# An equilibrium theory signature in the island biogeography of human parasites and pathogens

**Running title*:*** The island biogeography of human pathogens

**Key-words:** Infectious diseases, human, island biogeography, pathogen diversity, species-area relationship, disease ecology

**Authors**: *Kévin Jean*; William R. Burnside*; Lynn Carlson; Katherine Smith; Jean-François Guégan*

* both authors contributed equally to the work.

Kévin Jean*:* MRC Center for Outbreak Analysis, Department of Infectious Diseases Epidemiology, Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG, UK.
Center for Research in Epidemiology and Population Health, Team 4, INSERM U1018, 82 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex, France;
k.jean@imperial.ac.uk

William R. Burnside: National Socio-Environmental Synthesis Center (SESYNC); 1 Park Place, Suite 300; Annapolis, MD 21403, US.
bburnside@sesync.org

Lynn Carlson: Geological Sciences, Brown University, PO Box 1846 Providence, RI 02912, US.
Lynn_Carlson@brown.edu

Katherine Smith: Department of Ecology and Evolutionary Biology; Brown University; 91 Waterman St, Box G-W; Providence, RI 02912, US.
Katherine_Smith@brown.edu

Jean-François Guégan: UMR MIVEGEC IRD-CNRS-Université de Montpellier, Centre IRD de Montpellier, B.P. 64501, Montpellier, France; and International programme FutureEarth, ecoHEALTH initiative.
Jean-Francois.Guegan@ird.fr

Correspondence to: Kévin Jean,
Imperial College London, Dept. of Infectious Diseases Epidemiology;
St Mary's Campus
Norfolk Place
London, W2 1PG, UK
kevin.jean.lab@gmail.com,

**Words (abstract):** 300

45 **Words (main body):** 3785

46 **Number of references:** 36

## ABSTRACT

**Aim**: Our understanding of the ecology and biogeography of microbes, including those that cause human disease, lags behind that for larger species. Despite recent focus on the geographic distribution of viruses and bacteria, the overall environmental distribution of human pathogens and parasites on Earth remains incompletely understood. As islands have long inspired basic ecological insight, we aimed to assess whether the micro-organisms that cause human disease in modern times follow patterns common to insular plants and animals.

**Location**: Global and regional.

**Methods**: Relying on the publically-accessible GIDEON database, we use the spatial distribution of nearly 300 human parasites and pathogens across 66 island countries and territories to assess the current predictive value of the "equilibrium theory" of island biogeography (MacArthur & Wilson 1967). The relationships between species richness and: i) island surface area and, ii) distance to the nearest mainland were investigated with linear regression, and ANCOVAs were used to test for differences in these relationships with respect to pathogen ecology and taxonomy.

**Results**: Pathogen species richness increases with island surface area and decreases with distance to the nearest mainland. The effect of area is more than 10 times lower than that usually reported for macro-organisms but is greater than the effect of distance. The strongest relationships are for pathogens that are vector-borne, zoonotic (with humans as dead-end hosts) or protozoan.

**Main conclusion**: Our results support the theory's basic predictions: disease diversity is a positive function of island area and a negative function of island isolation. However, differences in the effects of area, distance, and pathogen ecology suggest that globalization, likely through human travel and the animal trade, has softened these relationships. Parasites that primarily target non-human species, whose distributions are more constrained by island life than are those restricted to human hosts, drive the island biogeography of human disease.

## Introduction

Infectious diseases remain one of the chief causes of human morbidity and mortality worldwide, especially among the young and the poor (Lozano *et al.*, 2012). Understanding the drivers of human pathogen diversity, a key predictor of infectious disease prevalence, is a critical challenge of the 21st century (Dunn *et al.*, 2010). The diversity of infectious agents and the burden of disease vary dramatically across the globe, as they have throughout human history (Wolfe *et al.*, 2007; Dunn *et al.*, 2010). This disease burden exerts a profound effect on the economic fortunes of entire nations and world regions (Bloom & Sachs, 1998; Bonds *et al.*, 2012). However, our understanding of the biogeography of human disease is surprisingly limited. Less than ten infectious diseases are mapped comprehensively (Hay *et al.*, 2013), and we know less about the distributions of many human parasites and pathogens than we do about those of most rare birds (Just *et al.*, 2014). As human populations grow and geographically change with urbanization and migration, exposing populations to novel social and ecological environments, there is an increasing need for first-order predictions to guide policy and future research.

Human parasites and pathogens interact both with their human hosts and with the broader environment, so their distributions should be a function of general ecological factors as well as of the specific ecology of *Homo sapiens*. Indeed, despite our sense of microbes' ubiquity, ecology still drives the worldwide distribution of human disease, the inspiration for the eponymous Baas-Becking hypothesis: "*Everything is everywhere*, *but the environment selects*" (Baas-Becking, 1934). As with species generally, the tropics have many more disease-causing species (Guernier *et al.*, 2004; Jones *et al.*, 2008; Peterson, 2008), and Earth can be divided into biogeographic human-disease regions (Just *et al.*, 2014). Considered broadly, our parasites and pathogens display patterns characteristic of animals and plants generally (Guernier *et al.*, 2004). At the same time, pestilence follows patterns of human dispersal and interaction. As anatomically modern humans migrated to new environments, such as from Africa to Eurasia and then to the Americas, our ancestors spread some pathogens, shed others, and acquired new ones along the way (Burnside *et al.*, 2012). Historic and continuing changes in human population density, promoted by agriculture and then by industrialization, engendered and supported the "crowd-epidemic diseases," such as seasonal influenza, measles and pertussis, that afflict urban residents (Bjørnstad & Harvill, 2005; Furuse *et al.*, 2010). With globalization, increasing travel, migration, and trade have spread pathogens and parasites specific to humans worldwide, though those with animals as their main reservoir and humans as secondary hosts remain more localized (Smith *et al.*, 2007). Illuminating the processes driving such large-scale epidemiological patterns is a growing focus of disease ecology (Guernier *et al.*, 2004; Dunn *et al.*, 2010; Bonds *et al.*, 2012).

A proven avenue for exploring the influence of spatial ecological and evolutionary processes is to study biodiversity patterns on islands. As Darwin argued, islands form natural laboratories where processes can be observed that are too complex to track on land masses (Darwin, 1859). MacArthur and Wilson formalized this insight in their influential "equilibrium theory of island biogeography" (MacArthur & Wilson, 1967). According to the theory, the number of species living on an island represents a dynamic equilibrium between species arriving from elsewhere (immigration) and those dying out some time after they arrive (extinction). The immigration rate declines with distance to the nearest mainland, the source

pool, while the extinction rate declines with island area, because larger islands can support larger populations with correspondingly lower probabilities of dying out. Once an island has reached ecological equilibrium, invasions will balance extinctions and the number of species will remain unchanged even though their composition may vary over time. The equilibrium theory of island biogeography has successfully explained a range of patterns of insular plant and animal species (Lomolino *et al.*, 2010) as well as of microbes with animal hosts (Bell *et al.*, 2005; Orrock *et al.*, 2011; Svensson-Coelho & Ricklefs, 2011), but its applicability to human pathogens is unclear. Previous research supports the existence of biogeographical patterns in microbes (e.g. Martiny, 2006; Hanson, 2012), but these studies were limited to free-living microbial taxa and not focused on host-associated pathogenic species. Recent, more-limited work on historic human populations supports the theorized effect of island size on the diversity of vector-borne pathogens (Cashdan, 2014), though the influence of distance and the effect of modern industrial lifestyles, with their enhanced mobility and access to medicine and public health, is less clear.

In this study, we use the Global Infectious Disease and Epidemiology Online Network Database (GIDEON) to examine whether the distribution of nearly 300 human pathogens occurring on different islands conforms to the general predictions of island biogeography theory, specifically that pathogen richness is a positive function of island size and a negative function of distance to the nearest mainland.

## Material & Methods

### *Data collection*

Analyses were based on a subset of data extracted and compiled from GIDEON (http://www.gideononline.com/). GIDEON provides clinical, geographical, and epidemiological information on 332 unique viruses, bacteria, fungi, protozoans, and helminths infecting humans in each of the 222 countries and administrative territories of the world. The database is updated regularly using publications from Medline based on a list of keywords and search information published by national Health Ministers, the World Health Organization (WHO), and the U.S. Centers for Disease Control and Prevention (CDC). As such, GIDEON is the most current, global database available for human infectious disease.

For simplicity, we use the term “pathogens” in this manuscript to cover both pathogens and parasites and consider disease names (e.g. measles) as synonymous with the infectious agents that cause them.

Following Guernier *et al.* (Guernier *et al.*, 2004) and Smith *et al*. (Smith *et al.*, 2007), we excluded pathogens causing infectious diseases that did not meet the following three criteria: (i) those with multiple etiological origins, (ii) those with major uncertainties surrounding national presence/absence, and (iii) vector- and reservoir-borne pathogens with imprecise information on hosts. The resulting database included 271 pathogens: 85 viruses, 87 bacteria, 15 fungi, 64 helminths, and 20 protozoans.

We categorized pathogens three ways to assess the importance of different ecological and evolutionary processes: by host association, by transmission mode, and by taxonomy. We assigned host associations following Smith *et al.* (2007) as: human specific pathogens (n=83), which circulate exclusively in the human reservoir and are transmitted from person to person and hence are contagious, e.g. measles; (ii) zoonotic pathogens (n=152), which develop, mature, and reproduce entirely in non-human hosts but can still infect humans, who are then dead-end hosts, e.g. rabies; and (iii) multi-host pathogens (n=36), which can use both human and non-human hosts to complete their life-cycle, e.g. Ebola virus disease. We assigned pathogen transmission mode as follows: pathogens that spread through an arthropod vector (n=82) versus those not transmitted through a vector (n=189). Finally, we categorized pathogens by major taxonomic group: viruses, bacteria, fungi, protozoans, and helminths (including both helminth worms and nematodes).

Our geographic choices are driven by island biogeography theory. From 222 administrative territories recorded in GIDEON, we extracted data from 66 island countries and territories (Fig. 1.A), the largest being Madagascar and the smallest Tokelau (Appendix S1). Inclusion or exclusion of islands was driven by the completeness of information for a set of geographic, socioeconomic and demographic indicators based on previous, complementary work (Guégan & Broutin, 2009).

Since including all islands in this sample together could introduce confounding effects, such as those related to latitude, and because the equilibrium theory was originally elaborated for a group of islands within an archipelago, we extracted from the whole island dataset two regional island subsets, one for Caribbean islands (n=25, Fig. 1.B) and one for Pacific islands (n=21, Fig. 1.C). Territorial surface areas (in square-km) and total human population size were extracted from the 2010 *World Factbook*, published by the U.S. Central Intelligence Agency and updated yearly. ArcGIS software, version 9.3.1 (Esri, Redlands, CA, USA) was used to compute the centroid of each island and the distance, in km, from that centroid to the nearest mainland shoreline.

*Statistical analysis*

We used univariate linear regression models to investigate the relationship between the total number of pathogenic species (Species Richness, or SR) and both island surface area and distance from an island to the nearest mainland. SR and surface area variables were normalized by log-transformation. This linear relationship expressed in logarithmic space corresponds to the classical power model of the species-area relationship, generally expressed as $SR=cA^z$, where $A$ is the surface area, $c$ is the intercept, and $z$ in the linear coefficient, or slope (Triantis *et al.*, 2012). The relatively small sample sizes prevented a reasonable use of multivariate analysis. Linear regression provided the most simple, robust method to test for monotony in the predicted relationships between pathogen diversity and the variables of interest. Although non-linear models may have explained more of the variation in some of the studied relationships, a comparison of models and discussion of their potential underlying mechanisms processes are beyond the scope of this research.

The analysis was first conducted on the whole set of island pathogen species and then on this set broken down by (i) host-requirement (human-only, zoonotic, multi-host), (ii)

transmission mode (vector-borne, directly transmitted) and (iii) taxonomy (bacteria, virus, fungi, protozoans, helminths). First, we calculated the SR for each of these three breakdowns. For transmission mode, for instance, we calculated SR for vector-borne pathogens and SR for directly-transmitted pathogens. Second, we estimated the linear relationship between these SR values and our covariates of interest, island surface area and distance to the nearest mainland. Finally, we assessed differences among these linear relationships and our covariates of interest using a generalized analysis of covariance (ANCOVA). For example, we tested for statistical difference in the linear relationship between SR and surface area (or distance to mainland) between vector-borne and directly-transmitted pathogens.

In the case of human-specific pathogens, one could consider the ultimate area occupied by a pathogen species as defined by the host population size. In order to test this hypothesis, we conducted a complementary analysis using univariate linear regression models to investigate the relationship between pathogen SR and island human population (log-transformed), hypothesizing that any relationship for the larger sample would be driven by that for human-only pathogens and that the relationship would be strongest for obligate human pathogens.

Analyses were conducted on the whole island dataset and then on both regional sub-datasets. Analyses were conducted using R software v2.15.1 (R Development Core Team, 2005).

## Results

*Species richness relationships with area and distance in the entire sample*

Our findings for the entire sample of island countries and territories supported predictions from the equilibrium theory of island biogeography, though the effect of area on pathogen diversity was much more pronounced than that of distance. Fig. 2 presents the island SR plotted against, respectively, surface area (Fig. 2.A) and distance to the mainland (Fig. 2.B). Larger islands support more species of pathogens, as shown in Fig. 2.A ($y$ = $1.695 \times 10^{-2}$ $x$ + 2.022, $p<10^{-3}$). Island surface area explained more than 40% of the total variance of pathogen SR ($R^2_{adj}$=0.407). In turn, more-isolated islands tended to support fewer pathogen species, as shown in Fig. 2.B ($y$ = - $6.394 \times 10^{-6}$ $x$ + 2.087, p=0.014), though this relationship explains less than 10% of the total variance of SR ($R^2_{adj}$=0.0766).

*Relationships between SR and host requirement, transmission pathway, and taxonomy*

Across all pathogen subcategories, SR increased with island surface area and decreased with distance to the nearest mainland. However, as Fig. 3 shows, the extent of these relationships, as indicated by differences among regression slopes, is driven by zoonotic status, vectorial transmission, and protozoan and helminthian taxonomy. Pathogens that infect humans obligately, those that do not require a vector for transmission, and those that are relatively small (viruses, bacteria) are affected much less by island biogeography.

The positive relationship between SR and surface area was significant for every pathogen subcategory (each $p<10^{-3}$, Table 1). However, as presented in Table 1, the strength of this relationship varied significantly across pathogen host-requirement categories (slope coefficients, Human Only pathogens: $5.768\text{x}10^{-3}$, Multi-Host pathogens: $1.210\text{x}10^{-2}$, Zoonotic pathogens: $3.788\text{x}10^{-2}$; ; ANCOVA-$p<10^{-3}$), transmission pathways (slope coefficients, Directly Transmitted pathogens: 0.0132, Vector-borne pathogens: 0.0469; ANCOVA-$p<10^{-3}$), and taxonomic categories (slope coefficients, Bacteria: $8.932\text{x}10^{-3}$, Viruses: $1.486\text{x}10^{-2}$, Fungi: $1.522\text{x}10^{-2}$, Protozoans: $3.733\text{x}10^{-2}$, Helminths: $2.416\ \text{x}10^{-2}$; ANCOVA-$p<10^{-3}$).

The negative relationship between SR and distance to the nearest mainland was significant or at the limit of significance for nine of the ten categories we considered (for 5 categories: $p<0.05$; for 4 categories: $p<0.10$; Table 1). The strength of this relationship varied significantly among pathogen transmission pathway categories (slope coefficients for Directly Transmitted and Vector-borne pathogens, respectively: $-4.380.10^{-6}$ and $-2.310.10^{-6}$; ANCOVA-$p=0.0343$).

*Complementary analysis on regional sub-datasets*

As for the dataset as a whole, we found that larger islands supported greater pathogen diversity in the Caribbean and Pacific subsets ($p<10^{-3}$and $p=0.001$, respectively). However, a significant negative relationship between SR and distance to the nearest mainland was only observed for the Pacific islands ($p=0.02$).

The effect of island size was driven by zoonotic and vector-borne pathogens in both Caribbean and Pacific islands and, for Pacific islands only, by protozoans and helminths (Table 2). For both Caribbean and Pacific islands, we did not find significant differences across pathogens categories in the relationship between SR and distance to the nearest mainland.

*Species richness relationships with human population size and density*

Although results for the entire sample support the hypothesized positive effect of human population on SR, the relationship was not driven by human-only pathogens (slope coefficients for Human Only, Multi-Host and Zoonotic pathogens, respectively: $6.583\text{x}10^{-3}$; $1.737\text{x}10^{-2}$ and $3.536\text{x}10^{-2}$; ANCOVA-$p<10^{-3}$).

Together, our findings suggest the area of an island is more important than the population size of potential human hosts living there. Larger islands support more people ($r$ = 0.767, and more people support more species of pathogens ($y = 1.742\text{x}10^{-2}\ x + 1.986$, $R^2_{adj}=0.538$, $p<10^{-3}$) (see Appendix S2). However, this relationship is largely a function of the relationship for more-populous island nations, corresponding to a “break” in the regression at a population of $\sim 10^5$ and thus perhaps reflecting a threshold of urbanization or more-general intensification. Tellingly, though, the relationship between human population density and pathogen SR is relatively smooth and weak ($y = 8.769\text{x}10^{-3}\ x + 2.060$, $R^2_{adj}=0.045$, $p=0.049$) (see Appendix S2), suggesting that human population size and pathogen SR are both responding to factors that vary with island area, such as environmental energy supply or the diversity of potential habitats.

## Discussion

We have shown here that the distribution of known human pathogens on islands follows the main predictions of MacArthur and Wilson's equilibrium theory of island biogeography (MacArthur & Wilson, 1967): pathogen species richness increases with island area and decreases with distance to the nearest mainland. However, the relative influence of area is much greater than that of isolation, and the extent and strength of the associations vary by host requirement, transmission pathway, and pathogen taxonomy. Importantly, pathogens whose primary hosts are not humans are more strongly affected by island biogeography than are those that primarily afflict people.

A limitation of this study is the relatively small sample size, an inherent constraint of focusing on a relatively small subset of the larger GIDEON dataset. The resulting lack of statistical power did not allow us to account simultaneously for different categorical factors or to take into account other factors previously identified as important drivers of pathogen richness, such as climate (Guernier *et al.*, 2004; Dunn *et al.*, 2010). However, our purpose was not to identify and assess the relative influence of a large set of variables but rather to test how well an influential biogeographical theory describes a pattern of contemporary human ecology. Conducting a complementary analysis on regional sub-datasets (Caribbean and Pacific islands) was a way to control for shared characteristics of islands from the larger sample, such as latitude and regional biotic influences. The fact that the results of these regional analyses were similar to those for the whole dataset supports the validity of the relationships we found.

Another limitation is that GIDEON is an evidence-based database, so the data could, potentially, reflect a reporting bias. Indeed, wider sampling or research efforts on larger or less-isolated islands could contribute to the results described here. Hypothetically, although such a reporting bias for this island dataset could influence our findings, it is unlikely this bias would produce the patterns we observed across pathogen categories. Furthermore, healthcare expenditure is a poor predictor of human pathogen SR at the country-level (Dunn *et al.*, 2010) even if it does predict infectious pathogen prevalence. Thus, our results are likely independent of any reporting effect.

According to the equilibrium theory of island biogeography, the positive relationship we found between human pathogen species richness and island area is due to lower extinction rates on larger islands. Larger islands contain larger habitat areas and a likelihood of greater habitat diversity. To human pathogens, larger habitat areas should support more host individuals, including more humans, and greater habitat diversity should support more species of alternative hosts. Indeed, nations with more people and more species of birds and mammals support more species of human pathogens (Dunn *et al.*, 2010), and habitat diversity drives the diversity of bacterial assemblages generally (Nemergut *et al.*, 2011). So the patterns we observed for human pathogens could simply mirror: (i) a species-human population relationship, in which the human population serves as the ''area'' that pathogen species occupy, and/or (ii) the species-area relationships for alternative insular host species.

However, the much greater effect of area on the species richness of zoonotics than on multi-host and human-only pathogens (Fig 2. A) suggests a strong mechanistic role for alternative insular hosts. In short, island animals for which pathogens are primary hosts help

drive the island biogeography of human disease. This is not surprising, because these animals are much more restricted in their ability to travel than are contemporary humans. This finding also corresponds with the strong relationship between pathogen diversity and bird and mammal species diversity (Dunn *et al.*, 2010) and the findings of a broader analysis that included continental nations (Smith *et al.*, 2007).

Greater immigration rates also likely moderate the effect of area on human-only pathogens through the rescue effect, a refinement of MacArthur and Wilson's theory. The rescue effect (Brown & Kodric-Brown, 1977) is the effect of immigration on extinction, which in the original theory was solely a function of island size. By continually bringing pathogens with us to islands when we migrate or travel, we "rescue" some pathogen species that might otherwise die out on smaller islands (i.e. when the number of susceptible individuals drops below a threshold necessary to sustain the disease, i.e. the Critical Community Size) (Rohani *et al.*, 1999).

Together, these features specific to human pathogens differentiate island microbes from other insular taxa. A positive relationship between island size and bacterial diversity holds in engineered systems (van der Gast et al., 2005), supporting the generality of this biogeographic pattern and the importance of species-area/volume effects among microbes generally (Green & Bohannan, 2006). Yet mass-related limitations on active dispersal and exceptional rates of diversification (Martiny et al., 2006), may explain the exceptionally low effect of area we observed: the value of the coefficient linking species richness and island surface (z-value) for insular human parasites and pathogens was about an order of magnitude lower than those of insular macro-organisms (Table 3).

The negative relationship we found between pathogen species richness and distance to the mainland is a function of varying immigration rates in MacArthur and Wilson's theory. Several factors may explain the relative weakness we found in the influence of isolation versus that of area. In the original theory, distance is that separating different islands to the same continental shore, viewed as the source of the same set of species. However, species considered here are pathogens of *Homo sapiens*, whose large-scale movement capacities have increased continuously, especially during the past five centuries (Smith & Guégan, 2010). This increase in connectedness, which has profoundly lessened effective isolation, likely explains the much smaller effect of distance. One way to test this idea would be to explore the relationship between island pathogen diversity and transport connectedness (Colizza *et al.*, 2006).

The finding of a greater role for area than for human host population and for distance can be seen as supporting the Baas-Becking' hypothesis, which posits that, regarding microbes, "everything is everywhere, but the environment selects" (Baas-Becking, 1934) and a corresponding non-stochastic view of microbial community assembly (Barberán *et al.*, 2014). In the case of zoonotics, the environment of interest is composed of non-human hosts and their habitats are not everywhere.

This work has several practical, broad-scale implications. The fact that our travel and trade swamps isolation so profoundly means there will remain few disease-free islands. However, the importance of area in supporting populations of vectors and alternative hosts means the proportion of island diseases that are zoonotic and vector-transmitted will tend to

decline with decreasing island size, with implications for public health management efforts. These implications should apply to current islands as well as to those, given time for equilibration, created or whose area or isolation is altered by rising sea levels. However, concluding that reducing or fragmentizing habitats is a viable public health strategy would misinterpret the broader lessons of ecology in the Anthropocene. The incursion of human populations into natural habitats is already associated with zoonotic outbreaks and emergence, and habitat loss would suppress biodiversity more broadly and have a disproportionate impact on larger taxa, such as mammals. As global biodiversity benefits human health and well-being in many ways, such a strategy would harm more than help.

Our results demonstrate how classic island biogeography theory applies to human pathogens, and our findings support the spirit of theoretical insight as much as the substance. Even if infectious diseases have been widely globalized because of large-scale human movements, area and isolation still affect macroscopic disease patterns. And the ways in which the results seem to show weak support—in the relative effect of isolation on human-only pathogens—highlights the importance of the underlying process, immigration, which is so strongly constrained by isolation. Globalization effectively increases pathogen immigration rates, reducing the historic barrier of isolation. Just as humans both follow and flout ecological patterns common to species generally (Burnside *et al.*, 2012), so do the parasites and pathogens that afflict us. And just as some aspects of biogeography are common to life generally, others may be unique to microbes. As it did for our understanding of assemblages of plant and animal communities, we hope that this test of “equilibrium theory” will be a stepping stone in the understanding of causal drivers behind global trends in human infectious disease and in the broader quest to understand the geography of life.

## ACKNOWLEDGEMENT

We are thankful to François Guilhaumon from Institut de Recherche pour le Développement (IRD) and Dov Sax from Brown University, for useful comments on an earlier draft. This research has benefited from an “Investissement d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA ANR-LABX-10-2501). In addition, JFG received support from both IRD and CNRS.

## REFERENCES

Baas-Becking, L. (1934) *Geobiologie of inleiding tot de milieukunde*, Van Stockum & Zoon. The Hague, the Netherlands.

Barberán, A., Casamayor, E.O. & Fierer, N. (2014) The microbial contribution to macroecology. *Frontiers in Microbiology*, **5**, 203.

Bell, T., Ager, D., Song, J.I., Newman, J.A., Thompson, I.P., Lilley, A.K. & van der Gast, C.J. (2005) Larger islands house more bacterial taxa. *Science*, **308**, 1884–1884.

Bjørnstad, O.N. & Harvill, E.T. (2005) Evolution and emergence of Bordetella in humans. *Trends in Microbiology*, **13**, 355–359.

Bloom, D.E. & Sachs, J.D. (1998) Geography, demography, and economic growth in Africa. *Brookings Papers on Economic Activity*, 207–295.

Bonds, M.H., Dobson, A.P. & Keenan, D.C. (2012) Disease ecology, biodiversity, and the latitudinal gradient in income. *PLoS biology*, **10**, e1001456.

Brown, J. & Kodric-Brown, A. (1977) Turnover Rates in Insular Biogeography - Effect of Immigration on Extinction. *Ecology*, **58**, 445–449.

Burnside, W.R., Brown, J.H., Burger, O., Hamilton, M.J., Moses, M. & Bettencourt, L.M.A. (2012) Human macroecology: linking pattern and process in big-picture human ecology. *Biological reviews of the Cambridge Philosophical Society*, **87**, 194–208.

Cashdan, E. (2014) Biogeography of Human Infectious Diseases: A Global Historical Analysis. *PLoS ONE*, **9**, e106752.

Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 2015–2020.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races In the Struggle for Life*, John Murray Publisher Ltd., London, UK.

Dunn, R.R., Davies, T.J., Harris, N.C. & Gavin, M.C. (2010) Global drivers of human pathogen richness and prevalence. *Proceedings. Biological sciences / The Royal Society*, **277**, 2587–2595.

Furuse, Y., Suzuki, A. & Oshitani, H. (2010) Origin of measles virus: divergence from rinderpest virus between the 11th and 12th centuries. *Virology Journal*, **7**, 52.

Van der Gast, C.J., Lilley, A.K., Ager, D. & Thompson, I.P. (2005) Island size and bacterial diversity in an archipelago of engineering machines. *Environmental Microbiology*, **7**, 1220–1226.

Green, J. & Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution*, **21**, 501–507.

Guégan, J.-F. & Broutin, H. (2009) *Microbial Communities: Patterns and Processes*. *Biodiversity Change and Human Health: From Ecosystem Services to Spread of Disease* Scientific Committee on Problems of the Environment (SCOPE) Series., pp. pp.193–210. Island Press; 1 edition.

Guernier, V., Hochberg, M.E. & Guégan, J.-F. (2004) Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, **2**, e141.

Hay, S.I., Battle, K.E., Pigott, D.M., Smith, D.L., Moyes, C.L., Bhatt, S., Brownstein, J.S., Collier, N., Myers, M.F., George, D.B. & Gething, P.W. (2013) Global mapping of infectious disease. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **368**, 20120250.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Just, M.G., Norton, J.F., Traud, A.L., Antonelli, T., Poteate, A.S., Backus, G.A., Snyder-Beattie, A., Sanders, R.W. & Dunn, R.R. (2014) Global biogeographic regions in a human-dominated world: the case of human diseases. *Ecosphere*, **5**, art143.

Lomolino, M.V., Riddle, B.R., Wittaker, R.J. & Brown, J.H. (2010) *Biogeography (Fourth Edition)*, Sinauer Associates, Inc.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., Alvarado, M., Anderson, H.R., Anderson, L.M., Andrews, K.G., Atkinson, C., Baddour, L.M., Barker-Collo, S., Bartels, D.H., Bell, M.L., Benjamin, E.J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S.S., Coffeng, L.E., Colan, S.D., Colquhoun, S., Colson, K.E., Condon, J., Connor, M.D., Cooper, L.T., Corriere, M., Cortinovis, M., de Vaccaro, K.C., Couser, W., Cowie, B.C., Criqui, M.H., Cross, M., Dabhadkar, K.C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D.C., Dharmaratne, S.D., Dorsey, E.R., Driscoll, T., Duber, H., Ebel, B., Erwin, P.J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A.D., Forouzanfar, M.H., Fowkes, F.G.R., Franklin, R., Fransen, M., Freeman, M.K., Gabriel, S.E., Gakidou, E., Gaspari, F., Gillum, R.F., Gonzalez-Medina, D., Halasa, Y.A., Haring, D., Harrison, J.E., Havmoeller, R., Hay, R.J., Hoen, B., Hotez, P.J., Hoy, D., Jacobsen, K.H., James, S.L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J.-P., Knowlton, L.M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S.E., Ohno, S.L., Mabweijano, J., MacIntyre, M.F., Mallinger, L., March, L., Marks, G.B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B.M., McAnulty, J.H., McDermott, M.M., McGrath, J., Mensah, G.A., Merriman, T.R., Michaud, C., Miller, M., Miller, T.R., Mock, C., Mocumbi, A.O., Mokdad, A.A., Moran, A., Mulholland, K., Nair, M.N., Naldi, L., Narayan, K.M.V., Nasseri, K., Norman, P., O'Donnell, M., Omer, S.B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J.D., Rivero, A.P., Padilla, R.P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope, C.A., 3rd, Porrini, E., Pourmalek, F., Raju, M., Ranganathan, D., Rehm, J.T., Rein, D.B., Remuzzi, G., Rivara, F.P., Roberts, T., De León, F.R., Rosenfeld, L.C., Rushton, L., Sacco, R.L., Salomon, J.A., Sampson, U., Sanman, E., Schwebel, D.C., Segui-Gomez, M., Shepard, D.S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J.A., Thomas, B., Tleyjeh, I.M., Towbin, J.A., Truelsen, T., Undurraga, E.A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G.R., Wang, M., Wang, W., Watt, K., Weinstock, M.A., Weintraub, R., Wilkinson, J.D., Woolf, A.D., Wulf, S., Yeh, P.-H., Yip, P., Zabetian, A., Zheng, Z.-J., Lopez, A.D., Murray, C.J.L., AlMazroa, M.A. & Memish, Z.A. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**, 2095–2128.

MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*, Princeton University Press.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreås, L., Reysenbach, A.-L., Smith, V.H. & Staley, J.T. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews. Microbiology*, **4**, 102–112.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L. & Knight, R. (2011) Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology*, **13**, 135–144.

Orrock, J.L., Allan, B.F. & Drost, C.A. (2011) Biogeographic and Ecological Regulation of Disease: Prevalence of Sin Nombre Virus in Island Mice Is Related to Island Area, Precipitation, and Predator Richness. *American Naturalist*, **177**, 691–697.

Peay, K.G., Bruns, T.D., Kennedy, P.G., Bergemann, S.E. & Garbelotto, M. (2007) A strong species-area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology Letters*, **10**, 470–480.

Peterson, A.T. (2008) Biogeography of diseases: a framework for analysis. *Die Naturwissenschaften*, **95**, 483–491.

R Development Core Team (2005) *R: a language and environment for statistical computing*, R Foundation for Statistical Computing. Vienna, Austria.

Rohani, P., Earn, D.J. & Grenfell, B.T. (1999) Opposite patterns of synchrony in sympatric disease metapopulations. *Science (New York, N.Y.)*, **286**, 968–971.

Smith, K.F. & Guégan, J.-F. (2010) *Changing Geographic Distributions of Human Pathogens*. *Annual Review of Ecology, Evolution, and Systematics, Vol 41* (ed. by D.J. Futuyma, H.B. Shafer, and D. Simberloff), pp. 231–250. Annual Reviews, Palo Alto.
Smith, K.F., Sax, D.F., Gaines, S.D., Guernier, V. & Guégan, J.-F. (2007) Globalization of human infectious disease. *Ecology*, **88**, 1903–1910.
Smith, V.H., Foster, B.L., Grover, J.P., Holt, R.D., Leibold, M.A. & Denoyelles, F. (2005) Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4393–4396.
Svensson-Coelho, M. & Ricklefs, R.E. (2011) Host phylogeography and beta diversity in avian haemosporidian (Plasmodiidae) assemblages of the Lesser Antilles. *Journal of Animal Ecology*, **80**, 938–946.
Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215–231.
Wolfe, N.D., Dunavan, C.P. & Diamond, J. (2007) Origins of major human infectious diseases. *Nature*, **447**, 279–283.

## BIOSKETCH

Kévin Jean holds a PhD in Epidemiology and was initially trained in Ecology and Evolution. His work focuses on epidemiology and prevention of infectious diseases, with a constant effort to address comprehensively the biological, ecological, behavioral and social determinants of infectious transmission. For more information: https://sites.google.com/site/kvnjean/home

William Burnside, a postdoctoral fellow at the National Socio-Environmental Synthesis Center (SESYNC), applies methods from macroecology and functional ecology to test the ability of ecological theory to inform our understanding of human-environment systems.

Jean-François Guégan is a senior research scientist at the French Institute for Research in Developing Countries (ww.ird.fr) and also a scientific adviser for the ecoHEALTH initiative from the international programme FutureEarth from the U.N. As a disease ecologist, his research interests focus on macroecology of infectious diseases and their hosts, and the links between climate change and biodiversity and emerging infections.

## SUPPORTING INFORMATION

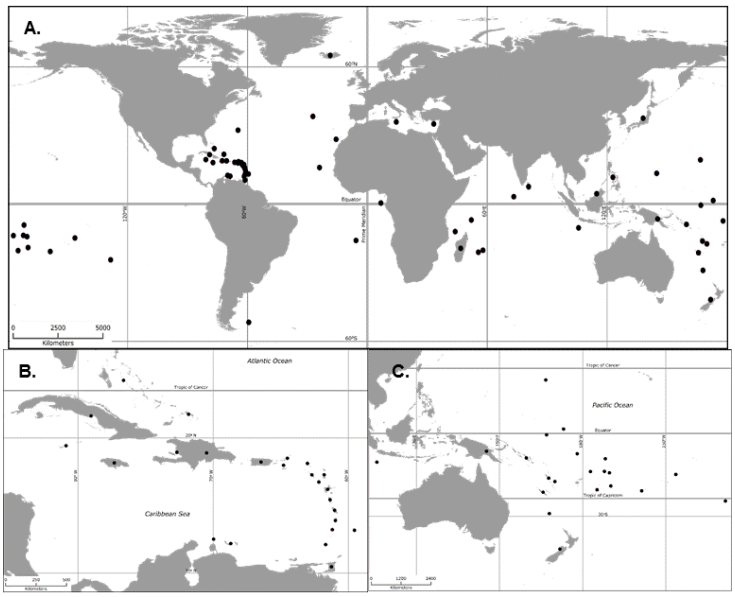
**Appendix S1:** Geographical characteristics of the 66 islands nations considered in the analysis.

**Appendix S2:** Species richness relationships with human population size and density.

**Appendix S3:** Reproduced graphs with countries labeled.

549 **Tables and Figures**

550 **Fig. 1:** Geographic location of the islands considered: **A)** whole dataset (n=66), **B)** Caribbean
551 dataset (n=24) and **C)** Pacific dataset (n=21).




552

553

**Fig. 2:** Pathogen species richness (log number of species) plotted against: **A)** island surface area (log km²) and **B)** distance to the nearest mainland (km). Linear regression parameters: **A)** $y = 1.695\text{x}10^{-2}x + 2.022$, $R^2_{adj}=0.407$, $p<0.0001$; **B)** $y = -6.394\text{x}10^{-6}x + 2.087$, $R^2_{adj}=0.0766$, $p=0.014$. Total pathogen species considered: n=271. Note that the influence of area is much stronger than that of distance ($|1.695\text{x}10^{-2}| >> |-6.394\text{x}10^{-6}|$). Appendix S3 includes these same graphs with the countries labeled.

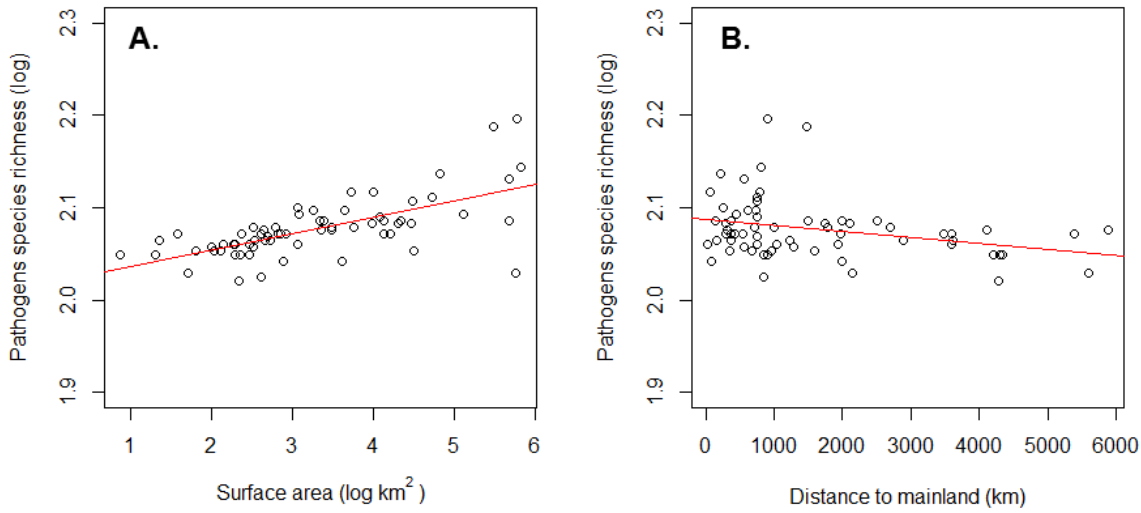


**Fig. 3:** Pathogen species richness (log number of species) as a function of island surface area (left) and distance to mainland (right) classified by host requirement (**A**, **B**), transmission pathway (**C**, **D**), and taxonomy **(E, F)**.

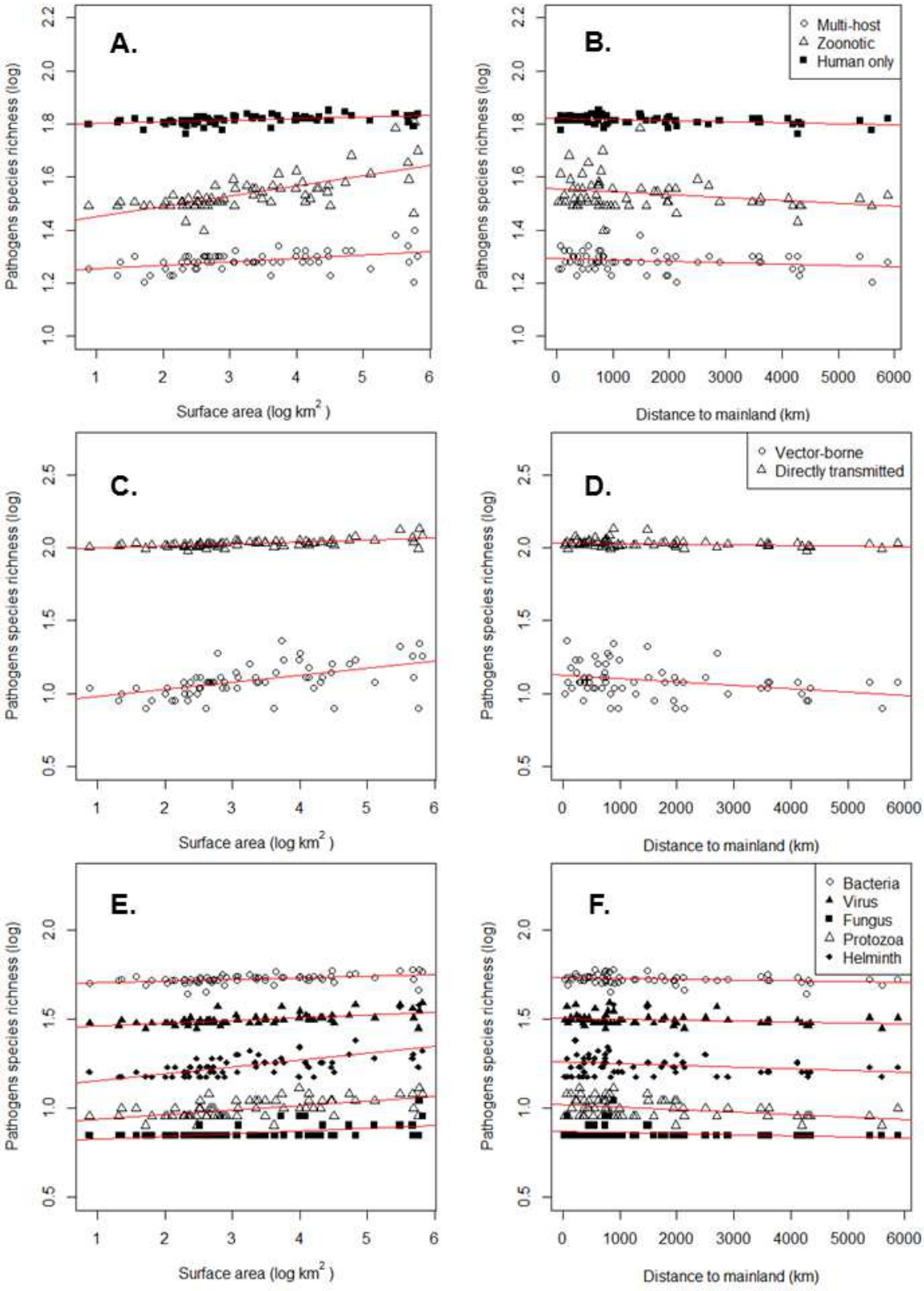


**Table 1: Results of univariate linear regressions of log number of pathogen species classified by host-requirement, transmission pathway, and taxonomy, as functions of a) island surface area (log) and b) distance to mainland in the total island sample (n=66).**

**NB: The linear relationship between Species Richness and Island surface area expressed in logarithmic space corresponds to the classical power model of the species-area relationship, generally expressed as $SR=cA^z$, where $A$ is the surface area, $c$ is the intercept, and $z$ in the linear coefficient, or slope.**

| | a) Island surface area (log km²) | | | | | b) Distance to Mainland (km) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Pathogen Species Richness classified by:* | Slope ($x10^{-2}$) | Intercept | $R^2_{adj}$ | p | ANCOVA p | Slope ($x10^{-6}$) | Intercept | $R^2_{adj}$ | p | ANCOVA p |
| **Host-requirement** | | | | | $<10^{-3}$ | | | | | 0.354 |
| Multi-host | 1.21 | 1.25 | 0.178 | $<10^{-3}$ | | -5.00 | 1.29 | 0.032 | 0.081 | |
| Zoonotic | 3.79 | 1.42 | 0.449 | $<10^{-3}$ | | -10.99 | 1.56 | 0.043 | 0.051 | |
| Human only | 0.58 | 1.80 | 0.185 | $<10^{-3}$ | | -4.10 | 1.82 | 0.13 | 0.002 | |
| **Transmission pathway** | | | | | $<10^{-3}$ | | | | | 0.034 |
| Vector-borne | 4.69 | 0.94 | 0.277 | $<10^{-3}$ | | -23.10 | 1.13 | 0.089 | 0.009 | |
| Directly transmitted | 1.32 | 1.99 | 0.392 | $<10^{-3}$ | | -4.38 | 2.04 | 0.051 | 0.038 | |
| **Taxonomy** | | | | | $<10^{-3}$ | | | | | 0.343 |
| Bacteria | 0.89 | 1.70 | 0.187 | $<10^{-3}$ | | -4.62 | 1.73 | 0.063 | 0.024 | |
| Viruses | 1.49 | 1.45 | 0.337 | $<10^{-3}$ | | -4.72 | 1.50 | 0.037 | 0.065 | |
| Fungi | 1.52 | 0.81 | 0.241 | $<10^{-3}$ | | -5.79 | 0.87 | 0.039 | 0.061 | |
| Protozoans | 3.73 | 1.12 | 0.363 | $<10^{-3}$ | | -13.88 | 1.02 | 0.154 | 0.001 | |
| Helminths | 2.42 | 0.92 | 0.329 | $<10^{-3}$ | | -9.95 | 1.26 | 0.025 | 0.110 | |

**Table 2: Results of univariate linear regressions of log number of pathogen species classified by host-requirement, transmission pathway, and taxonomy as functions of a) island surface area (log) and b) distance to the nearest mainland for two regional island subsets, Caribbean islands (n=24) and Pacific islands (n=21).**

**NB: The linear relationship between Species Richness and Island surface area expressed in logarithmic space corresponds to the classical power model of the species-area relationship, generally expressed as $SR=cA^z$, where $A$ is the surface area, $c$ is the intercept, and $z$ in the linear coefficient, or slope.**

| | a) Island surface area (log km²) | | | | | b) Distance to Mainland | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Slope ($x10^{-2}$) | Intercept | $R^2$ | p | **ANCOVA p** | Slope ($x10^{-6}$) | Intercept | $R^2$ | p | **ANCOVA p** |
| **Caribbean Islands (n=24)** | | | | | | | | | | |
| **All pathogens** | 2.12 | 2.01 | 0.54 | $<10^{-3}$ | - | -5.19 | 2.08 | 0.00 | 0.782 | **-** |
| *Pathogen Species Richness classified by:* | | | | | | | | | | |
| **Host-requirement** | | | | | **$<10^{-3}$** | | | | | **0.797** |
| Multi-host | 0.97 | 1.26 | 0.11 | 0.112 | | 1.68 | 1.29 | 0.00 | 0.93 | |
| Zoonotic | 4.51 | 1.38 | 0.56 | $<10^{-3}$ | | -20.6 | 1.54 | 0.01 | 0.60 | |
| Human only | 1.20 | 1.78 | 0.35 | 0.002 | | 0.21 | 1.82 | 0.00 | 0.99 | |
| **Transmission pathway** | | | | | **0.038** | | | | | **0.581** |
| Vector-borne | 6.03 | 0.91 | 0.28 | 0.007 | | -40.8 | 1.12 | 0.01 | 0.58 | |
| Directly transmitted | 1.62 | 1.97 | 0.61 | $<10^{-3}$ | | 0.54 | 2.02 | 0.00 | 0.97 | |
| **Taxonomy** | | | | | **0.446** | | | | | **0.707** |
| Bacteria | 1.66 | 1.67 | 0.36 | 0.002 | | -11.5 | 1.73 | 0.02 | 0.52 | |
| Viruses | 1.93 | 1.43 | 0.40 | $<10^{-3}$ | | -13.5 | 1.50 | 0.02 | 0.50 | |
| Fungi | 2.50 | 0.79 | 0.29 | 0.006 | | -2.94 | 0.87 | 0.00 | 0.92 | |
| Protozoans | 2.92 | 1.15 | 0.22 | 0.020 | | 35.1 | 1.22 | 0.03 | 0.39 | |
| Helminths | 3.51 | 0.88 | 0.54 | $<10^{-3}$ | | -15.4 | 1.00 | 0.01 | 0.62 | |
| **Pacific Islands (n=21)** | | | | | | | | | | |
| **All pathogens** | 1.21 | 2.03 | 0.56 | $<10^{-3}$ | - | -7.10 | 2.09 | 0.24 | 0.02 | **-** |
| *Pathogen Species Richness classified by:* | | | | | | | | | | |
| **Host-requirement** | | | | | **0.001** | | | | | **0.352** |
| Multi-host | 1.39 | 1.24 | 0.34 | 0.006 | | -4.87 | 1.29 | <0.01 | 0.324 | |
| Zoonotic | 2.53 | 1.45 | 0.64 | $<10^{-3}$ | | -12.80 | 1.56 | 0.20 | 0.041 | |
| Human only | 0.43 | 1.80 | 0.16 | 0.078 | | -4.56 | 1.82 | 0.22 | 0.033 | |
| **Transmission pathway** | | | | | **$<10^{-3}$** | | | | | **0.246** |
| Vector-borne | 4.49 | 0.92 | 0.61 | $<10^{-3}$ | | -18.90 | 1.11 | 0.13 | 0.103 | |
| Directly transmitted | 0.84 | 2.00 | 0.46 | $<10^{-3}$ | | -5.65 | 2.04 | 0.26 | 0.019 | |
| **Taxonomy** | | | | | **$<10^{-3}$** | | | | | **0.191** |
| Bacteria | 0.79 | 1.70 | 0.19 | 0.046 | | -6.74 | 1.74 | 0.18 | 0.058 | |
| Viruses | 1.16 | 1.46 | 0.54 | $<10^{-3}$ | | -5.97 | 1.51 | 0.18 | 0.056 | |
| Fungi | 0.62 | 0.83 | 0.25 | 0.021 | | -5.10 | 0.87 | 0.21 | 0.035 | |
| Protozoans | 2.26 | 1.15 | 0.80 | $<10^{-3}$ | | -5.50 | 1.24 | 0.06 | 0.289 | |
| Helminths | 2.21 | 0.92 | 0.47 | $<10^{-3}$ | | -16.90 | 1.04 | 0.34 | 0.005 | |

**Table 3 : Summary of *z*-values documented among different organisms for Island Species-Area Relationships**.

***z* corresponds to the coefficient of the power model $SR=cA^z$, where SR is the species richness and A the surface area. Equivalently, *z* corresponds to the slope of the log-linear relationship linking SR and A.**

| Organisms | *z*-value | reference |
|---|---|---|
| Plants | 0.355 | Triantis *et al.*, 2012 |
| Invertebrates | 0.323 | Triantis *et al.*, 2012 |
| Vertebrates | 0.284 | Triantis *et al.*, 2012 |
| Bacteria | 0.104 - 0.295 | Green & Bohannan, 2006 |
| Phytoplankton | 0.134 | Smith *et al.*, 2005 |
| Fungi | 0.20 - 0.23 | Peay *et al.*, 2007 |
| | | |
| Human parasites and pathogens | | *Present study* |
| **Overall** | **0.017** | |
| Bacteria | 0.009 | |
| Viruses | 0.015 | |
| Fungi | 0.015 | |
| Protozoans | 0.037 | |
| Helminths | 0.024 | |