

# The coevolutionary mosaic of bat-betacoronaviruses spillover risk

Norma Forero Rocio Munoz<sup>1,2</sup> Renata L. Muylaert<sup>3</sup> Stephanie N. Seifert<sup>4</sup> Gregory F. Albery<sup>5</sup> Colin J. Carlson<sup>5</sup> **Timothée Poisot**<sup>1,2,‡</sup>

<sup>1</sup> Université de Montréal <sup>2</sup> Québec Centre for Biodiversity Sciences <sup>3</sup> Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand

<sup>4</sup> Paul G. Allen School for Global Health, Washington State University, Pullman, WA, United States <sup>5</sup> ???

<sup>‡</sup> These authors contributed equally to the work

## Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Driven by the need to understand the ecological factors involved in the emergence of betacoronavirus (the genus causing the SARS and MERS disease in human) through bat hosts, we develop an approach to the assessment of spillover risk based on the Geographic Mosaic Theory of Coevolution. In doing so, we provide a global mapping of the spillover risk posed by betacoronaviruses, reflecting the fact that this risk has many ecological and evolutionary origins. Our framework reveals that reservoir richness alone, although a component of viral hazard, is not a sufficiently integrative predictor of risk. We offer alternative insights based on viral sharing, host compositional uniqueness, and host phylogenetic diversity. By comparing our aggregated measure of risk to a proxy for human density, namely the proportion of each pixel that is covered by urban or built land, we provide a synthetic risk map, allowing the identification of hotspots where the bat-betacoronavirus system may originate novel viruses in humans.

1 Spillover risk is multidimensional and complicated. From the standpoint of an animal community, *i.e.* a  
2 pool of suitable hosts, it is driven by a multiplicity of factors (Plowright et al. 2017). Proxies for the global  
3 richness of hosts is one such component commonly analysed (see *e.g.* Anthony et al. 2017 for  
4 coronaviruses), but there is an argument to be made that species who are not competent (or known) hosts  
5 of a specific virus genus may not factor into this (Plowright et al. 2015); this is especially true as  
6 competence data increases predictive power when the taxonomic scope of hosts of a viral family increases  
7 (**Becker2020InfInt?**, **Mull2022VirIso?**). Similarly, host species who are assumed to share viruses at  
8 different rates should be weighted accordingly (Albery et al. 2020). In mammals, key functional traits (for  
9 which phylogeny is a reasonable proxy) are determinants of the spillover potential (Olival et al. 2017);  
10 these include, notably, body mass and affinity for urban environments (Albery et al. 2022). Finally,  
11 especially when the pool of potential hosts spans the entire globe, there may be local host pools that are  
12 highly unique; not having been observed in other locations, these can act on the overall risk either by  
13 providing novel contact opportunities, reflecting unique host-environment combinations (Engering et al.  
14 2013), or facilitating rapid evolutionary changes in specialism of their pathogens (Agosta et al. 2010). In  
15 the specific case of generalist pathogens (which is the case many viruses in the betacoronavirus genus, see  
16 *e.g.* **MacLean2021NatSel?**), there is conceptual and empirical support to the idea that these  
17 community-level mechanisms are even more important in driving the overall risk (Power and Mitchell  
18 2004).

19 Bats are important reservoir hosts for different classes of microorganisms (Chu 2008, Donaldson 2010, Li  
20 2010), many of which can threaten human health (**Letko2020BatVir?**, **VanBrussel2022ZooDis?**).  
21 Chiropterans emerged around 64 million years ago and are one of the most diverse mammalian orders,  
22 with an estimated richness of more than 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020).  
23 They exhibit a broad variety of habitat use, behaviour, and feeding strategies, putting them at key positions  
24 in the delivery and provisioning of several ecosystem services, tied to important ecosystem-derived  
25 benefits to human (Kasso and Balakrishnan 2013). For example, bats are an essential component of many  
26 seed-dispersal networks (**Mello2011MisPar?**). Over two-thirds of bats are known to be either obligate or  
27 facultative insectivores, therefore playing an important role in the regulation of insect pests that can affect  
28 crops (Williams-Guillén et al. 2008, Voigt and Kingston 2016), and vectors of pathogens that put a risk on  
29 human health (Gonsalves et al. 2013a, b). Because bats are globally distributed and have a long  
30 evolutionary history, phylogeographic and biogeographic approaches are required to shed light on the

31 contemporary distribution of coevolutionary processes between bats and the pathogens they host. Not all  
32 areas in which bats, viruses, and human are co-occurring are facing a risk of spillover towards human  
33 populations, and the areas in which this risk exist may not be facing risks of the same nature and  
34 magnitude.

35 In this paper, we examine the biogeographic structure of bats-betacoronaviruses associations, based on a  
36 curated dataset of all confirmed bat hosts of betacoronaviruses. This work is important both as a  
37 description of the bats-betacoronavirus complex, but also because more broadly, bats are known reservoirs  
38 for a variety of emerging viruses and other pathogens (Calisher et al. 2006, Melaun et al. 2014), making  
39 balancing the needs for bat conservation and disease prevention most likely very difficult and a source of  
40 human-wildlife conflicts, especially in more densely populated areas (Stone et al. 2015, Rego et al. 2015).  
41 By drawing on concepts from the Geographic Mosaic Theory of Coevolution (Thompson 2005), we turn  
42 these associations into a spatially explicit additive mapping of zoonotic risk components, revealing the  
43 extreme heterogeneity of risk at the global scale. We identify the Amazon and South-Eastern Asia as  
44 hotspots where the phylogenetic distinctiveness of betacoronaviruses is the highest (Anthony et al. 2017);  
45 surprisingly, current data suggest that viral sharing between hosts is high in the Amazon and low in  
46 South-Eastern Asia, which has the potential to result in different evolutionary dynamics between these  
47 two regions.

## 48 **Methods**

### 49 **Known betacoronavirus hosts**

50 We downloaded the data on bats hosts of betacoronaviruses assembled by Becker et al. (2022) from  
51 <https://www.viralemergence.org/betacov> on Apr. 2022, and filtered it to “known” hosts (established  
52 before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling since the  
53 emergence of SARS-CoV-2). The original database was assembled by a combination of data mining and  
54 literature surveys, including automated alerts on the “bats” and “coronavirus” keywords to identify novel  
55 empirical evidence of bats-betacoronaviruses associations.

56 **Bats occurrences**

57 We downloaded the rangemap of every extant bat species that was either classified as an empirically  
58 documented host of beta-coronaviruses from the previous step, according to recent IUCN data (IUCN  
59 2021). The range maps were subsequently rasterized using the `rasterize` function from GDAL (Rouault et  
60 al. 2022) at a resolution of approximately **TK TP**. For every pixel in the resulting raster where at least one  
61 bat host of betacoronavirus was present, we extract the species pool (list of all bat species), which was used  
62 to calculate the following risk assessment components: phylogenetic diversity, bat compositional  
63 uniqueness, and predicted viral sharing risk.

64 **Bats phyogeography**

65 For every pixel, we measured Faith's Phylogenetic Diversity (Faith 1992) based on a recent synthetic tree  
66 with robust time calibration, covering about 6000 mammalian species (Upham et al. 2019). Faith's PD  
67 measures the sum of unique branches from an arbitrary root to a set of tips, and comparatively larger  
68 values indicate a more phylogenetic diverse species pool. We measured phylogenetic diversity starting  
69 from the root of the entire tree (and not from Chiroptera); this bears no consequences on the resulting  
70 values, since all branches leading up to Chiroptera are only counted one per species pool, and (as we  
71 explain when describing the assembly of the composite risk map), all individual risk components are  
72 ranged in [0,1]. This measure incorporates a richness component, which we chose not to correct for; the  
73 interpretation of the phylogenetic diversity is therefore a weighted species richness, that accounts for  
74 phylogenetic over/under-dispersal in some places.

75 **Bats compositional uniqueness**

76 For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres  
77 2013); LCBD works from a species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites  
78 are columns, and a value of 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel  
79 represents a unique location, and following best practices (Legendre and Condit 2019) transformed it  
80 using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw  
81 community data is particularly important for two reasons: first, it prevents the artifact of richer sites  
82 having higher importance; second, it removes the effect of overall species richness, which is already

83 incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a  
84 community that is on average more dissimilar in species composition than what is expected knowing the  
85 entire matrix, i.e. a more unique community. Recent results by Dansereau et al. (2022) shows that LCBD  
86 measures are robust with regards to spatial scale, and are therefore applicable at the global scale.

87 **Viral sharing between hosts**

88 For all bat hosts of betacoronaviruses, we extracted their predicted viral sharing network (Albery et al.  
89 2020). This network stores pairwise values of viral community similarity. To project viral sharing values  
90 into a single value for every pixel, we averaged the pairwise scores. High values of the average sharing  
91 propensity means that this specific extant bat assemblage is likely to be proficient at exchanging viruses.

92 **Composite risk map**

93 To visualize the aggregated risk at the global scale, we combine the three individual risk components  
94 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model  
95 (Seekell et al. 2018). In this approach, every risk component gets assigned a component in the RGB color  
96 model (phylogenetic diversity is green, compositional uniqueness is red, and viral sharing is blue). In  
97 order to achieve a valid RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel  
98 with no sharing, no phylogenetic diversity, and no compositional uniqueness is black, and a pixel with  
99 maximal values for each is white. This additive model conveys both the intensity of the overall risk, but  
100 also the nature of the risk as colors diverge towards combinations of values for three risk components. Out  
101 of the possible combinations, the most risky in terms or rapid diversification and spillover potential is high  
102 phylogenetic diversity and low viral sharing (Gomulkiewicz et al. 2000, Cavender-Bares et al. 2009), in  
103 that this allows multiple independent host-virus coevolutionary dynamics to take place in the same  
104 location. In the colorimetric space, this correspond to yellow – because the HSV space is more amenable  
105 to calculations for feature extraction (see e.g. Keke et al. 2010), we measured the risk level by calculating  
106 the angular distance of the hue of each pixel to a reference value of 60, and weighted this risk level by the  
107 value component. Specifically, given a pixel with colorimetric coordinates  $(h, s, v)$ , its ranged weighted  
108 risk value is

$$v \times \left[ 1 - \frac{|\text{atan}(\cos(\text{rad}(h)), \sin(\text{rad}(h))) - X|}{2\pi} \right],$$

109 where X is  $\text{atan}(\cos(\text{rad}(60)), \sin(\text{rad}(60)))$ , a constant approximately equal to 0.5235.

## 110 Viral phyogeography and evolutionary diversification

111 We used the following query to pull all betacoronavirus sequence data from the GenBank Nucleotide  
 112 database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR betacoronavirus[All Fields]) NOT  
 113 (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR sars-cov-2[All Fields]). We added a  
 114 single representative sequence for SARS-CoV-2 and manually curated to remove sequences without the  
 115 RNA-dependent RNA polymerase (RdRp) sequence or that contained words indicating recombinant or  
 116 laboratory strains including “patent,” “mutant,” “GFP,” and “recombinant.” We filtered over-represented  
 117 taxa including betacoronavirus 1, hCoV-OC43, Middle East respiratory syndrome coronavirus, Murine  
 118 hepatitis virus, and hCoV-HKU1. Curated betacoronavirus RdRp sequences were then aligned using  
 119 MAFFT v 1.4.0 (**Katoh and Standley 2013**, parameters in text?) and a maximum likelihood tree  
 120 reconstructed in IQ-TREE v 1.6.12 (Nguyen et al. 2015) with ModelFinder (Kalyaanamoorthy et al. 2017)  
 121 ultrafast bootstrap approximation (Hoang et al. 2018) and the following parameters (**STEPH WILL ADD**,  
 122 parameters in text?).

123 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat  
 124 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the  
 125 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary  
 126 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary  
 127 distinctiveness (Isaac et al. 2007) for each of the viruses in the tree, then averaged these at the bat species  
 128 level, projected these values onto their geographic distributions, and averaged across every bat found in a  
 129 given pixel. As such, this can be thought of as a map of the mean evolutionary distinctiveness of the  
 130 known viral community believed to be associated with a particular subset of bats present.

131 **Co-distribution of hosts and viral hotspots**

132 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
133 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,  
134 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their  
135 original method, a distance matrix - where each row or column represents a geographic raster's grid cell,  
136 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes  
137 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected  
138 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel  
139 methodology. First, we measure the phylogenetic distance between the different viruses in the  
140 betacoronavirus tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently, we  
141 take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this  
142 case) to project the viral tree into an n-dimensional space. We then take the first two principal  
143 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value  
144 and projected them using a four-color bivariate map.

145 **Results and discussion**

146 **Host richness does not predict virus distinctiveness**

147 Bats are found worldwide and are both one of the most diverse groups among mammals (**Moratelli &**  
148 **Calisher, 2015**), and one of the main animal reservoir for different strains of betacoronaviruses (Drexler et  
149 al. 2014). This has attracted attention to areas where high diversity of bats, and therefore presumably high  
150 diversity of betacoronaviruses, can be an important issue for human health (Calisher et al. 2006, Moratelli  
151 and Calisher 2015). By overlaying the IUCN rangempas for confirmed bat hosts of betacoronaviruses  
152 [fig. 1; top], we see that the the main hotspots of host richness are primarily South-Eastern Asia, parts of  
153 Southern Europe, and to a lesser extent parts of Africa in the -25-0 range of latitudes. The description of  
154 host richness is an important first step towards understanding risk, as previous research (**Anthony et al.,**  
155 **2017; Mollentze & Streicker, 2020**) states that locally diverse bat communities could maintain more  
156 viruses and hence, a higher probability of having a pathogen that could represent a risk for human health.

158 Nevertheless, locally diverse and virus-rich bat communities could represent an increased risk of spillover  
159 under climate change through the creation of novel interactions (**Ice ice berg berg**), and therefore the  
160 diversity of betacoronavirus strains should similarly be accounted for. In fig. 1 (bottom), we contrast the  
161 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness  
162 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as microchiroptera and  
163 macrochiroptera, where macrochiroptera have an older history from an evolutionary perspective  
164 compared to macrochiroptera (Teeling et al. 2005, Springer 2013). Specifically, we would expect that the  
165 so-called “New World” group of bats, being more evolutionary distinct, would also have evolutionary  
166 distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a distinct  
167 bat-betacoronavirus complex. By contrast, South-Eastern Asia has a lot of non-evolutionary distinct bats,  
168 but evolutionary-distinct viruses.

169 It is noteworthy that outside of South America, viral evolutionary distinctiveness does not accurately  
170 tracks host diversity, with some areas having over-distinct viruses (southern China but, oddly, not the rest  
171 of southeast Asia). There are a number of likely explanations. First, given the richness of bats in southeast  
172 Asia, many betacoronaviruses likely remain to be discovered in this region. Indeed, global predictions by  
173 (**Becker?**) highlight that southeast Asia is a likely hotspot of unconfirmed hosts of betacoronaviruses,  
174 which would likely result in additional viral discoveries. This idea is unsurprising given the growing  
175 realization, around the emergence of SARS-CoV-2, that a unique lineage of similar viruses are widespread  
176 in bats but still mostly undescribed. The most distinct bats/betacoronavirus complex is found in South  
177 America, a region with a comparatively lower number of hosts; this matches with the isolation through  
178 variance of the host group, and may highlight a different co-evolutionary dynamic. Alternatively, this  
179 distinctiveness hotspot may be a product of under-sampling: South-America is one of the places where the  
180 fewest betacoronaviruses have been discovered (Anthony et al. 2017, Olival et al. 2017, Allen et al. 2017),  
181 resulting in sparser phylogenetic tree, thereby artificially inflating distinctiveness. Adding more viruses  
182 would bring the distinctiveness of known sequences down.

### 183 **The phylogeographic regions of hosts and their viruses overlap**

184 Despite the difference in evolutionary distinctiveness globally, there are reasons to expect that the  
185 phylogeography of bats and betacoronaviruses should show some degree of congruence. High density of  
186 hosts sharing the same virus (albeit possibly different strains) can drive or result from evolution of the bat

187 antiviral immune system, resulting in spatially distinct immunological responses, as evidenced in several  
188 bat species (Banerjee et al. 2020). Immune characteristics that allow bats to be better adapted to infection  
189 by emerging viruses (Gorbunova et al., 2020; Irving et al., 2021), in addition to being hardcoded in their  
190 genome (**Jebb et al. six...**), may be related to a wide variety of diets (Jones et al., 2022; Moreno Santillán  
191 et al., 2021; Banerjee et al., 2020; Schneeberger et al., 2013; Muylaert et al., 2021), themselves likely to be  
192 driven by spatial effects, especially at the local scale – bats, indeed, occupy a variety of environments, and  
193 therefore display a variety of adaptations to these environments.

194 [Figure 2 about here.]

195 In fig. 2, we show a projection of the phylogeographic signal of bats (top) and viruses (bottom) in space;  
196 the distinct groupings (represented by different colors symbolizing positions in the subspace formed by  
197 the first two axes of the PCoA) are essentially equivalent between the two groups, and can be coarsely  
198 delineated as southeast Asia, Eurasia above a northing of 25, and Africa and south America. These results  
199 suggest that, although the evolutionary distinctiveness of the bat/betacoronavirus complex varies spatially,  
200 the system shows an important degree of spatial consistency, with a reduced number of bioregions.  
201 Available information describing the spillover of zoonotic betacoronaviruses of bat origin where data was  
202 available before and up through the COVID-19 pandemic puts spillover events of SARS-CoV-2 in Wuhan,  
203 China; SARS-CoV in **XXX** based on the presence of closest known viruses circulating in nature, and a  
204 nearby location where serological (antibody) evidence has indicated human exposure to SARS-like viruses  
205 (Wang et al. 2018 *Virologica Sinica*); MERS-CoV in **XXX** based on index cases available from a  
206 recently-published compendium of cases (Ramshaw et al. 2019). For the latest event, most if not all index  
207 cases are presumed to be camel-to-human transmission, and the precise origin point (if it exists) of  
208 MERS-CoV in bats is uncertain. Recent recombinant canine coronavirus spillover events in Haiti (**ref**)  
209 and Europe (**ref**) are not relevant here, as bats' involvement in these cycles of transmission have been  
210 supposed to be non-existent.

## 211 **Coevolution-informed spillover risk is different in space**

212 As host richness, joint distinctiveness, or phylogeographic structure suggest that the bat/betacoronavirus  
213 complex is globally fragmented enough to give rise to both different levels of risk (as evidenced by the  
214 spatial location of spillover events) and different types of co-evolutionary dynamics, we turn to the

215 Geographic Mosaic Theory of Coevolution **REF** to provide a measure of risk accounting for multiple  
216 processes. In fig. 3, we overlapped three components of spillover risk: viral sharing, *i.e.* the chance that  
217 two bats will share viruses overall; Local Contribution to Beta Diversity, *i.e.* the fact that a bat community  
218 is compositionally unique compared to the average compositional similarity across the entire system;  
219 finally, host phylogenetic diversity, *i.e.* how dispersed the bats in a location are within the tree of life. This  
220 approach leads to the definition of broad biogeographic regions of risk, where the same color represents  
221 the same type of risk. By way of contrast to figures fig. 1 and fig. 2, these regions do not necessarily  
222 overlap with previous spatial partitions of the bat/betacoronavirus complex.

223 [Figure 3 about here.]

224 From the perspective of spillover risk, the most important combination of factors is a high phylogenetic  
225 diversity of hosts with low viral sharing; this, essentially, means that very different betacoronavirus could  
226 co-exist within the same place. This is particularly the case given that betacoronaviruses often evolve and  
227 even achieve host shifts through recombination, which requires the co-occurrence of sufficiently distinct  
228 viruses to be a major driver of emergence. In fig. 3, this corresponds to yellow to pale green areas, which  
229 are essentially limited to South-Eastern Asia, and to some part of Sub-Saharan Africa. Adopting a  
230 geographic mosaic theory perspective on risk, other regions of the world are of lesser concern fig. 4. Our  
231 risk decomposition does not account for viral diversity or distinctiveness. The simple rationale behind it is  
232 that the acquisition of viral data is rarely disconnected from the acquisition of host data; furthermore,  
233 there are more sources of information on hosts than on viruses, allowing to develop a host-centric  
234 perspective on risk. Areas with high bat diversity and high turnover *may* facilitate the evolutionary  
235 radiation of viruses, matching previous findings that the diversification of bat coronaviruses is driven  
236 largely by host shifts (inter-genus or higher levels of cross-species transmission) and, to a lesser degree,  
237 cospeciation and sharing (intra-genus cross-species transmission; **Anthony et al. 2017**). This  
238 diversification is not an actual risk factor for spillover itself, but acts downstream of a spillover event by  
239 increasing the random chance of the emergence of a virus with the raw genomic components required for  
240 the potential to infect humans.

241 [Figure 4 about here.]

242 **Human occupancy drives different levels of effective risk globally**

243 Based on the previous result, we extracted the yellow component of the risk map (TP add methods), to  
244 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. However,  
245 this maps the potential risk, which must be weighed by the potential for contacts with humans. As a proxy  
246 for this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a  
247 reasonable proxy for the density of humans per unit area, which increases the probability of pathogen  
248 spread more widely (Hazarie et al., 2021). Since human activity is required to amplify the frequency of  
249 virus encounters and thus create areas of viral amplification, mapping the potential risk against measures  
250 of land use is required to generate a more actionable assessment of risk. This map is presented in fig. 5.  
251 Most of South America and Europe are at low risk, as although densely populated, settlements tend to be  
252 in areas with lower potential risk. However, this mapping reveals that South-East Asia, the Indian  
253 subcontinent, and parts of sub-Saharan Africa, are at high risk due to the overlap between built areas and  
254 bat communities representing more opportunities for cross-species transmission of betacoronaviruses.

255 [Figure 5 about here.]

256 **Conclusion**

257 Our study focuses largely on the biogeography of hosts. Yet, we know that viruses with high host  
258 plasticity, that is, the ability of a given virus to adapt to various taxonomic orders and ecological groups  
259 (Kreuder Johnson et al., 2015); are more likely to amplify viral spillover, followed by secondary  
260 human-to-human transmission, and geographical spread (Hazarie et al., 2021). High viral host plasticity is  
261 an especially important trait for RNA viruses such as betacov (Kreuder Johnson et al., 2015; Haddad et al.,  
262 2021). Indeed, our analysis of viral sequences reveals that Latin America is a hotspot of viral  
263 distinctiveness, suggesting that this part of the bats-betacov system may be undergoing independent  
264 evolutionary dynamics (related species sharing viruses that are different from the rest of the global pool).  
265 The other hotspot of viral distinctiveness is S.E. Asia, in which richness is high but sharing is low; this  
266 suggests a different type of evolutionary dynamics (unrelated viruses coevolving with evolutionarily  
267 distinct hosts, generating high diversity locally). Both of these areas should be priority areas for sampling,  
268 especially since (**Becker?**) advance that they harbor undiscovered hosts of beta-coronaviruses. This

269 diversity of hosts, and the mechanisms by which the exchange of viruses occurs between species, is largely  
270 affected by the local environmental conditions and environmental change.

271 There are several factors that drive changes in the diversity of bats (Alves et al., 2018), but human  
272 activities' effects on the ecosystem (like modifications of land use) could significantly decrease it.

273 Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their  
274 biogeographic variation, and human population density and other anthropogenic factors are decisive  
275 moderators for its implications in public health. With the increase of contact between humans and  
276 potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al., 2020), as previous  
277 studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal  
278 (Gryseels et al., 2017). One of these scenarios where interaction between bats and humans can occur can  
279 be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats  
280 not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and  
281 humans (Deshpande et al., XX). This represents a challenge for conservation strategies and disease  
282 ecology since we have areas with both potential for the acquisition of zoonotic viruses and bat-human  
283 interaction; in particular, the challenge lies in the fact that actual exposure must be quantified from  
284 several scenarios, including both direct and indirect bat - human interaction.

285 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
286 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
287 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research  
288 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des  
289 Données (IVADO). This research was enabled in part by support provided by Calcul Québec  
290 ([www.calculquebec.ca](http://www.calculquebec.ca)) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)). NF is funded by the NSERC  
291 BIOS<sup>2</sup> CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by  
292 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation.

## 293 **References**

- 294 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and  
295 implications for emerging infectious disease. - *Zoologia (Curitiba)* 27: 151–162.  
296 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -

- 297 Nature Communications 11: 2260.
- 298 Albery, G. F. et al. 2022. Urban-adapted mammal species have more known pathogens. in press.
- 299 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - Nature  
300 Communications in press.
- 301 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - Virus Evolution in press.
- 302 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - Frontiers in Immunology 11: 26.
- 303 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -  
304 The Lancet Microbe in press.
- 305 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - Clinical Microbiology  
306 Reviews 19: 531–545.
- 307 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - Ecol. Lett.  
308 12: 693–715.
- 309 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species  
310 distribution modelling. - Oikos n/a: e09063.
- 311 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of  
312 SARS. - Antiviral Research 101: 45–56.
- 313 Engering, A. et al. 2013. Pathogen–host–environment interplay and disease emergence. - Emerging  
314 Microbes & Infections 2: e5.
- 315 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1–10.
- 316 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -  
317 The American Naturalist 156: 156–174.
- 318 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:  
319 10–24.
- 320 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE  
321 8: e77183.
- 322 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.  
323 - PLOS ONE 2: e296.

- 324 IUCN 2021. The IUCN Red List of Threatened Species.
- 325 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -  
326 ISRN Biodiversity 2013: e187415.
- 327 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010  
328 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.
- 329 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - Proceedings of  
330 the National Academy of Sciences 104: 5925–5930.
- 331 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species  
332 distributions. - Journal of Biogeography 37: 2029–2053.
- 333 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity  
334 coefficients and partitioning (H Morlon, Ed.). - Ecology Letters 16: 951–963.
- 335 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado  
336 Island forest dynamics plot, Panama. - Forest Ecosystems 6: 7.
- 337 Melaun, C. et al. 2014. Bats as Potential Reservoir Hosts for Vector-Borne Diseases. - In: Klimpel, S. and  
338 Mehlhorn, H. (eds), Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths.  
339 Parasitology Research Monographs. Springer, pp. 25–61.
- 340 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with  
341 emerging deadly viruses? - Memórias do Instituto Oswaldo Cruz 110: 1–22.
- 342 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - Nature 546:  
343 646–650.
- 344 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary  
345 analyses in R. - Bioinformatics 35: 526–528.
- 346 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across  
347 spatial scales. - BMC ecology 18: 18.
- 348 Plowright, R. K. et al. 2015. Ecological dynamics of emerging bat virus spillover. - Proceedings of the  
349 Royal Society B: Biological Sciences 282: 20142124.
- 350 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - Nature Reviews Microbiology 15: 502–510.

- 351 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - The American  
352 Naturalist 164: S79–S89.
- 353 Rego, K. M. da C. et al. 2015. Assessing human-bat interactions around a protected area in northeastern  
354 Brazil. - Journal of Ethnobiology and Ethnomedicine 11: 80.
- 355 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 356 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - Limnology and Oceanography Letters 3:  
357 49–56.
- 358 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic  
359 database.
- 360 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - Current Biology 23: R999–R1001.
- 361 Stone, E. et al. 2015. Managing Conflict between Bats and Humans: The Response of Soprano Pipistrelles  
362 (*Pipistrellus pygmaeus*) to Exclusion from Roosts in Houses. - PLoS ONE 10: e0131825.
- 363 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.  
364 - Science (New York, N.Y.) 307: 580–584.
- 365 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.
- 366 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in  
367 ecology, evolution, and conservation. - PLOS Biology 17: e3000494.
- 368 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,  
369 Eds.). - Springer International Publishing.
- 370 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - Science 320:  
371 70–70.

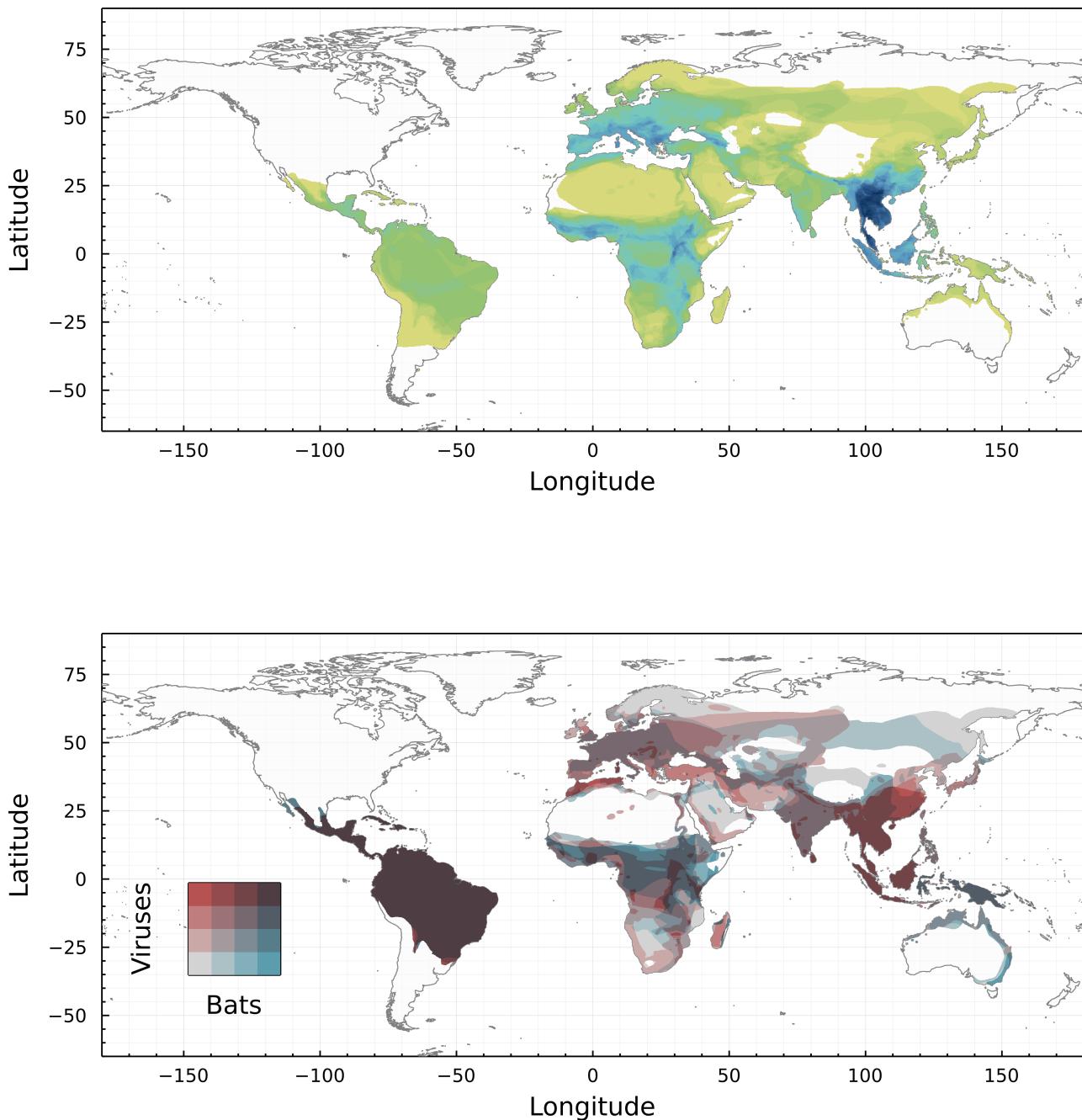


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red). By contrast to the richness map, this reveals that South America has the most evolutionary distinct hosts *and* viruses, whereas South-Eastern Asia has mostly distinct viruses. This is congruent with known results about New World bats being evolutionary distinct, and suggests that they similarly have distinct viruses.

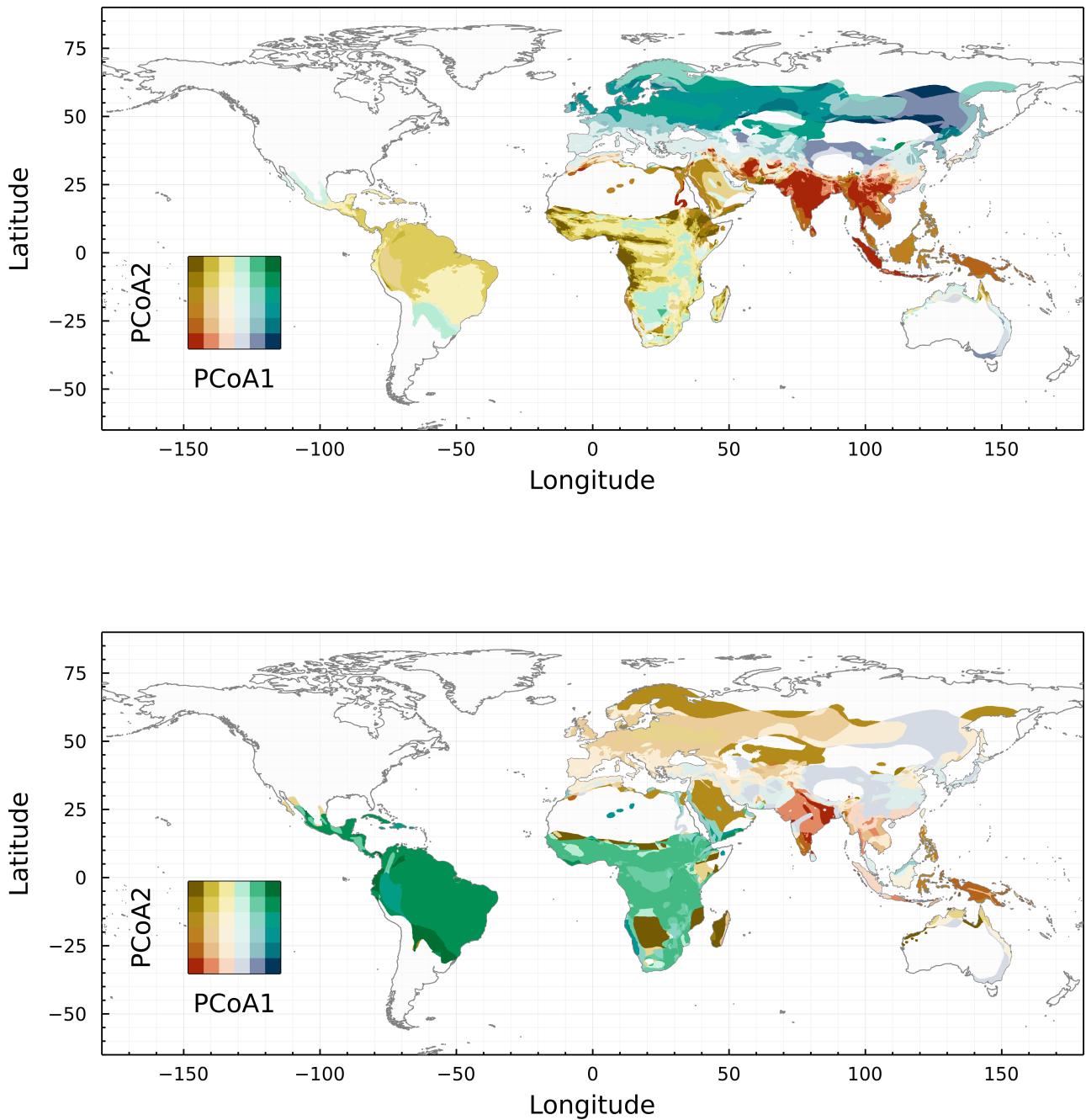


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent, and so cannot be compared directly – that being said, the regions can be compared across maps.

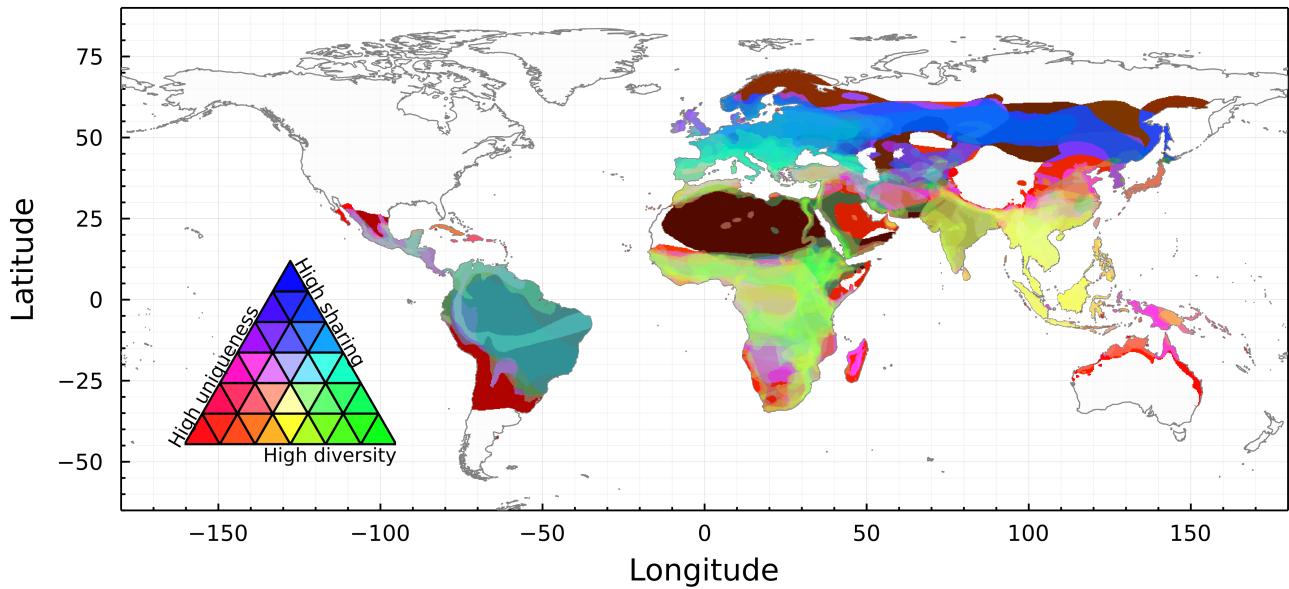


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures would be a pure white (specifically  $\text{RGB}(1.0, 1.0, 1.0)$ ), and a pixel with the lowest possible values would be pure black (specifically  $\text{RGB}(0.0, 0.0, 0.0)$ ). Therefore, lighter values (the sum of the three channels gets closer to 3) indicate higher risk, and the color indicates the proportional distribution of the factors making up the total risk.

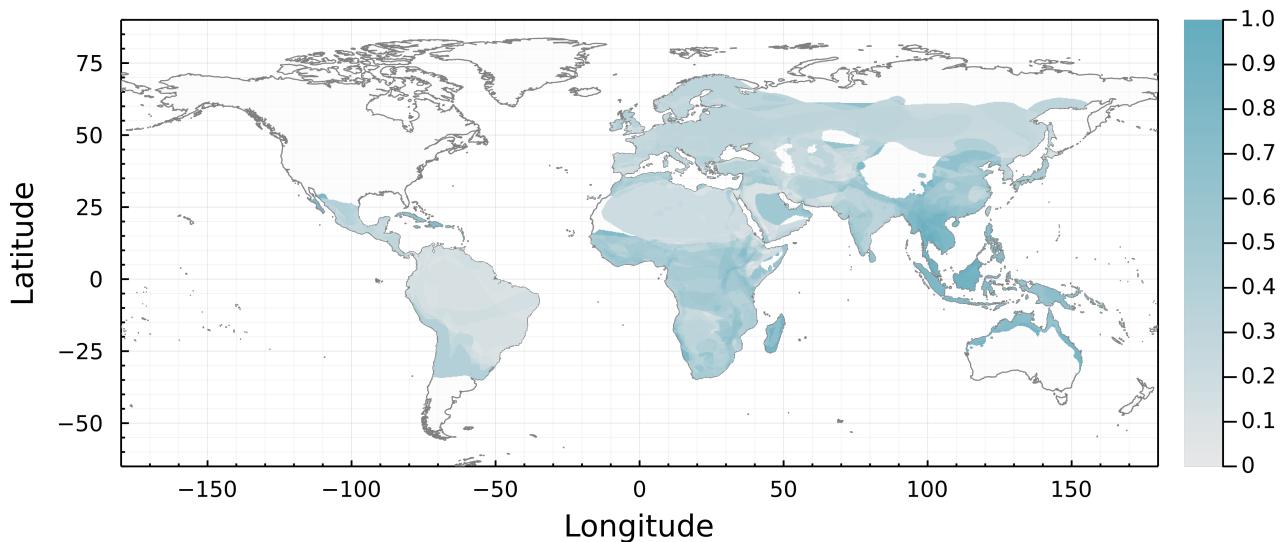


Figure 4: Extraction of a measure of risk based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions at high risk of spillover are southeast Asia and Madagascar.

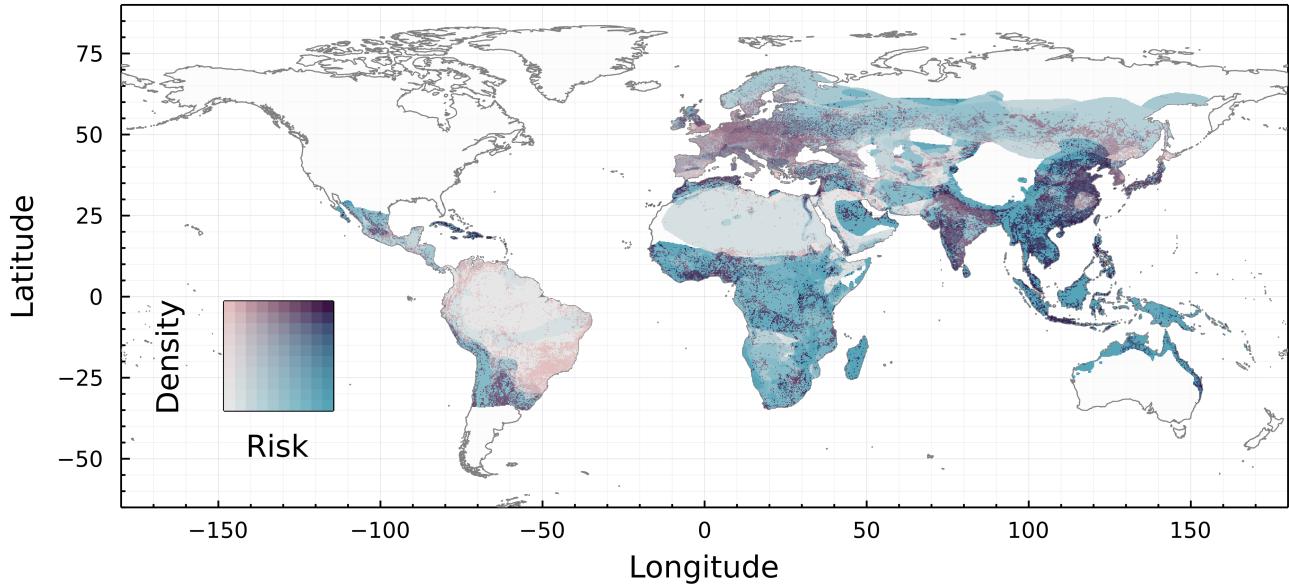


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the risk map. Darker pixels correspond to more risk, in that the GMTC-derived risk of fig. 4 is high *and* the pixel is densely occupied by human populations. This approach increases the relative risk of several regions in Africa, and highlights the risk in India, southeast China, and the arabic peninsula where areas of high to moderate risk overlap with areas of denser population.