# Biogeography of human infectious diseases: a global historical analysis

Elizabeth Cashdan
University of Utah

Geographic variation in infectious disease has played a major role in determining history's political and demographic winners and losers [9, 25], and remains a significant factor shaping differential welfare across the world today. We know a great deal about the ecological conditions that influence the distribution of particular pathogens in particular parts of the world, but there have been comparatively few analyses of global pathogen distributions and their determinants. On the other hand, theory in geographical ecology has addressed global patterning in species distributions across a wide range of taxa. The aim of this paper is to evaluate some of those arguments in the context of human pathogens, by assessing the relative influence of environmental variables that have been found to shape species diversity in other taxa. Among the factors considered are climate (temperature and precipitation), island size and isolation, and human factors that enhance disease transmission (population density, sedentism, and roads).

The dataset is unusual in being historical (mid-twentieth century or earlier) and in taking as units of observation the local pathogen and environmental conditions prevailing at 186 mostly small-scale non-industrial societies around the globe (the Standard Cross-Cultural Sample). The data are specific to these locations, which are for the most part not near major population centers and transportation hubs. While the use of historical pathogen data poses obvious limitations in accuracy and precision, it has the potential to give a clearer picture of the role of the physical environment, since influential moderators (global travel, modern medicine and public health) played a smaller role than they do today. The more that global disease patterns rest on differential access to vaccines and antibiotics, good sanitation, and clean water, the more difficult it becomes to isolate the effect of climate and other biogeographical variables in a global analysis. The dataset also has the advantage, when compared to national data such as GIDEON, of being spatially focused and on a consistent scale. Finally, a number of relevant cultural variables have been coded for the SCCS, including several that are likely to affect pathogen abundance and diversity. The present study, therefore, complements global biogeographical pathogen analyses that have used modern datasets [11, 16].

Latitudinal gradients in species richness are reported for a wide range of taxa, including human pathogens [11, 16], but the reason for the gradient remains a subject of debate [35, 44]. Energy and water availability affect organism abundance because they are central to metabolism, but it is less clear why more energy or water leads to greater number of species; it is likely that there are several mechanisms, and that they vary by taxa [7]. Empirical studies have shown that temperature (used as a proxy for energy availability) is often correlated with species richness, but other studies have shown similar patterning with precipitation and habitat diversity. This study assesses the relative importance of these variables as predictors of historical pathogen number and prevalence. In addition to developing a global model, the study tests the hypothesis that temperature is more important in areas where it is limiting (e.g., areas far from the equator), while water [17] and habitat diversity [18] are more important in areas of energy abundance.

Species richness on islands is also shaped by island size and isolation. The MacArthur & Wilson [24] equilibrium model of island biogeography explained this relationship as a consequence of immigration and extinction rates: smaller is-

lands have fewer species due to higher extinction rates and fewer habitats, and more isolated islands have fewer species due to lower colonization rates. Larger islands also attract more immigrants (target effect) and less isolated islands receive repeated immigration and so are less vulnerable to extinction (rescue effect). While the assumption of equilibrium is problematic and new dynamic theories have been developed [20, 43], the influence of island area and isolation remain important. A separate analysis of the 37 islands in the sample was therefore conducted to see whether the size and isolation of islands shape pathogen number and severity, and, if so, whether greater habitat diversity on larger islands could explain the relationship.

Finally, the SCCS also allows us to include in the models aspects of human demography and culture likely to affect pathogen growth and transmission. High host population densities are associated with greater pathogen diversity across a range of non-human primate taxa [31], and the same is likely to be the case for humans. Skeletal and other evidence suggests that the neolithic transition to settled farming and husbandry was often accompanied by an increase in infectious disease; proposed reasons include the larger pool of susceptible hosts and wider contacts arising from larger, denser, and more permanent settlements, as well as exposure to new zoonoses and vectors associated with food production [2, 8, 12]. Similar factors are likely to lead to variation in pathogen exposure among the nonindustrial societies of the SCCS. These factors are evaluated here by modeling the effects of population density, sedentism, and road quality.

# 1 Methods

**Pathogen Data** The SCCS is a worldwide sample of 186 non-industrial cultures pointpointed both in time and space [27] (see figure 1). The ethnographic present of most of the SCCS societies is in the early part of the twentieth century (interquartile range 1880–1939). Cashdan & Steele [6] developed pathogen codes for the SCCS using historical sources, chiefly global maps created between the 1930s and 1950s. The codes reflect the prevalence of eight pathogens: malaria, dengue, filariae, typhus, trypanosomes, leishmanias, schistosomes, and plague. Most of these pathogens include several related species, due to limitations of the source material. Prevalence was inferred from qualitative map codes (e.g., high prevalence or heavy infection, low prevalence or slight infection, rare but present, absent). The coding procedure followed that used by Murray & Schaller [30] in their historical cross-national pathogen codes, but was made specific to local conditions by recording the highest pathogen value (1-4) within a 100 km. radius of each SCCS society. The main sources were the three volume series of maps in Rodenwaldt & Bader [34] and the maps and data in Simmons et al. [36], supplemented by data in Faust & Russell [14]. Low [22, 23] developed historical codes for the SCCS using different sources. The two codes are highly correlated, but Low's include two pathogens (leprosy and spirochetes) not in the Cashdan and Steele database. A combined index was therefore created by converting Low's three-point scale for leprosy and spirochetes and Cashdan and Steele's four-point scale for the other eight pathogens to z-scores, and using the mean of the 10 z-scores as the index of pathogen prevalence. A high score in both codes indicates both many types of pathogens and severe exposure. In order to get a measure that more closely reflects species richness, a second index was created in which pathogens were dichotomized as either present or absent. The score here is the number of pathogens out of a total possible of 10. All analyses were done with both the pathogen prevalence index and with the number of pathogens.

Because of limitations in the source material, both codes are biased toward pathogens that

Figure 1: Standard cross-cultural sample locations

are transmitted through arthropod and other vectors. A few of these also have non-human hosts. The prevalence of such diseases is strongly shaped by the geographic distribution of the vectors that transmit them and the species that host them. This bias has the disadvantage that a number of important diseases (e.g., measles and cholera) are omitted. It also has the advantage that the geographic patterning of this sample of diseases will be less affected by international travel and by socioeconomic and public health measures than are diseases spread via droplet and oral-fecal transmission.

**Environmental and Island Data** Energy measures included in this study were mean annual temperature, number of frost-free months, and within-year measures of temperature extremes, all coded by Whiting [41]. Water availability was measured by yearly mean precipitation over a 20-year period, coded by Cashdan [5], and within-year measures of wet and dry extremes, including lowest precipitation in dryest month and highest in wettest month, coded by Whiting [41]. All data were taken from weather station records closest in time and place to the focus of each SCCS society. Habitat diversity was coded as the number of vegetation types within a radius of 100 through 250 miles [4, 5], based on world maps published in the 1960s by Eyre [13].

Many sociocultural factors affect pathogen spread, directly or indirectly, and three are used in these analyses. Population density [29] is coded as a 7-level ordinal scale ranging from 1 (less than 1 person/5 sq. mi) to 7 (more than 500 persons/sq. mi). Road quality [28] is a 4-level ordinal scale, and sedentism [29] is a 6-level ordinal scale of residential mobility. The environmental and cultural data analyzed here come from the 2003 World Cultures 14(1) data disks, although the original published sources were consulted for full variable definitions and coding procedures. The repository is available online at `http://www.escholarship.org/uc/wc_worldcultures`.

Island area and various measures of isolation were obtained from the UNEP (United Nations Environment Programme) Island Directory at `http://islands.unep.ch`, supplemented in a

few cases by other sources. A few islands were so small that the 100 km radius used to calculate pathogens extended beyond the island border. In these cases, if there was another island within that radius, the area of that island was added to the focal island.

**Analysis** Island area and some climate variables were transformed with a natural log transform prior to regression to make relationships linear and improve residual distributions. Where necessary, a constant was added so that the minimum value was at least 1.0. For variables with a negative skew, the data were reflected about zero and then translated to achieve a minimum value of 1.0. After the log transform they were again reflected to restore the original order. The pathogen prevalence index is specific to the data sources that were available and is not readily generalizable, so regressions for pathogen prevalence report only standardized (beta) coefficients. SAS was used for all analyses.

# 2 Results

The upper graph in Figure 2 shows that pathogen prevalence is negatively correlated with distance from the equator, and that island societies have lower pathogen scores than those on the mainland. Because the pathogen prevalence index conflates number of species and abundance, the lower graph uses an index based solely on pathogen presence or absence; it shows a similar picture, with islands having fewer pathogens than expected given their latitude. This result is consistent with the broader literature on island biogeography, which finds species richness to be reduced on islands, and will be dscussed further in a later section. First we turn to the climatic factors that might be influencing the latitudinal gradient. In this dataset, mean annual temperature and precipitation are both correlated with distance from the equator (mean annual temperature: $r = -.80, p < .0001, n = 180$; mean annual precipitation: $r = -.50, p < .0001, n = 186$), so the first question is which variable is more important in shaping pathogen distributions, and to what extent associated variables (climate extremes and variation) also play a role. We then turn to the question of whether the determinants at high and low latitudes differ.

## 2.1 Climate

Figure 3 shows pathogens as a function of log mean annual temperature, subset in two ways to illustrate additional effects on the relationship. The upper graph shows that islands have lower pathogen scores that would be expected from their temperature, the same pattern seen with latitude. The lower graph, which excludes islands, shows that a year-round frost-free climate predisposes to more pathogens than would be expected from the climate's average temperature. Other measures of within-year temperature extremes were also analyzed, but were too highly correlated with mean annual temperature to be included in regressions. Temperature, frost, and islands have independent effects when included together in a multiple regression model: log mean annual temperature, frost months (dummy coded as some vs none), and islands (dummy coded as island vs. mainland) together explain 39% of the variance in pathogen prevalence and 40% of the variance in number of pathogens, with pathogen prevalence being higher on the mainland ($\beta = .33, p < .0001$), in areas with high mean annual tempeature ($\beta = .46, p < .0001$), and in frost-free climates ($\beta = -.26, p = .0002$).

Precipitation shows a more complicated relationship to pathogens, because two variables have independent effects: (a) mean annual precipitation and (b) the amount of precipitation in the dryest part of the year. As will be shown below, these two precipitation variables affect different kinds of pathogens. Mean annual precipitation showed a modest ($R^2 = .13$) curvilinear

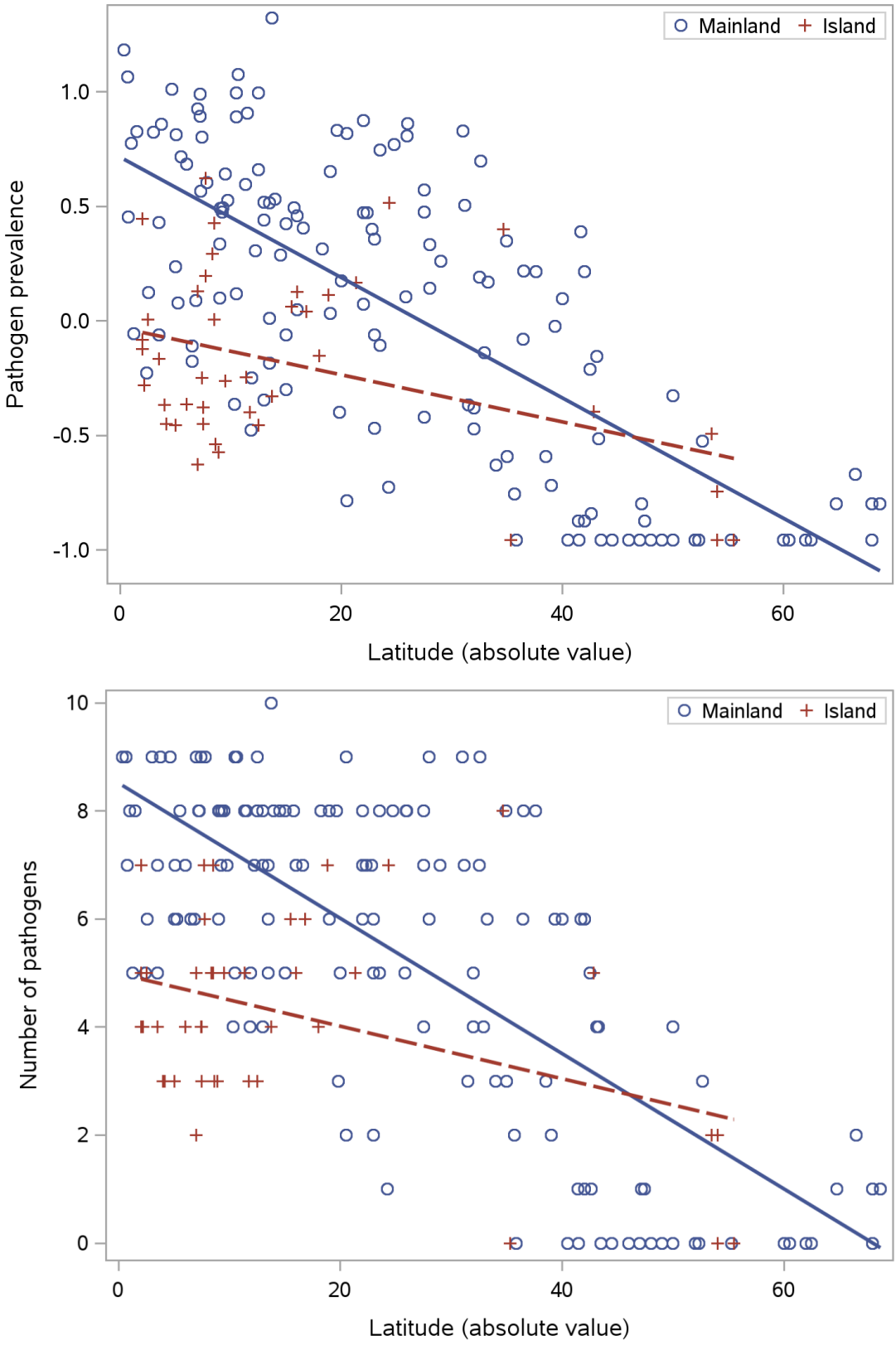
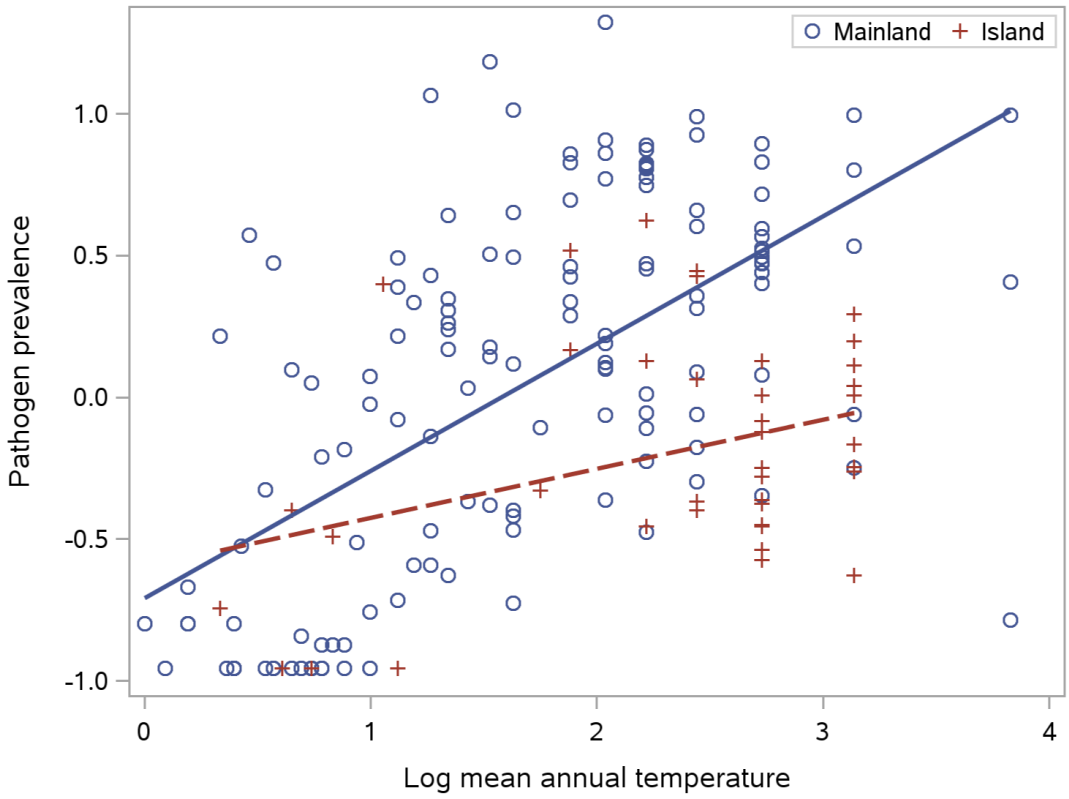



Figure 2: Pathogens by Latitude



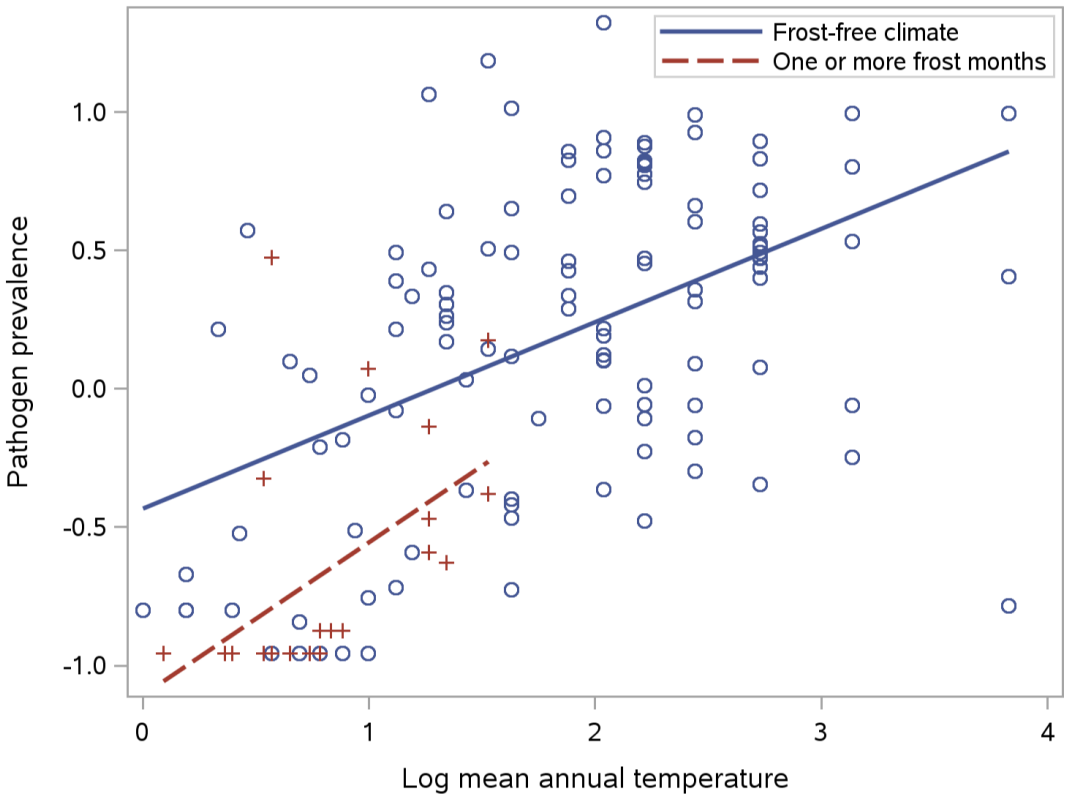




Figure 3: Pathogens by natural log of mean annual temperature (c). Regression lines in upper graph are subset by mainland vs. island sites. Regression lines in lower graph are subset by presence or absence of frost.

relationship with pathogens best approximated with a third-order polynomial (see Figure 4; one influential rainfall value was removed from this graph and from the analyses). Extreme dryness during part of the year (measured as log lowest precipitation during driest month) also enhances pathogen prevalence. Figure 5 shows this relationship, which is clarified when controlling for latitude (the figure shows dry extremes against pathogen by latitude residuals). Adding the two precipitation variables (mean annual precipitation and seasonal dry extremes) to the previous model increases the variance explained to 50% for both pathogen prevalence and number of pathogens.

Temperature and precipitation variation and extremes also vary with latitude, and it has been suggested [39] that greater climate variation leads to lower diversity and larger geographic ranges because organisms in such climates have evolved to be generalists, broadly tolerant of a wide range of climates. On the other hand, Guernier et al. [16] found for six groups of human pathogens that greater seasonal range in precipitation was associated with greater species diversity. In this dataset, greater precipitation range throughout the year (measured as maximum precipitation in wettest month minus lowest precipitation in dryest month) was correlated with significantly higher pathogen prevalence. However, precipitation range was no longer a significant predictor when temperature and other variables are included in the model, while lowest precipitation in dryest month (one componenent of precipitation range) remains significant.

The effects of temperature, mean precipitation, and dryness differ for the different pathogens, and the patterning appears to reflect the ecology of the vector more than the type of pathogen (bacteria, virus, nematode, etc). The mosquito-borne pathogens are a variable lot, including malaria (protozoans), dengue (virus), and filariae (nematodes), but all were worse in hot wet climates. Typhus (rickettsia) leishmanias (protozoans), and schistosomes (flukes) were all worse in areas with dry months, perhaps because of greater aggregations of vectors and hosts during drought. Bivariate correlations between the various pathogen groups and environmental predictors are summarized in Table 2.1.

## 2.2 Roads, Population Density, and Mobility

Pathogen distributions are affected by cultural as well as physical environmental factors. Better roads can be expected to broaden the geographic reach of pathogens by facilitating the movement of people, and of insect vectors carried inadvertently in the goods they carry. Road quality was coded for the SCCS societies on a 4-point scale, dichotomized here into societies where only footpaths were present ($n = 124$) and societies with roads of varying quality ($n = 57$). The average pathogen prevalence score was higher in the latter, as expected: .29 for societies with roads, -.14 for societies without ($t(139) = -5.17, p < .0001$).

High population density and sedentism can also be expected to enhance pathogen risk, and their effects are shown in Figure 6. Population density, coded as a 7-point ordinal scale, was strongly and linearly correlated with pathogens: $r = .46$ ($r_s = .47$), $p < .0001, n = 184$. Sedentism, dichotomized into the 117 societies that maintain permanent camps (5-6 in the original scale) and the 69 that move during the year (1-4 in the original scale) was also an important predictor, with the sedentary groups having pathogen prevalence scores averaging .22 as compared with -.37 for the mobile groups ($t(184) = -7.7, p < .0001$). As Figure 6 shows, increases in population density are accompanied by a trend toward increasing sedentism, although mobile populations have lower pathogens at the same degree of density. A model with just density and sedentism explains 26% of the variance in pathogen prevalence and number of pathogens.

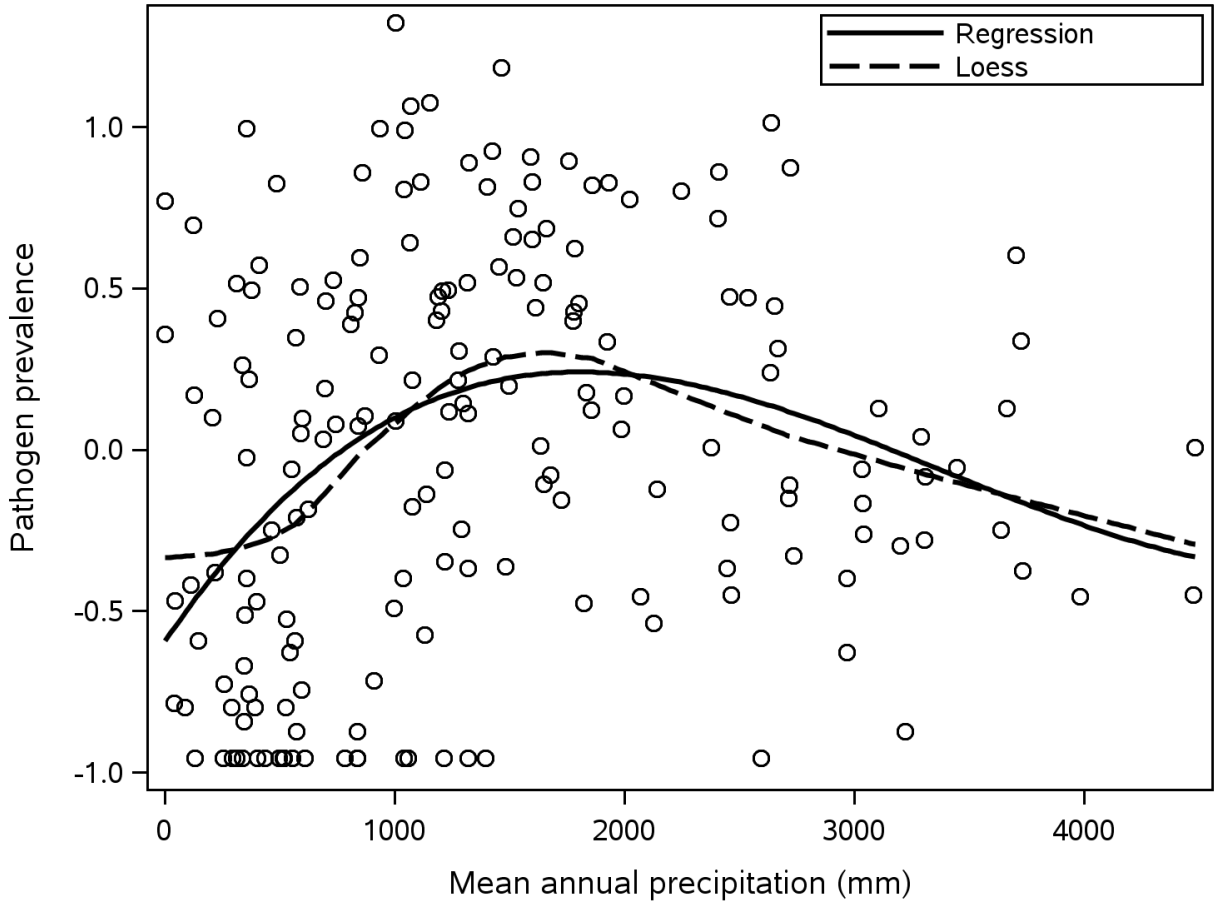



Figure 4: Pathogens by Mean Annual Precipitation

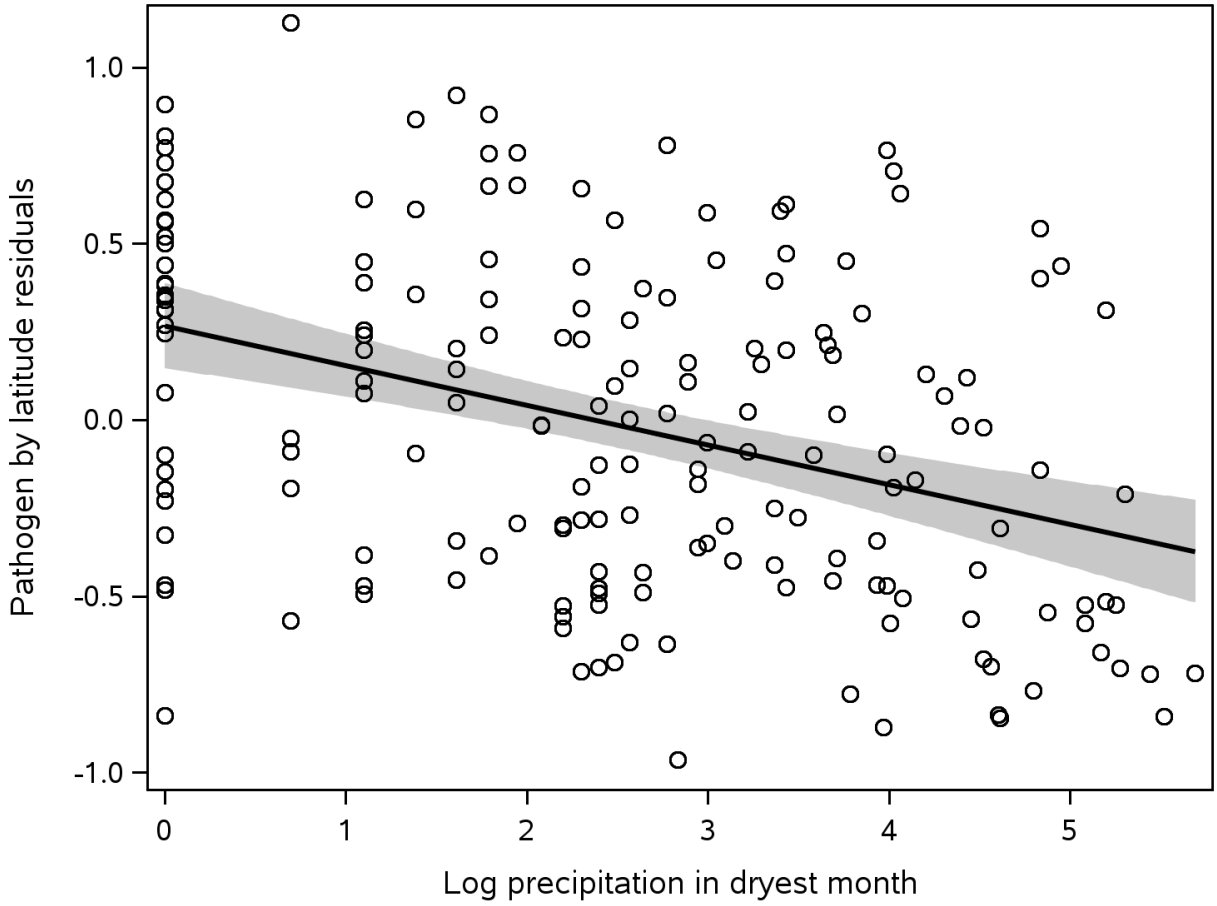



Figure 5: Pathogens by Seasonal Dry Extremes. Regression with 95% confidence limits

Table 1: Pathogen-specific correlations

| | Mean annual Temperature | Yrly Precip (linear) | Yrly Precip (polynomial) | Precipitation Dryest month | Number of Frost months | Population Density |
|---|---|---|---|---|---|---|
| Malaria | **.53***** | **.42***** | .41 | -.07 | **-.41***** | **.36***** |
| Dengue | **.56***** | **.48***** | .43 | .06 | **-.38***** | **.48***** |
| Filaria | **.53***** | **.42***** | .41 | -.02 | **-.39***** | **.46***** |
| Typhus | -.07 | -.01 | .22 | **-.39***** | **-.16*** | **.35***** |
| Trypanosomes | .06 | **.18*** | .20 | -.03 | -.12 | .05 |
| Leishmanias | **.21**** | .08 | .14 | **-.22**** | **-.33***** | .11 |
| Schistosomes | **.26***** | **.17*** | .19 | **-.24***** | **-.24**** | **.18*** |
| Plague | .07 | **.27**** | .17 | .01 | -.12 | **.19*** |
| Spirochetes | **.35***** | **.26**** | .20 | -.13 | **-.35***** | **.38***** |
| Leprosy | **.38***** | **.35***** | .23 | **-.19*** | **-.31***** | **.32***** |

Note. Pathogen prevalence by predictors in the final model of 2.3. Two correlations are given for mean annual precipitation ("Yrly Precip"): the polynomial model uses the full sample, but without significance values, which are probably unreliable for the individual pathogen scores. The linear model is limited to the 78% of the sample where pathogens increase with precipitation (i.e., up to 2000 mm). Other variables were transformed as indicated in the text, except that frost was not dichotomized. Spearman's rank-order correlations were used for all variables other than the polynomial. Sample sizes are 180 for log temperature, log precipitation in dryest month, and number of frost months; sample sizes for the polynomial and linear correlations with mean annual precipitation are 185 and 143 respectively.
$* < .05, ** < .01, *** < .001$

For pathogen prevalence, the standardized coefficients were $\beta = .28, p = .001$ for density, $\beta = .27, p = .002$ for sedentism, VIF=1.84 (the significance of road quality drops out with these variables in the model).

Examination of outliers in the climate analysis underscores the importance of considering the cultural as well as physical environment. For example, there was a highly influential point in the temperature and frost model. This point represents the Teda, a nomadic group in Chad with an unusually low pathogen score, given their local temperature (very warm) and rainfall (very low). None of the physical environmental factors in the dataset explain the discrepancy adequately, but their comparatively low pathogens are consistent with their very low density (less than 1 person per 5 sqare miles, 1 on the 7-point scale) and high mobility (1 on the 6-point scale) at the time and place of their SCCS ethnographic description. Population density, sedentism, and roads are all correlated, and in a combined model with the environmental variables only density remains statistically significant. However, figure 6 suggests that the effect of density is due both to its direct effects and to indirect effects resulting from associated decreased mobility.

### 2.3 The global model

Taken together, the results indicate that there are more pathogens and pathogen types on the mainland than on islands, and that pathogens increase with mean temperature, population density, and a frost-free climate. The relationship with precipitation is more complex, peaking at intermediate levels of mean annual precipitation but also increasing in seasonally dry climates. The full model, with both physical and cultural environmental variables, explains 58% of the variance in both number of pathogens and pathogen prevalence. The regression statistics of this model, with standardized regression coefficients ($\beta$), are in Table 2.3.

As a check, the final model was run against each of the two databases from which the combined pathogen score was derived. Low [22, 23] coded data on seven of the pathogens using different historical sources, only two of which were used in the combined index. Using Low's summary score as the dependent variable with this model produces an identical $R^2 = .58$, although the coefficient for density was smaller and that for temperature was larger. The greater influence of temperature using Low's data is probably because typhus and plague, which are unrelated to temperature in table 1, were not included in that dataset. The other coefficients were similar to those of the combined index. The same model applied to the index of Cashdan & Steele [6], which sums scores for 8 pathogens, yields an $R^2 = .60$, with coefficients very similar to those of the combined index.

Another check was done to see whether the relationship between the independent and dependent variables in the model was due to spatial autocorrelation (e.g., whether independent and dependent variables were associated only because they vary similarly across space). Where this is the case, there will be spatial autocorrelation in the residuals. In order to evaluate this, the squared difference between the residuals of each pair of points was plotted against their great circle geographic distance. Visual inspection indicated that the relationship was flat at all scales of distance. Correlations at various scales of distance (full sample and points less than 3000, 1000, 500, and 200 km. apart) averaged $-.02$ and did not show any trends with distance.

**Differences between high and low latitude regions.** The model above is the best fit for global pathogen distributions, but recent literature suggests that more specific models may be appropriate at high and low latitudes. Energy availability appears to have a greater effect on species richness farther from the equator, whereas water [17] has been proposed as more important where energy is abundant. Habitat diversity [18] may also be more important at low latitudes. The sample was divided into tropical

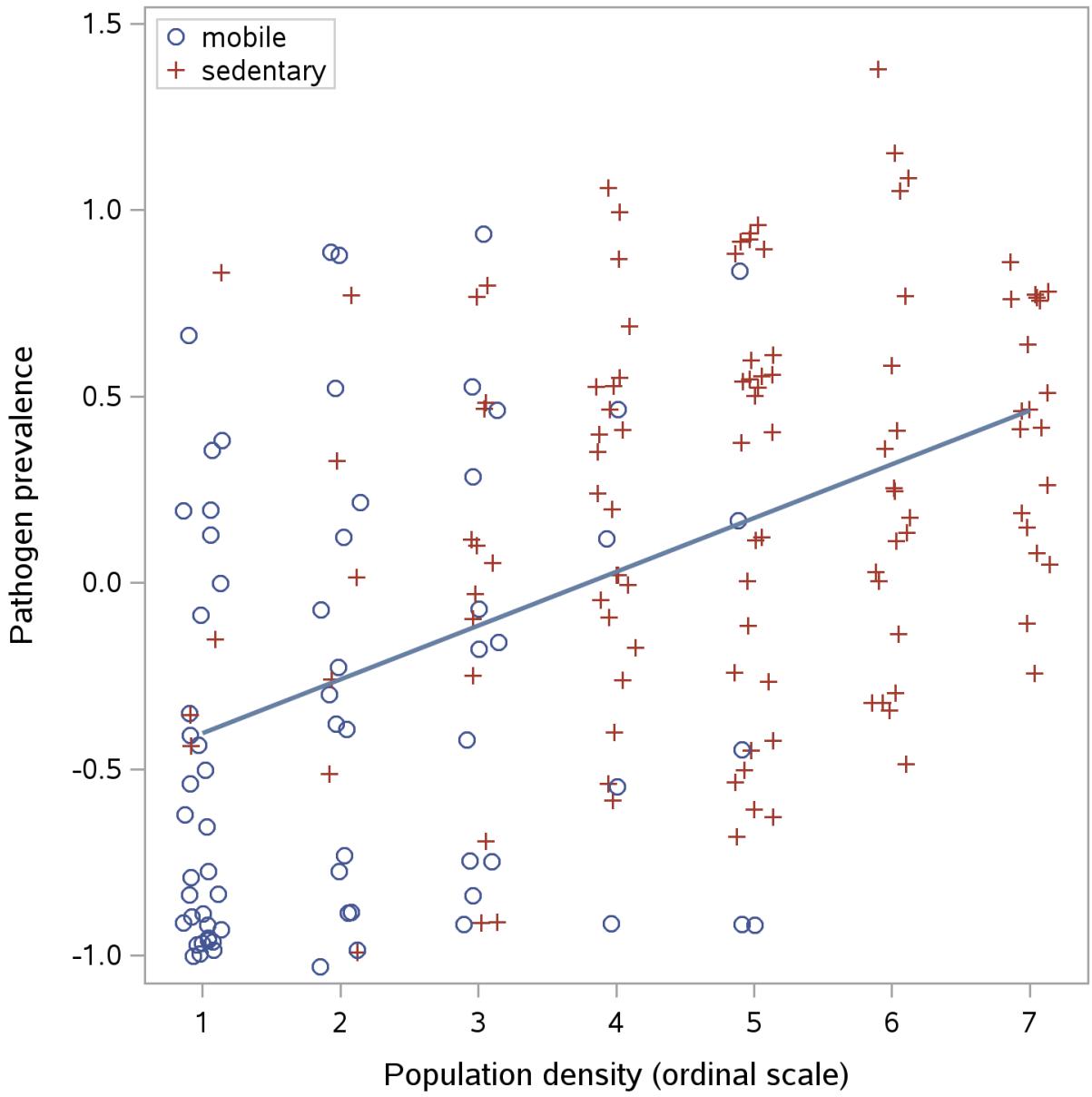



Figure 6: Pathogens by Density and Residential Mobility. Points have been jittered to avoid overlap. Regression line is based on all points.

Table 2: Final regression model for predictors of pathogen prevalence and number of pathogens

| | Pathogen Prevalence | | Number of Pathogens | |
|---|---|---|---|---|
| Variable | $\beta$ | $p$ | $\beta$ | $p$ |
| Temperature | 0.30 | $< .001$ | 0.31 | $< .001$ |
| Frost months | -0.19 | .001 | -0.20 | $< .001$ |
| Island | 0.31 | $< .001$ | 0.31 | $< .001$ |
| Driest month | -0.18 | .009 | -0.19 | .006 |
| Pop Density | 0.31 | $< .001$ | 0.30 | $< .001$ |
| Precipitation (P) | 1.78 | $< .001$ | 1.62 | $< .001$ |
| $P^2$ | -2.73 | .001 | -2.45 | .044 |
| $P^3$ | 1.15 | .024 | 1.03 | .006 |
| | $F(8,168)=29.57$ | | $F(8,168)=29.23$ | |
| | $R^2 = .58$ | | $R^2 = .58$ | |

Variable definitions:
Temperature: Log mean annual temperature (c)
Frost months: 1=presence 0=absence
Island: 1=mainland 0=island
Driest month: Log lowest rainfall in driest month (mm)
Pop Density: Population density (ordinal scale, 1-7)
Precipitation: Mean annual precipitation (mm)

(low latitude) and non-tropical (high latitude) zones, and the results supported these expectations. Bivariate correlations by latitude zone are shown in Table 3.

As expected, temperature was a significant predictor only at high latitudes. In this region, the relationship was strongly linear (Pearson's $r = .63$ for number of pathogens, $r = .66$ for pathogen prevalence). An unanticipated result was that the same is true for population density: it is a strong predictor of pathogens outside the tropics only. A multivariate model using only those two variables (log mean annual temperature and population density) explains 56% of the variance in pathogen prevalence and 61% in pathogen number at high latitudes. No other variables add significantly when those are in the model, although the best climate-only model for this region also includes lowest rainfall in dryest month.

The pattern in the tropical locations, in contrast, is shaped more by precipitation than by temperature. The relationship between precipitation and pathogens in the tropics is similar in shape to that shown in 4 for the full sample, but the relationship is much tighter, the peak is at somewhat lower precipitation, and the pathogen decline at higher precipitation is more apparent. The best multivariate model of pathogen prevalence in the tropics includes mean annual precipitation as a third-degree polynomial together with population density (notwithstanding its weak bivariate relationship) and the dummy-coded island vs. continent. This model explains 44% of the variance in pathogen prevalence and 50% in pathogen number.

Table 3 shows that habitat diversity (measured as number of vegetation zones in a given radius) is also a factor in shaping pathogen diversity in the tropics. Habitat diversity remains significant when added to the other variables (rainfall, islands, density) in the tropical model. However, the overall $R^2$ is reduced, perhaps because of reduced sample size when that variable was included. It is worth noting that habitat diversity, alone among the variables considered in this study, is a stronger predictor of number of pathogens than it is of prevalence. This presumably reflects the fact that niche differentiation associated with habitat diversity affects the number of species that can inhabit an area, whereas temperature, precipitation and population density also affect prevalence through effects on pathogen growth and transmission.

### 2.4 Islands

Island biogeography predicts that islands will have fewer species than continents, both because extinction rates are higher in small, bounded areas and because distance from reservoirs poses a barrier to new immigrants. The smaller the island and the greater the distance, therefore, the fewer species are expected. Figure 2 showed that both pathogen prevalence and number (an indicator of richness) were lower on islands than would be expected from their latitude.

The relationship of size to pathogen number is also consistent with expectations, with larger islands having more pathogens than small islands, controlling for latitude. The analysis is based on 37 islands (New Guinea was excluded because it is home to four societies in the sample). Controlling for distance from the equator, the partial correlation of log island area with pathogens is $r = .63, p < .0001$ for pathogen prevalence and $r = .61, p < .0001$ for number of pathogens. Figure 7 shows the relationship for islands in the tropics; the non-tropical islands are included in the statistics but are not shown in the figure because they span a wide latitudinal range.

As Figure 7 indicates, the linear relationship breaks down for the smallest islands. This is often the case in small islands, where the effect of area on species richness is overshadowed by stochastic factors [19, 21]. In such cases, however, species richness typically plateaus at the lowest level, which is not the case in these data. It may, therefore, be an artifact of the

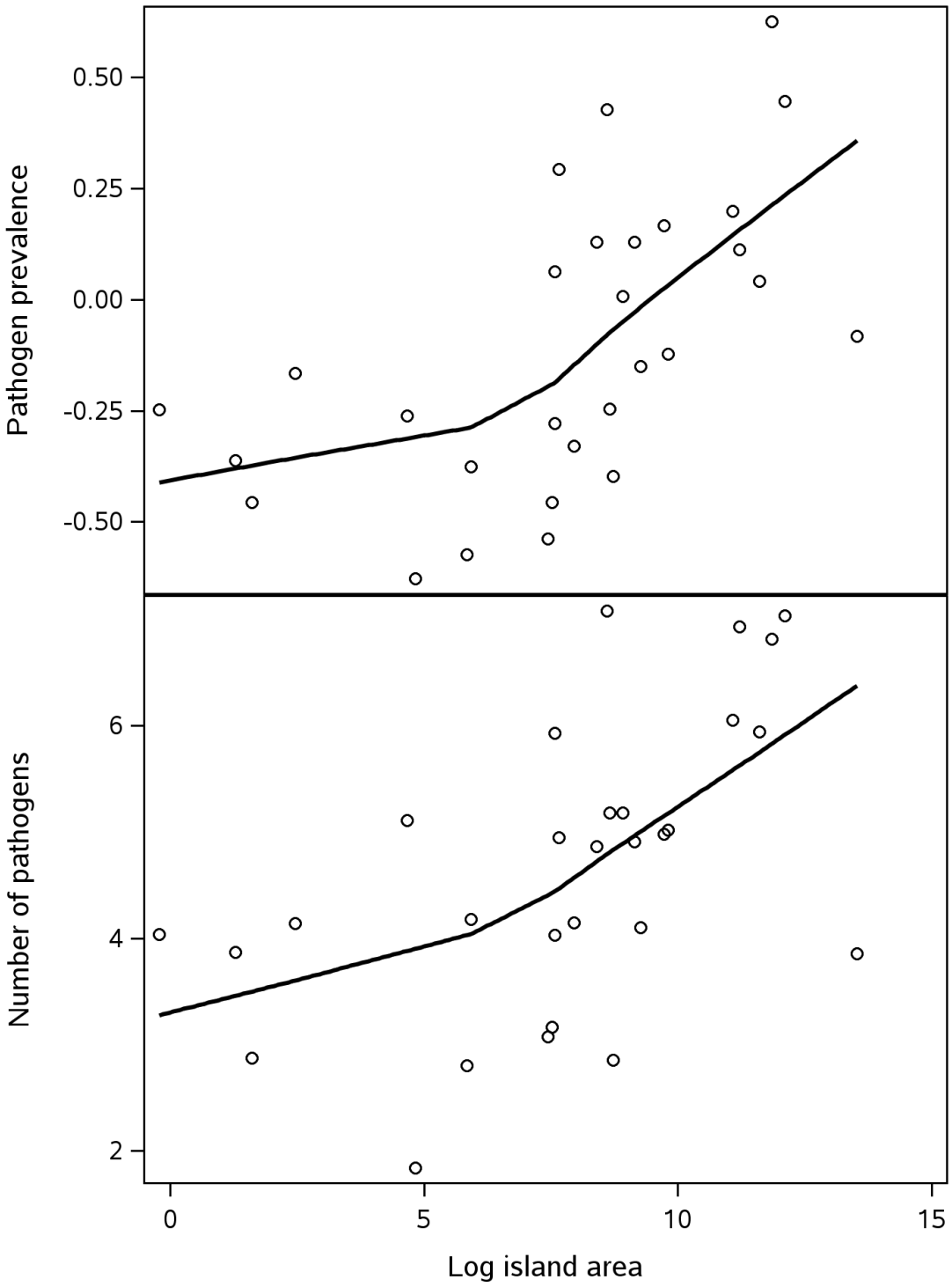



Figure 7: Pathogens by Island Area (tropics), with loess curves.

Table 3: Bivariate correlations between independent variables and pathogen prevalence and number of pathogens, by latitude zone

| | Low latitudes | | High latitudes | |
|---|---|---|---|---|
| | Number | Prevalence | Number | Prevalence |
| Mean temperature | -.20* | -.18 | **.59***** | **.61***** |
| Mean precipitation | **.44***** | **.40**** | .30 | .35 |
| Low precip dryest month | **-.33***** | **-.31**** | **-.33**** | **-.31**** |
| Population density | .14 | .17 | **.67***** | **.69***** |
| Habitat diversity | **.36***** | **.24***** | .05 | -.01 |

Note. Table shows Spearman's rank order correlation coefficients except for mean annual precipitation, where the correlation is based on the adjusted $R^2$ of a third-degree polynomial regression (no significance values are given for the nontropics because of poor fit diagnostics). Sample sizes for low/high latitudes: temperature 107/73, precipitation 111/75, precipitation dryest month 107/73, habitat diversity 97/75, density 109/75. Habitat diversity calculated at 150 miles radius; the correlation was slightly less at 100 miles.
$* < .05, ** < .01, *** < .001$

poor resolution of the historical (mid 20th century) pathogen data; it seems likely that direct disease data was not collected for many of the smallest islands, and that the maps reflect inferences made from larger islands in the region. A regression without the four smallest islands probably presents a more accurate picture of the relationship between pathogens and island area; in this model the largest island (Borneo) is also best removed as it is a highly influential point. Within this range of values, the relationship is linear with the log of island area, and the regression of log area and latitude on pathogen number provides a better fit: $F(2, 29) = 23.51, p < .0001$, $R^2 = .62$ (adjusted $R^2 = .59$). The unstandardized coefficient for log island area on number of kinds of pathogens is 0.64, and the relationship with pathogen prevalence is similar.

Theory predicts that pathogens will also decrease with distance from the mainland. The relationship is weak in this dataset, and its independent effect is hard to evaluate since the smallest islands are also farthest from the mainland. Getting a good measure of isolation is difficult, since it involves not just distance to the nearest continent but to nearby islands that could be links to sources of greater diversity. Various distance measures were used to try to capture this, but none showed more than a weak correlation with pathogen load, or remained significant when island area was also included in the model. However, this could reflect measurement difficulties rather than relative importance. It is also possible that some of the area effect reflects the greater isolation of many of the smaller islands.

# 3 Discussion and Conclusions

A single model with 5 variables explained most (58%) of the variance in both historical pathogen prevalence and richness. Pathogens were more numerous on the mainland than on islands, and on large as opposed to small islands. Both temperature and precipitation were significant predictors. Pathogens increased with mean annual temperature and, controlling for mean temperature, in climates that remained free of frost throughout the year. The effect of temperature was highly significant, but only outside the tropics. Within the tropics, mean annual precipita-

tion was a more important predictor. The relationship with mean annual precipitation was curvilinear, peaking at slightly less than 2000 mm/year, and was associated most strongly with mosquito-borne diseases (malaria, dengue, filariasis). Extreme seasonal dryness was also associated with more pathogens, increasing primarily the prevalence of typhus, leishmaniasis, and schistosomiasis. Finally, pathogens were worse in areas with high population density. Most of the predictions derived from species diversity patterns in other taxa were supported, and are discussed in turn below, beginning with the island results.

**Island biogeography.** The classic model of island biogeography[24] predicts that there will be fewer species on islands that are small (due to higher extinction rates) and isolated (due to lower colonization rates). The model and its later refinements use simplifying assumptions [20], and the assumption of equilibrium is particularly problematic when studying the distribution of human infectious diseases. Nonetheless, and notwithstanding the heterogeneity of the 37 islands in this dataset, the classic predictions of island biogeography were upheld: islands have fewer kinds of pathogens than expected given their climate, and pathogen number (and prevalence) are negatively correlated with island area. Island isolation did not have an independent effect, but may have had an indirect influence since many of the smaller islands in this dataset were also more isolated. Smaller islands also offer fewer types of habitat, but this appears not to be driving variation in pathogen richness in this dataset: although habitat diversity and area of islands were correlated, only island area had any relationship to pathogens.

The lower pathogen load on islands is consistent with Curtin's (1989) meticulous accounting of historical troop mortality, which found that early in the 19th century at least some Pacific islands had a far lower pathogen load than would be expected by their climate, particularly Tahiti "which gave French troops a 100% mortality improvement over France" in the 1840s and continued to give benefits into the early 20th century [10, p. 12]. A similar pattern existed for Hawaii and New Zealand. (This ecological protection later made the islanders vulnerable to novel European diseases, which caused huge mortality throughout the Pacific).

Most of the pathogens in this dataset are vector-borne, and this is likely to enhance the influence of island area and isolation since two species, pathogen and vector, need to be present at the same time [38]. Some pathogens also require a minimum size of host population in order to remain endemic. This triple challenge can be expected to amplify the effect of isolation and extinction on pathogenic species on small islands, and among small isolated populations generally [3].

**Effects of climate and latitude.** Both prevalence and number of pathogens show a strong latitudinal gradient in these data. A latitudinal gradient in species richness exists across a very wide range of taxa, although the cause of this pattern remains a topic of debate [35, 44]. A number of mediators have been discussed in the literature; those discussed below include temperature (often used as a proxy for energy availability) and frost, precipitation, and habitat diversity.

Pathogen prevalence increased with temperature in 7 of the 10 pathogen groups, including the mosquito-borne pathogens, which are known to be highly temperature sensitive [33]. Although growth and survival of insect vectors, hence pathogen transmission, decline in extreme heat [26], the relationship with log temperature was linear in these data. There were also more kinds of pathogens as temperature increased. Locations with a year-round frost-free climate had more pathogens than would be expected from their mean annual temperature, probably by enabling pathogens and their vectors to overwinter.

Precipitation had a more complex relationship with pathogen prevalence, because mean yearly precipitation and extreme seasonal dryness had different effects and affected different pathogens. Mean annual precipitation had a curvilinear relationship with pathogen prevalence and number, peaking at intermediate values and declining in very high precipitation areas. The association between pathogens and mean annual precipitation was especially strong in the tropics and for mosquito-borne pathogens; perhaps the decline in very wet areas is associated with mosquito larvae being washed out due to heavy rains and flooding.

Areas with little or no precipitation during the dryest part of the year (lowest precipitation in dryest month) also had more pathogens. Seasonal dryness affected a different group of pathogens than did mean temperature and precipitation. The greatest effect was on typhus. Typhus is transmitted by fleas and lice and can become worse in crowded conditions with poor sanitation, and when drought causes rodents (and the fleas they carry) to move near human habitation in search of water. An analysis of tree-rings in pre-industrial Central Mexico found that a significant drought occurred during the first year of all 22 large typhus outbreaks studied [1]. A similar effect via aggregation of people and the sand fly vector during dry periods has been associated with temporal changes in leishmaniasis in Brazil [40].

Guernier et al. [16] found precipitation range to be the single best predictor of species richness across six categories of human pathogens. In the present dataset, precipation range (highest precipitation in wettest month minus lowest precipitation in dryest month) was also associated with significantly higher pathogen number (and prevalence). However, range was not a significant predictor when other climate variables were included in the model, while seasonal dryness (one component of range) remained significant. In these data, therefore, the more influential aspect of precipitation range on pathogen distributions appears to be seasonal dryness.

The relative importance of temperature and water as predictors of species richness has been shown to vary with latitude, with temperature being more important at high latitudes and availability of water more important at low latitudes, where energy is already abundant [17, 42]. These patterns were also found in the pathogen data. In non-tropical locations, mean annual temperature was the strongest climatic predictor of pathogen number and prevalence, whereas in tropical areas mean annual precipitation was the key climate variable. Surprisingly, population density was also a much stronger predictor in high latitude areas; one plausible reason for this finding is that there are more alternate animal hosts in the tropics, so that zoonotic pathogens may remain endemic there even when people are at low density. Habitat diversity was also correlated with pathogen number in tropical regions.

**Historical data: Advantages and limitations.** Cross-national differences in pathogen prevalence today have been shown to reflect differential access to disease prevention measures more than environmental variables, although pathogen richness still shows the latitudinal gradient found with other taxa [11]. Historical data on remote populations, as used here, reduces the influence of public health and modern medicine, allowing for a clearer picture of the way environmental variables shape pathogen distributions. A second advantage of this dataset for environmental analysis is that the data describe local conditions at a consistent scale, rather than being based on national averages. In the present study, pathogen prevalence and number of kinds of pathogens show similar patterns, and are strongly environmentally determined.

Use of historical data have limitations as well as advantages, due primarily to lack of precision in the historical source material: pathogen distributions were for the most part not available at the species level, and prevalence was assessed by a four-point ordinal scale for each

pathogen, rather than by direct counts of infected individuals. Limitations of the source material also bias the pathogen sample toward vector-borne pathogens, which are less global than other pathogens [37] and are likely to show a stronger environmental signature. For this reason, prevalence and richness are more likely to be correlated with each other in this group of pathogens, even in modern datasets [15], and latitudinal gradients and climatic correlates are likely to reflect vector as well as pathogen ecology [32].

The results of this study show strong support for several theoretical and empirical findings in geographical ecology, and show that they explain human pathogen distributions on a global level. The results offer insights into past and present patterns of infectious disease, and provide information relevant to the likely effects of global warming on pathogens sensitive to temperature, frost, and seasonal dry extremes.

## 4 Acknowledgments

I am grateful to Adrian Bell, Damian Murray, Bobbi Low, Alan Rogers, and Randy Thornhill for comments, suggestions, and support. Alan Rogers also helped with the spatial autocorrelation analysis.

## References

[1] Acuna-Soto, R., Stahle, D., Villanueva Diaz, J., & Therrell, M. (2007). Association of drought with typhus epidemics in Central Mexico. In *AGU Spring Meeting Abstracts*, volume 1.

[2] Barrett, R., Kuzawa, C. W., McDade, T., & Armelagos, G. J. (1998). Emerging and re-emerging infectious diseases: the third epidemiologic transition. *Annual review of anthropology*, (pp. 247–271).

[3] Black, F. L. (1975). Infectious diseases in primitive societies. *Science*, 187(4176), 515–518.

[4] Cashdan, E. (2001). Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity. *American Anthropologist*, 103(4), 968–991.

[5] Cashdan, E. (2004). Rainfall, habitat diversity, and ethnic diversity codes for the standard cross-cultural sample. *World Cultures*, 13(2).

[6] Cashdan, E. & Steele, M. (2013). Pathogen prevalence, group bias, and collectivism in the standard cross-cultural sample. *Human Nature*, 24(1), 59–75.

[7] Clarke, A. & Gaston, K. J. (2006). Climate, energy and diversity. *Proceedings of the Royal Society B: Biological Sciences*, 273(1599), 2257–2266.

[8] Cohen, M. N. & Armelagos, G. J. (1984). *Paleopathology at the Origins of Agriculture*. Academic Press.

[9] Crosby, A. W. (1986). *Ecological imperialism: the biological expansion of Europe, 900-1900*. Cambridge University Press.

[10] Curtin, P. D. (1989). *Death by migration: Europe's encounter with the tropical world in the nineteenth century*. Cambridge University Press.

[11] Dunn, R. R., Davies, T. J., Harris, N. C., & Gavin, M. C. (2010). Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences*, 277(1694), 2587–2595.

[12] Eshed, V., Gopher, A., Pinhasi, R., & Hershkovitz, I. (2010). Paleopathology and the origin of agriculture in the Levant. *American journal of physical anthropology*, 143(1), 121–133.

[13] Eyre, S. R. (1968). *Vegetation and soils: A world Picture*. E. Arnold Ltd., second edition.

[14] Faust, E. & Russell, P. (1964). *Craig and Faust's Clinical Parasitology.* Philadelphia: Lea and Febiger.

[15] Fincher, C. & Thornhill, R. (2008). Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religion diversity. *Proceedings of the Royal Society B: Biological Sciences*, 275(1651), 2587–2594.

[16] Guernier, V., Hochberg, M. E., & Guégan, J.-F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, 2(6), 0740–0746.

[17] Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E., & Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84(12), 3105–3117.

[18] Kerr, J. T. & Packer, L. (1997). Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature*, 385(6613), 252–254.

[19] Lomolino, M. & Weiser, M. (2001). Towards a more general species-area relationship: diversity on all islands, great and small. *Journal of Biogeography*, 28(4), 431–445.

[20] Lomolino, M. V. (2000a). A call for a new paradigm of island biogeography. *Global Ecology and Biogeography*, 9(1), 1–6.

[21] Lomolino, M. V. (2000b). Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, 27(1), 17–26.

[22] Low, B. (1990). Marriage systems and pathogen stress in human societies. *American Zoologist*, 30(2), 325–340.

[23] Low, B. S. (1994). Pathogen intensity cross-culturally: SCCS. *World Cultures*, 8(2).

[24] MacArthur, R. H. & Wilson, E. O. (1967). *The theory of island biogeography.* Princeton University Press.

[25] McNeill, W. H. (1976). *Plagues and Peoples.* Garden City, NY: Anchor Press.

[26] Mordecai, E. A., Paaijmans, K. P., Johnson, L. R., Balzer, C., Ben-Horin, T., Moor, E., McNally, A., Pawar, S., Ryan, S. J., Smith, T. C., & Lafferty, K. D. (2013). Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology letters*, 16(1), 22–30.

[27] Murdock, G. & White, D. (1969). Standard cross-cultural sample. *Ethnology*, 8(4), 329–369. Updated at `http://www.escholarship.org/uc/item/62c5c02n`.

[28] Murdock, G. P. & Morrow, D. O. (1970). Subsistence economy and supportive practices: Cross-cultural codes 1. *Ethnology*, 9, 302–330.

[29] Murdock, G. P. & Wilson, S. F. (1972). Settlement patterns and community organization: Cross cultural codes 3. *Ethnology*, 11, 254–295.

[30] Murray, D. & Schaller, M. (2010). Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology*, 41(1), 99–108.

[31] Nunn, C. L., Altizer, S., Jones, K. E., & Sechrest, W. (2003). Comparative tests of parasite species richness in primates. *The American Naturalist*, 162(5), 597–614.

[32] Nunn, C. L., Altizer, S. M., Sechrest, W., & Cunningham, A. A. (2005). Latitudinal gradients of parasite species richness in primates. *Diversity and Distributions*, 11(3), 249–256.

[33] Patz, J. A., Epstein, P. R., Burke, T. A., & Balbus, J. M. (1996). Global climate change and emerging infectious diseases. *JAMA: the*

*journal of the American Medical Association*, 275(3), 217–223.

[34] Rodenwaldt, E. & Bader, R. E. (1961). *World-Atlas of Epidemic Diseases (1952-1961)*. Hamburg, Germany: Falk-Verlag.

[35] Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press.

[36] Simmons, J. S., Whayne, T. F., Anderson, G. W., & Horack, H. M. (1944). *Global Epidemiology: A Geography of Disease and Sanitation*. Philadelphia: J. B. Lippincott.

[37] Smith, K. F., Sax, D. F., Gaines, S. D., Guernier, V., & Guégan, J.-F. (2007). Globalization of human infectious disease. *Ecology*, 88(8), 1903–1910.

[38] Spurgin, L. G., Illera, J. C., Padilla, D. P., & Richardson, D. S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across island populations of Berthelots pipit (*Anthus berthelotii*). *Oecologia*, 168(3), 691–701.

[39] Stevens, G. C. (1989). The latitudinal gradient in geographical range: how so many species coexist in the tropics. *American Naturalist*, (pp. 240–256).

[40] Thompson, R. A., de Oliveira Lima, J. W., Maguire, J. H., Braud, D. H., & Scholl, D. T. (2002). Climatic and demographic determinants of American visceral leishmaniasis in northeastern Brazil using remote sensing technology for environmental categorization of rain and region influences on leishmaniasis. *The American journal of tropical medicine and hygiene*, 67(6), 648–655.

[41] Whiting, J. W. M. (1989). Climate data from weather stations. *World Cultures*, 1(1), 179–199.

[42] Whittaker, R. J., Nogués-Bravo, D., & Araújo, M. B. (2007). Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins *et al.* (2003) using European data for five taxa. *Global Ecology and Biogeography*, 16(1), 76–89.

[43] Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, 35(6), 977–994.

[44] Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, (pp. 273–309).