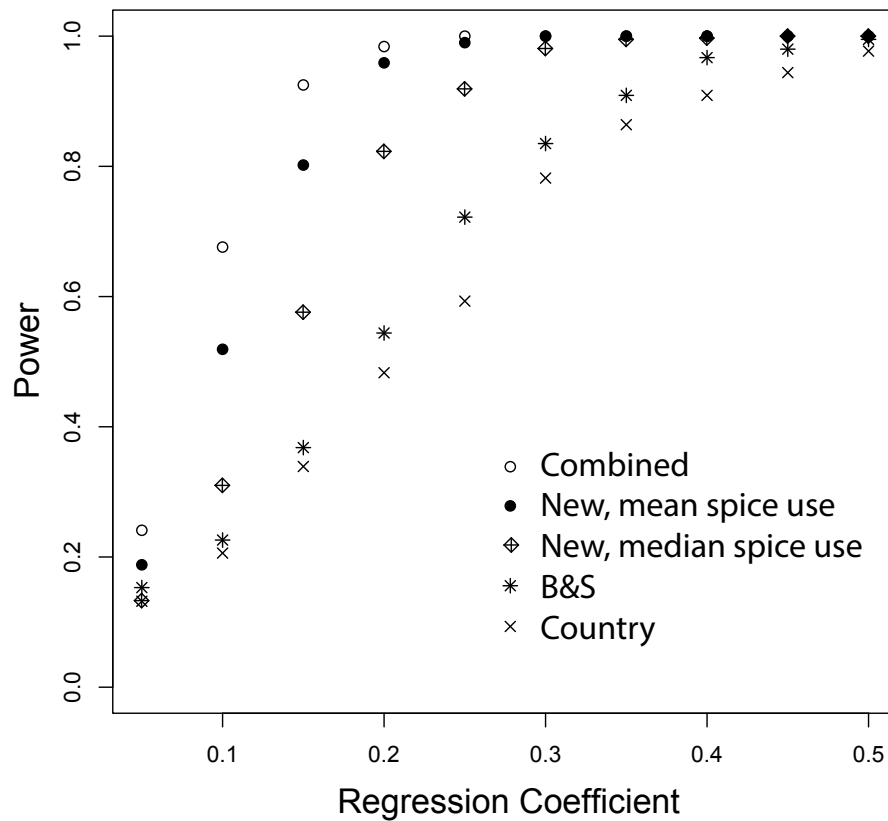
Supplementary Table 2: List of spice blends. Many recipes list a spice blend rather than separate spices, so we needed to count all the spices that would typically be included in that blend. This is an approximate list of spice blend ingredients used for the purposes of counting spices per recipe, but note that not all spice blends may correspond exactly to this list of ingredients. However, listing the likely ingredients is preferable to failing to count any of the spices likely to be included in a blend.

Region	Spice blend	Number	Commonly included spices
India	chaat masala	6	cumin, coriander, ginger, pepper, asafoetida, chilli, amchoor
	chicken masala powder	5	coriander, fenugreek, fennel, cumin, pepper
	chole masala	8	bay, mustard, cumin, coriander, cloves, chilli, pepper, pomegranate
	curry powder	8	coriander, cumin, tumeric, ginger, mustard, fenugreek, pepper, chilli
	dabeli masala	5	chilli, coriander, cinnamon, cloves, cumin
	garam masala	8	coriander, cumin, cardamom, pepper, fennel, mustard, cloves, chili
	goda masala	7	cumin, caraway, cardamom, cloves, pepper, bay, coriander
	jal jeera powder	7	cumin, citrus, mint, pepper, ginger, asafoetida, tamarind
	kebab masala	8	bay, ginger, cinnamon, cloves, pepper, coriander, fennel, chilli
	kitchen king masala	17	cumin, caraway, chili, turmeric, coriander, ginger, pepper, cloves, mustard, cardamom, fenugreek, fennel, cinnamon, star anise, nutmeg, poppy
	madras curry powder	8	coriander, cumin, cinnamon, chilli, tumeric, cardamom, pepper, cloves, pepper, fenugreek
	panch phoron	5	cumin, mustard, fenugreek, nigella, fennel
	pulao masala	6	pepper, cloves, cumin, cinnamon, cardamom, coriander
	rasam powder	5	pepper, coriander, cumin, chilli, asafoetida
	sambar powder	7	coriander, cumin, mustard, pepper, chilli, fenugreek, cinnamon
	tandoori masala	11	coriander, cumin, cardamom, pepper, fennel, mustard, cloves, chilli, garlic, ginger, onion
	Japan	Japanese seven spice	6
wasabi		2	japanese horseradish, mustard
China	bag of spices (小滷包)	6	cumin, cinnamon, licorice, bay, anise, tea
	BBQ sauce	2	garlic, pepper
	beef sauce (醬牛肉汁)	8	cloves, cardamom, cinnamon, anise, pepper, onion, ginger, garlic
	black pepper sauce (黑椒汁)	2	garlic, pepper
	cantonese-style duck wings (广式滷鴨翅滷包)	6	citrus, chilli, licorice, cloves, cumin, cinnamon
	cattle hoof yellow (牛蹄黃調料)	4	garlic, onion, ginger, chilli
	char sui sauce (叉燒醬)	5	cinnamon, cloves, fennel, star anise, sichuan pepper
	chicken sauce (雞汁)	2	garlic, chili
	Chilli garlic paste (辣椒蒜泥醬)	2	chilli, garlic
	Egret Island sauce (鷺島醬)	2	chilli, garlic
	garlic sauce (蒜辣醬)	2	chilli, garlic
	Ginger vinegar (姜醋汁)	2	ginger, galangal, vinegar
	hoisin sauce	2	chilli, garlic
	hot pot (火鍋調料)	5	sesame, chilli, onions, ginger, pepper, vinegar
	Hu spicy powder (胡辣粉)	4	coriander, onion, ginger, pepper

	kimchi	5	chilli, ginger, garlic, radish, onion
	Hot sauce (辣酱油)	3	tamarind, onions, garlic, vinegar
	korean hot sauce (韩式辣酱)	2	chilli, garlic
	mala sauce (麻辣酱)	9	sichuan pepper, chilli, cloves, garlic, star anise, cardamom, fennel, ginger, cinnamon
	milk soup (奶汤调料)	2	ginger, onions
	Miyamoto sauce (宫本酱)	2	chilli, sesame, ginger, vinegar
	Mother of Pork chilli sauce (干妈香辣酱)	3	chilli, pepper, sesame
	mushroom oil (菌油)	4	tea, ginger, pepper, star anise
	oil chilli sauce (油辣酱)	3	chilli, sesame, onion
	Oil crisp pepper (油酥辣椒调料)	4	garlic, ginger, sesame, chilli
	OK sauce (ok 酱)	2	citrus, garlic
	onion juice (洋葱汁)	3	onion, sesame, pepper
	pickle sauce (泡椒汁)	2	chilli, garlic
	pickled vegetables (泡青菜香料包)	4	garlic, celery, pepper, aniseed
	plum sauce (梅酱)	2	garlic, chili
	Red drum oil (红鼓油)	4	sichuan pepper, star anise, cinnamon, licorice
	red soy sauce (红酱油)	4	pepper, cinnamon, licorice, star anise
	red tank sauce (红糟汁)	5	onion, ginger, pepper, cinnamon, cloves
	ribs sauce (排骨酱)	2	sesame, garlic
	Sesame oil chili sauce (香油辣酱)	2	sesame, chilli
	Shacha sauce (沙茶酱)	3	garlic, onions, chilli
	Sichuan red oil sauce (红油汁)	2	chilli, onion
	spicy flowers (香辣姜花酱)	5	chilli, pepper, onion, ginger, garlic
	spicy lobster sauce (香辣豆豉)	4	ginger, chilli, garlic, star anise
	spicy marinade (辣味卤汁)	7	cumin, pepper, licorice, cloves, cinnamon, chilli, ginger
	spicy paste (辣糊)	5	seaweed, coriander, ginger, mustard, pepper
	spicy sauce (香辣酱)	5	chilli, garlic, ginger, sesame, allspice
	Spicy sesame paste (辣芝麻酱)	2	chilli, sesame
	sweet and sour (酸甜酱)	2	chilli, garlic
	Szechuan Kung Po (宫保酱)	3	chilli, sesame, ginger
	thousand island sauce	4	citrus, paprika, mustard, chilli
	wing lobster sauce (永豆豉)	3	garlic, onion, chilli
	worcester sauce	3	tamarind, onions, garlic
	Yellow chilli sauce (黄辣酱)	2	capsicum, chilli
	XO sauce	2	chilli, garlic
Global	Old Bay Seasoning	11	bay, celery, mustard, pepper, ginger, nutmeg, clove, allspice, chilli, cardamom, cinnamon
	Herbes de Provence	11	rosemary, lavender, fennel, parsley, savory, oregano, thyme, tarragon, basil, bay, marjoram
	Cajun seasoning	7	cayenne, garlic, oregano, paprika, thyme, pepper, onion
	ketchup	7	Vinegar, onion, allspice, coriander, clove, cumin, garlic, mustard
	Creole seasoning	7	onion, pepper, garlic, oregano, chilli, basil, thyme

Stock	6	onion, celery, thyme, bay, pepper, garlic
poultry seasoning	6	sage, thyme, marjoram, rosemary, nutmeg, pepper
Chinese five spice	5	cinnamon, star anise, fennel, pepper, clove
Ranch Dressing	5	garlic, onion, parsley, pepper, dill
Italian seasoning	4	basil, oregano, rosemary, thyme
Vegetable broth	3	onion, celery, sage
BBQ sauce	3	vinegar, onion, mustard, pepper
Adobo	3	chilli, garlic, vinegar, cumin
Hot sauce	1	chilli, vinegar

Supplementary Figure 1: Power of the GLS analysis on different datasets as listed in Table 1. The power (y axis) indicates how often the GLS analysis will detect significant effect of an independent variable that has certain amount of effect (x axis) on spice use in each dataset. The effect is measured as the amount of change in standard deviations in spice use with a standard deviation change in the independent variable, which is comparable to the estimated regression coefficients in the GLS analysis.



Supplementary References

- 1 Billing, J. & Sherman, P. W. Antimicrobial functions of spices: why some like it hot. *The Quarterly review of biology* **73**, 3-49 (1998).
- 2 Bagler, G. CulinaryDB: <https://cosylab.iiitd.edu.in/culinarydb/>. *Indraprastha Institute of Information Technology Delhi (IIIT-Delhi)* (2017).
- 3 Jain, A., Rakhi, N. & Bagler, G. Analysis of food pairing in regional cuisines of India. *PloS one* **10**, e0139539 (2015).
- 4 Zhu, Y.-X. *et al.* Geography and similarity of regional cuisines in China. *PloS one* **8**, e79161 (2013).
- 5 Ohtsubo, Y. Adaptive ingredients against food spoilage in Japanese cuisine. *International Journal of Food Sciences and Nutrition* **60**, 677-687 (2009).
- 6 Lewis, M. P., Simons, G. F. & Fennig, C. D. (SIL International, Dallas, Texas, 2016).
- 7 Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. *Glottolog 2.7*, <<http://glottolog.org/>> (2016).
- 8 Billing, J. & Sherman, P. W. Antimicrobial functions of spices: why some like it hot. *Quarterly review of biology*, 3-49 (1998).
- 9 Bromham, L., Hua, X., Cardillo, M., Schneemann, H. & Greenhill, S. J. Parasites and politics: why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society Open Science* **5**, 181100 (2018).
- 10 Bouckaert, R. *et al.* Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957-960 (2012).
- 11 Grollemund, R. *et al.* Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* **112**, 13296-13301, doi:10.1073/pnas.1503793112 (2015).
- 12 Kitchen, A., Ehret, C., Assefa, S. & Mulligan, C. J. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences* **276**, 2703-2710 (2009).
- 13 Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479-483, doi:10.1126/science.1166858 (2009).
- 14 Honkola, T. *et al.* Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* **26**, 1244-1253 (2013).
- 15 Blench, R. & Post, M. W. Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages. *Trans-Himalayan Linguistics* **266**, 71-104 (2014).
- 16 Lee, S. & Hasegawa, T. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B: Biological Sciences*, doi:10.1098/rspb.2011.0518 (2011).
- 17 Holman, E. W. *et al.* Automated dating of the world's language families based on lexical similarity. *Current Anthropology* **52**, 000-000 (2011).
- 18 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 1965-1978 (2005).
- 19 Ellis, E. C., Antill, E. C. & Krefl, H. All is not loss: plant biodiversity in the Anthropocene. *PlosONE* **7**, e30535 (2012).

- 20 Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* **22**, GB1022. (2008).
- 21 Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**, 5925-5930, doi:10.1073/pnas.0608361104 (2007).
- 22 Murray, D. R. & Schaller, M. Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology* **41**, 99-108 (2010).
- 23 Hua, X., Greenhill, S. J., Cardillo, M., Schneemann, H. & Bromham, L. The ecological drivers of variation in global language diversity. *Nature Communications* **10**, 2047, doi:10.1038/s41467-019-09842-2 (2019).
- 24 Dunn, R. R., Davies, T. J., Harris, N. C. & Gavin, M. C. Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences* **277**, 2587-2595, doi:10.1098/rspb.2010.0340 (2010).
- 25 Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**, 744-751 (2019).
- 26 Ushey, K., Allaire, J. J. & Tang, Y. reticulate: Interface to 'Python'. R package version 1.14. <https://cran.r-project.org/package=reticulate>. (2019).
- 27 Simpson, E. H. Measurement of diversity. *Nature* **163**, 688-688 (1949).
- 28 Klein Goldewijk, K., Beusen, A. & Janssen, P. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene* **20**, 565-573 (2010).
- 29 Suryanarayana, M. H., Agrawal, A. & Prabhu, K. S. Inequality-Adjusted Human Development Index for India's States 2011 (United Nations Development Programme, New Delhi, 2013).
- 30 Kinyoki, D. *et al.* Mapping child growth failure across low-and middle-income countries. *Nature* **577**, 231-234 (2020).
- 31 Department of Medicine and Therapeutics Centre for Nutritional Studies, T. C. U. o. H. K. & The Department of Health Hong Kong, S. G. A Survey of Infant and Young Child Feeding in Hong Kong: Diet and Nutrient Intake. (2012).
- 32 Sensarma, K., Chopra, H. S., Sharma, N. K., Balani, N. & Rawat, S. S. Road Accidents in India. (Ministry of Road Transport and Highways, Government of India, New Delhi, 2008).
- 33 rgeos: Interface to Geometry Engine - Open Source (GEOS) v. 0.3-21 (R package, 2016).
- 34 Vavrek, M. J. fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* **14:1T** (2011).