

# 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. Within a pool of competent hosts, it is driven by a  
2 multiplicity of factors (Plowright et al. 2017). Although proxies for the local richness of hosts is commonly  
3 analysed (see e.g. Anthony et al. 2017 for coronaviruses), there is an argument to be made that species  
4 who are not (known as) competent hosts of a specific virus genus may not factor into this (Plowright et al.  
5 2015), calling for species-level information. This is especially true as competence data increases predictive  
6 power when the taxonomic scope of hosts of a viral family increases (Becker et al. 2020, Mull et al. 2022).  
7 Similarly, host species who share viruses at different rates should be weighted accordingly (Albery et al.  
8 2020). In mammals, key functional traits (for which phylogeny is a reasonable proxy) are determinants of  
9 the spillover potential (Olival et al. 2017); these include, notably, body mass, and affinity for urban  
10 environments (Albery et al. 2022). Finally, especially when the pool of potential hosts spans the entire  
11 globe, there may be local host pools that are highly unique; not having been observed in other locations,  
12 these can act on the overall risk either by providing novel contact opportunities, reflecting unique  
13 host-environment combinations (Engering et al. 2013), or facilitating rapid evolutionary changes in  
14 specialism of their pathogens (Agosta et al. 2010). In the specific case of generalist pathogens (which is  
15 the case many viruses in the betacoronavirus genus, see e.g. MacLean et al. 2021), there is conceptual and  
16 empirical support to the idea that these community-level mechanisms are even more important in driving  
17 the overall risk (Power and Mitchell 2004).

18 In this paper, we examine the biogeographic structure of bats-betacoronaviruses associations, based on a  
19 curated dataset of all confirmed bat hosts of betacoronaviruses. By drawing on concepts from the  
20 Geographic Mosaic Theory of Coevolution (GMTC; Thompson 2005), we turn these associations into a  
21 spatially explicit additive mapping of zoonotic risk components, revealing the extreme heterogeneity of  
22 risk at the global scale. Explicitely framing the notion of spillover risk based on propositions from the  
23 GMTC (which is to say, based on a framework linking interactions between species to change within  
24 species) is a novel idea, that should be relatively general. Indeed, it only assumes the action of well  
25 described evolutionary mechanisms. The benefit of this approach is to provide the potential for a more  
26 dynamic and nuanced understanding of risk: not only on ecological timescales, but also by providing  
27 clues about which areas can change over micro-evolutionary timescales. This provides a way to look at  
28 spatial structure by accounting for more notions than species richness/similarity, but also a way to identify  
29 spatial areas of higher risk.

30 We identify the Amazon and South-Eastern Asia as hotspots where the phylogenetic distinctiveness of

31 betacoronaviruses is the highest (Anthony et al. 2017); surprisingly, current data suggest that viral sharing  
32 between hosts is high in the Amazon and low in South-Eastern Asia, which has the potential to result in  
33 different evolutionary dynamics between these two regions, hinting at different futures for their viral  
34 communities. This work is important both as a description of the bats-betacoronavirus complex, but also  
35 because more broadly, bats are known reservoirs for a variety of emerging viruses and other pathogens  
36 (Calisher et al. 2006, Melaun et al. 2014), making balancing the needs for bat conservation and disease  
37 prevention most likely very difficult and a source of human-wildlife conflicts, especially in more densely  
38 populated areas (Stone et al. 2015, Rego et al. 2015).

## 39 Methods

### 40 Known betacoronavirus hosts

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

### 47 Bats occurrences

48 We downloaded the rangemap of every extant bat species that was either classified as an empirically  
49 documented host of beta-coronaviruses from the previous step, according to recent IUCN data (IUCN  
50 2021). The range maps were subsequently rasterized using the rasterize function from GDAL (Rouault et  
51 al. 2022) at a resolution of approximately **TK TP**. For every pixel in the resulting raster where at least one  
52 bat host of betacoronavirus was present, we extract the species pool (list of all bat species), which was used  
53 to calculate the following risk assessment components: phylogenetic diversity, bat compositional  
54 uniqueness, and predicted viral sharing risk.

55 **Bats phyogeography**

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

66 **Bats compositional uniqueness**

67 For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres  
68 2013); LCBD works from a species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites  
69 are columns, and a value of 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel  
70 represents a unique location, and following best practices (Legendre and Condit 2019) transformed it  
71 using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw  
72 community data is particularly important for two reasons: first, it prevents the artifact of richer sites  
73 having higher importance; second, it removes the effect of overall species richness, which is already  
74 incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a  
75 community that is on average more dissimilar in species composition than what is expected knowing the  
76 entire matrix, i.e. a more unique community. Recent results by Dansereau et al. (2022) shows that LCBD  
77 measures are robust with regards to spatial scale, and are therefore applicable at the global scale.

78 **Viral sharing between hosts**

79 For all bat hosts of betacoronaviruses, we extracted their predicted viral sharing network (Albery et al.  
80 2020). This network stores pairwise values of viral community similarity. To project viral sharing values

81 into a single value for every pixel, we averaged the pairwise scores. High values of the average sharing  
82 propensity means that this specific extant bat assemblage is likely to be proficient at exchanging viruses.

83 **Composite risk map**

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

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

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

101 **Viral phyogeography and evolutionary diversification**

102 We used the following query to pull all betacoronavirus sequence data from the GenBank Nucleotide  
103 database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR betacoronavirus[All Fields]) NOT  
104 (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR sars-cov-2[All Fields]). We added a

105 single representative sequence for SARS-CoV-2 and manually curated to remove sequences without the  
106 RNA-dependent RNA polymerase (RdRp) sequence or that contained words indicating recombinant or  
107 laboratory strains including “patent,” “mutant,” “GFP,” and “recombinant.” We filtered over-represented  
108 taxa including betacoronavirus 1, hCoV-OC43, Middle East respiratory syndrome coronavirus, Murine  
109 hepatitis virus, and hCoV-HKU1. Curated betacoronavirus RdRp sequences were then aligned using  
110 MAFFT v 1.4.0 (**Katoh and Standley 2013**, parameters in text?) and a maximum likelihood tree  
111 reconstructed in IQ-TREE v 1.6.12 (Nguyen et al. 2015) with ModelFinder (Kalyaanamoorthy et al. 2017)  
112 ultrafast bootstrap approximation (Hoang et al. 2018) and the following parameters (**STEPH WILL ADD,**  
113 **parameters in text?**).

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

## 122 Co-distribution of hosts and viral hotspots

123 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
124 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,  
125 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their  
126 original method, a distance matrix - where each row or column represents a geographic raster’s grid cell,  
127 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes  
128 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected  
129 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel  
130 methodology. First, we measure the phylogenetic distance between the different viruses in the  
131 betacoronavirus tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently, we  
132 take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this  
133 case) to project the viral tree into an n-dimensional space. We then take the first two principal

134 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value  
135 and projected them using a four-color bivariate map.

## 136 **Results and discussion**

### 137 **Host richness does not predict virus distinctiveness**

138 Bats are found worldwide and are both one of the most diverse groups among mammals and one of the  
139 main animal reservoir for different strains of betacoronaviruses (Drexler et al. 2014). This has attracted  
140 attention to areas where high diversity of bats, and therefore presumably high diversity of  
141 betacoronaviruses, can be an important issue for human health (Calisher et al. 2006, Moratelli and  
142 Calisher 2015). By overlaying the IUCN rangempas for confirmed bat hosts of betacoronaviruses [fig. 1;  
143 top], we see that the the main hotspots of host richness are primarily South-Eastern Asia, parts of  
144 Southern Europe, and to a lesser extent parts of Africa in the -25-0 range of latitudes. The description of  
145 host richness is an important first step towards understanding risk, as previous research (Anthony et al.  
146 2017, Mollentze and Streicker 2020) states that locally diverse bat communities could maintain more  
147 viruses and hence, a higher probability of having a pathogen that could represent a risk for human health.

148 [Figure 1 about here.]

149 Nevertheless, locally diverse and virus-rich bat communities could represent an increased risk of spillover  
150 under climate change through the creation of novel interactions (Carlson et al. 2022), and therefore the  
151 diversity of betacoronavirus strains should similarly be ccounted for. In fig. 1 (bottom), we contrast the  
152 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness  
153 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as microchiroptera and  
154 macrochiroptera, where macrochiroptera have an older history from an evolutionary perspective  
155 compared to macrochiroptera (Teeling et al. 2005, Springer 2013). Specifically, we would expect that the  
156 so-called “New World” group of bats, being more evolutionary distinct, would also have evolutionary  
157 distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a distinct  
158 bat-betacoronavirus complex. By contrast, South-Eastern Asia has a lot of non-evolutionary distinct bats,  
159 but evolutionary-distinct viruses.

160 It is noteworthy that outside of South America, viral evolutionary distinctiveness does not accurately  
161 tracks host diversity, with some areas having over-distinct viruses (southern China but, oddly, not the rest  
162 of southeast Asia). There are a number of likely explanations. First, given the richness of bats in southeast  
163 Asia, many betacoronaviruses likely remain to be discovered in this region. Indeed, global predictions by  
164 Becker et al. (2022) highlight that southeast Asia is a likely hotspot of unconfirmed hosts of  
165 betacoronaviruses, which would likely result in additional viral discoveries. This idea is unsurprising given  
166 the growing realization, around the emergence of SARS-CoV-2, that a unique lineage of similar viruses are  
167 widespread in bats but still mostly undescribed. The most distinct bats/betacoronavirus complex is found  
168 in South America, a region with a comparatively lower number of hosts; this matches with the isolation  
169 through variance of the host group, and may highlight a different co-evolutionary dynamic. Alternatively,  
170 this distinctiveness hotspot may be a product of under-sampling: South-America is one of the places  
171 where the fewest betacoronaviruses have been discovered (Anthony et al. 2017, Olival et al. 2017, Allen et  
172 al. 2017), resulting in sparser phylogenetic tree, thereby artificially inflating distinctiveness. Adding more  
173 viruses would bring the distinctiveness of known sequences down.

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

175 Despite the difference in evolutionary distinctiveness globally, there are reasons to expect that the  
176 phylogeography of bats and betacoronaviruses should show some degree of congruence. High density of  
177 hosts sharing the same virus (albeit possibly different strains) can drive or result from evolution of the bat  
178 antiviral immune system, resulting in spatially distinct immunological responses, as evidenced in several  
179 bat species (Banerjee et al. 2020). Immune characteristics that allow bats to be better adapted to infection  
180 by emerging viruses (Gorbunova et al. 2020, Irving et al. 2021), in addition to being hardcoded in their  
181 genome (Jebb et al. 2020), may be related to a wide variety of diets (Banerjee et al. 2020, Moreno Santillán  
182 et al. 2021, Muylaert et al. 2021), themselves likely to be driven by spatial effects, especially at the local  
183 scale – bats, indeed, occupy a variety of environments, and therefore display a variety of adaptations to  
184 these environments.

185 [Figure 2 about here.]

186 In fig. 2, we show a projection of the phylogeographic signal of bats (top) and viruses (bottom) in space;  
187 the distinct groupings (represented by different colors symbolizing positions in the subspace formed by

188 the first two axes of the PCoA) are essentially equivalent between the two groups, and can be coarsely  
189 delineated as southeast Asia, Eurasia above a northing of 25, and Africa and south America. These results  
190 suggest that, although the evolutionary distinctiveness of the bat/betacoronavirus complex varies spatially,  
191 the system shows an important degree of spatial consistency, with a reduced number of bioregions.  
192 Available information describing the spillover of zoonotic betacoronaviruses of bat origin where data was  
193 available before and up through the COVID-19 pandemic puts spillover events of SARS-CoV-2 in Wuhan,  
194 China; SARS-CoV in Guangdong, China based on the presence of closest known viruses circulating in  
195 nature, and a nearby location where serological (antibody) evidence has indicated human exposure to  
196 SARS-like viruses (Wang et al. 2018); MERS-CoV in Saudi Arabia based on index cases available from a  
197 recently-published compendium of cases (Ramshaw et al. 2019). For the latest event, most if not all index  
198 cases are presumed to be camel-to-human transmission, and the precise origin point (if it exists) of  
199 MERS-CoV in bats is uncertain. Recent recombinant canine coronavirus spillover events in Haiti  
200 (Lednicky et al. 2021) and Europe (Vlasova et al. 2022) are not relevant here, as bats' involvement in these  
201 cycles of transmission have been supposed to be non-existent. These index cases fall within different  
202 phylogeographic bioregions (fig. 2), which further highlight the issue that different host-virus sub-systems  
203 may lead to widespread emergence.

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

205 As host richness, joint distinctiveness, or phylogeographic structure suggest that the bat/betacoronavirus  
206 complex is globally fragmented enough to give rise to both different levels of risk (as evidenced by the  
207 spatial location of spillover events) and different types of co-evolutionary dynamics, we turn to the  
208 Geographic Mosaic Theory of Coevolution (Thompson 2005) to provide a measure of risk accounting for  
209 multiple processes. In fig. 3, we overlapped three components of spillover risk: viral sharing, *i.e.* the  
210 chance that two bats will share viruses overall; Local Contribution to Beta Diversity, *i.e.* the fact that a bat  
211 community is compositionally unique compared to the average compositional similarity across the entire  
212 system; finally, host phylogenetic diversity, *i.e.* how dispersed the bats in a location are within the tree of  
213 life. This approach leads to the definition of broad biogeographic regions of risk, where the same color  
214 represents the same type of risk. By way of contrast to figures fig. 1 and fig. 2, these regions do not  
215 necessarily overlap with previous spatial partitions of the bat/betacoronavirus complex.

[Figure 3 about here.]

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

[Figure 4 about here.]

234 From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide  
235 hotspots of betacoronavirus risk through mixing of unique viruses (via codivergence) and in turn  
236 recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat  
237 species are endemic following evolutionary divergence from sister species in both African and Asian  
238 continents (e.g. Shi et al. 2014). Recent surveillance (Kettenburg et al. 2022) has identified a novel  
239 betacoronavirus (in the subgenus *Nobecovirus*) in Madagascar-endemic pteropid bat species (*Pteropus*  
240 *rufus*, *Rousettus madagascariensis*), emphasizing strong proof of principle in model predictions.

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

242 Based on the previous result, we extracted the risk component from the composite map (see Methods), to  
243 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. However,  
244 this maps the potential risk, which must be weighed by the potential for contacts with humans. As a proxy  
245 for this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a  
246 reasonable proxy for the density of humans per unit area, which increases the probability of pathogen  
247 spread more widely (Hazarie et al. 2021). Since human activity is required to amplify the frequency of  
248 virus encounters and thus create areas of viral amplification, mapping the potential risk against measures  
249 of land use is required to generate a more actionable assessment of risk. This map is presented in fig. 5.  
250 Most of South America and Europe are at low risk, as although densely populated, settlements tend to be  
251 in areas with lower potential risk. However, this mapping reveals that South-East Asia, the Indian  
252 subcontinent, and parts of sub-Saharan Africa, are at high risk due to the overlap between built areas and  
253 bat communities representing more opportunities for cross-species transmission of betacoronaviruses. In  
254 looking for the origins of SARS in China, Xu et al. (2004) present serological evidence that strongest  
255 human-animal contact results in higher risk of virus exposure, regardless of the animal species, but that  
256 different types of contact had different impacts. Ideally, finer-grained information about human activity  
257 (rather than human presence through anthropisation) could allow to partition this risk further.

258 [Figure 5 about here.]

259 **Conclusion**

260 Our study focuses largely on the biogeography of hosts. Yet, we know that viruses with high host  
261 plasticity, that is, the ability of a given virus to adapt to various taxonomic orders and ecological groups  
262 (Kreuder Johnson et al. 2015), are more likely to amplify viral spillover, followed by secondary  
263 human-to-human transmission, and geographical spread (Hazarie et al. 2021). High viral host plasticity is  
264 an especially important trait for RNA viruses like betacoronaviruses (Haddad et al. 2021). Indeed, our  
265 analysis of viral sequences reveals that Latin America is a hotspot of viral distinctiveness, suggesting that  
266 this part of the bats-betacoronavirus system may be undergoing independent evolutionary dynamics  
267 (related species sharing viruses that are different from the rest of the global pool). The other hotspot of

268 viral distinctiveness is S.E. Asia, in which richness is high but sharing is low; this suggests a different type  
269 of evolutionary dynamics (unrelated viruses coevolving with evolutionarily distinct hosts, generating high  
270 diversity locally). Both of these areas should be priority areas for sampling, especially since Becker et al.  
271 (2022) advance that they harbor undiscovered hosts of beta-coronaviruses. This diversity of hosts, and the  
272 mechanisms by which the exchange of viruses occurs between species, is largely affected by the local  
273 environmental conditions and environmental change.

274 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to  
275 human health (Letko et al. 2020, Van Brussel and Holmes 2022). Chiropterans emerged around 64 million  
276 years ago and are one of the most diverse mammalian orders, with an estimated richness of more than  
277 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020). They exhibit a broad variety of habitat  
278 use, behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of  
279 several ecosystem services, tied to important ecosystem-derived benefits to human (Kasso and  
280 Balakrishnan 2013). For example, bats are an essential component of many seed-dispersal networks  
281 (Mello et al. 2011). Over two-thirds of bats are know to be either obligate or facultative insectivores,  
282 therefore playing an important role in the regulation of insect pests that can affect crops (Williams-Guillén  
283 et al. 2008, Voigt and Kingston 2016), and vectors of pathogens that put a risk on human health  
284 (Gonsalves et al. 2013a, b). Because bats are globally distributed and have a long evolutionary history,  
285 phylogeographic and biogeographic approaches are required to shed light on the contemporary  
286 distribution of coevolutionary processes between bats and the pathogens they host. Not all areas in which  
287 bats, viruses, and human are co-occurring are facing a risk of spillover towards human populations, and  
288 the areas in which this risk exist may not be facing risks of the same nature and magnitude.

289 There are several factors that drive changes in the diversity of bats (Alves et al. 2018), but human  
290 activities' effects on the ecosystem (like modifications of land use) could significantly decrease it.  
291 Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their  
292 biogeographic variation, and human population density and other anthropogenic factors are decisive  
293 moderators for its implications in public health. With the increase of contact between humans and  
294 potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al. 2020), as previous  
295 studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal  
296 (Gryseels et al. 2017). One of these scenarios where interaction between bats and humans can occur can  
297 be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats

298 not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and  
299 humans (Deshpande et al. 2022). This represents a challenge for conservation strategies and disease  
300 ecology since some areas can have both potential for the acquisition of zoonotic viruses and bat-human  
301 interactions; in particular, the challenge lies in the fact that actual exposure must then be quantified  
302 accounting for several transmission scenarios, including both direct and indirect bat - human interaction.

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

## 311 References

- 312 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and  
313 implications for emerging infectious disease. - *Zoologia* (Curitiba) 27: 151–162.
- 314 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -  
315 *Nature Communications* 11: 2260.
- 316 Albery, G. F. et al. 2022. Urban-adapted mammal species have more known pathogens. in press.
- 317 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - *Nature*  
318 *Communications* in press.
- 319 Alves, D. M. C. C. et al. 2018. Geographic variation in the relationship between large-scale environmental  
320 determinants and bat species richness. - *Basic and Applied Ecology* 27: 1–8.
- 321 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - *Virus Evolution* in press.
- 322 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - *Frontiers in Immunology* 11: 26.
- 323 Becker, D. J. et al. 2020. Beyond Infection: Integrating Competence into Reservoir Host Prediction. -  
324 *Trends in Ecology & Evolution* 35: 1062–1065.

- 325 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -  
326 The Lancet Microbe in press.
- 327 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - Clinical Microbiology  
328 Reviews 19: 531–545.
- 329 Carlson, C. J. et al. 2022. Climate change increases cross-species viral transmission risk. - Nature: 1–1.
- 330 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - Ecol. Lett.  
331 12: 693–715.
- 332 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species  
333 distribution modelling. - Oikos n/a: e09063.
- 334 Deshpande, K. et al. 2022. Forbidden fruits? Ecosystem services from seed dispersal by fruit bats in the  
335 context of latent zoonotic risk. - Oikos (Copenhagen, Denmark): oik.08359.
- 336 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of  
337 SARS. - Antiviral Research 101: 45–56.
- 338 Engering, A. et al. 2013. Pathogen–host–environment interplay and disease emergence. - Emerging  
339 Microbes & Infections 2: e5.
- 340 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1–10.
- 341 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -  
342 The American Naturalist 156: 156–174.
- 343 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:  
344 10–24.
- 345 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE  
346 8: e77183.
- 347 Gorbunova, V. et al. 2020. The World Goes Bats: Living Longer and Tolerating Viruses. - Cell Metabolism  
348 32: 31–43.
- 349 Gryseels, S. et al. 2017. When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure  
350 in the Spatial Spread of Arenaviruses (JH Kuhn, Ed.). - PLOS Pathogens 13: e1006073.

- 351 Haddad, D. et al. 2021. SARS-CoV-2: Possible recombination and emergence of potentially more virulent  
352 strains (H Attoui, Ed.). - PLOS ONE 16: e0251368.
- 353 Hazarie, S. et al. 2021. Interplay between population density and mobility in determining the spread of  
354 epidemics in cities. - Communications Physics 4: 191.
- 355 Irving, A. T. et al. 2021. Lessons from the host defences of bats, a unique viral reservoir. - Nature 589:  
356 363–370.
- 357 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.  
358 - PLOS ONE 2: e296.
- 359 IUCN 2021. The IUCN Red List of Threatened Species.
- 360 Jebb, D. et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. - Nature 583:  
361 578–584.
- 362 Johnson, C. K. et al. 2020. Global shifts in mammalian population trends reveal key predictors of virus  
363 spillover risk. - Proceedings of the Royal Society B: Biological Sciences 287: 20192736.
- 364 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -  
365 ISRN Biodiversity 2013: e187415.
- 366 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010  
367 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.
- 368 Kettenburg, G. et al. 2022. Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a  
369 Unique Evolutionary History for This Coronavirus Clade. - Frontiers in Public Health in press.
- 370 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - Proceedings of  
371 the National Academy of Sciences 104: 5925–5930.
- 372 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species  
373 distributions. - Journal of Biogeography 37: 2029–2053.
- 374 Kreuder Johnson, C. et al. 2015. Spillover and pandemic properties of zoonotic viruses with high host  
375 plasticity. - Scientific Reports 5: 14830.
- 376 Lednicky, J. A. et al. 2021. Isolation of a Novel Recombinant Canine Coronavirus from a Visitor to Haiti:  
377 Further Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. - Clinical

- 378 Infectious Diseases: An Official Publication of the Infectious Diseases Society of America: ciab924.
- 379 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity  
380 coefficients and partitioning (H Morlon, Ed.). - *Ecology Letters* 16: 951–963.
- 381 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado  
382 Island forest dynamics plot, Panama. - *Forest Ecosystems* 6: 7.
- 383 Letko, M. et al. 2020. Bat-borne virus diversity, spillover and emergence. - *Nature Reviews Microbiology*  
384 18: 461–471.
- 385 MacLean, O. A. et al. 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist  
386 virus and highly capable human pathogen. - *PLOS Biology* 19: e3001115.
- 387 Melaun, C. et al. 2014. Bats as Potential Reservoir Hosts for Vector-Borne Diseases. - In: Kliment, S. and  
388 Mehlhorn, H. (eds), *Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths.*  
389 *Parasitology Research Monographs*. Springer, pp. 25–61.
- 390 Mello, M. A. R. et al. 2011. The Missing Part of Seed Dispersal Networks: Structure and Robustness of  
391 Bat-Fruit Interactions. - *PLOS ONE* 6: e17395.
- 392 Mollentze, N. and Streicker, D. G. 2020. Viral zoonotic risk is homogenous among taxonomic orders of  
393 mammalian and avian reservoir hosts. - *Proceedings of the National Academy of Sciences* 117:  
394 9423–9430.
- 395 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with  
396 emerging deadly viruses? - *Memórias do Instituto Oswaldo Cruz* 110: 1–22.
- 397 Moreno Santillán, D. D. et al. 2021. Large-scale genome sampling reveals unique immunity and metabolic  
398 adaptations in bats. - *Molecular Ecology*: mec.16027.
- 399 Mull, N. et al. 2022. Virus isolation data improve host predictions for New World rodent  
400 orthohantaviruses. - *Journal of Animal Ecology* in press.
- 401 Muylaert, R. L. et al. 2021. Present and future distribution of bat hosts of sarbecoviruses: Implications for  
402 conservation and public health.
- 403 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - *Nature* 546:  
404 646–650.

- 405 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary  
406 analyses in R. - *Bioinformatics* 35: 526–528.
- 407 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across  
408 spatial scales. - *BMC ecology* 18: 18.
- 409 Plowright, R. K. et al. 2015. Ecological dynamics of emerging bat virus spillover. - *Proceedings of the  
410 Royal Society B: Biological Sciences* 282: 20142124.
- 411 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - *Nature Reviews Microbiology* 15: 502–510.
- 412 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - *The American  
413 Naturalist* 164: S79–S89.
- 414 Ramshaw, R. E. et al. 2019. A database of geopositioned Middle East Respiratory Syndrome Coronavirus  
415 occurrences. - *Scientific Data* 6: 318.
- 416 Rego, K. M. da C. et al. 2015. Assessing human-bat interactions around a protected area in northeastern  
417 Brazil. - *Journal of Ethnobiology and Ethnomedicine* 11: 80.
- 418 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 419 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - *Limnology and Oceanography Letters* 3:  
420 49–56.
- 421 Shi, J. J. et al. 2014. A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with  
422 Evidence for Widespread Panmixia. - *Acta Chiropterologica* 16: 279–292.
- 423 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic  
424 database.
- 425 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - *Current Biology* 23: R999–R1001.
- 426 Stone, E. et al. 2015. Managing Conflict between Bats and Humans: The Response of Soprano Pipistrelles  
427 (*Pipistrellus pygmaeus*) to Exclusion from Roosts in Houses. - *PLoS ONE* 10: e0131825.
- 428 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.  
429 - *Science (New York, N.Y.)* 307: 580–584.
- 430 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.

- 431 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in  
432 ecology, evolution, and conservation. - PLOS Biology 17: e3000494.
- 433 Van Brussel, K. and Holmes, E. C. 2022. Zoonotic disease and virome diversity in bats. - Current Opinion  
434 in Virology 52: 192–202.
- 435 Vlasova, A. N. et al. 2022. Animal alphacoronaviruses found in human patients with acute respiratory  
436 illness in different countries. - Emerging Microbes & Infections 11: 699–702.
- 437 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,  
438 Eds.). - Springer International Publishing.
- 439 Wang, N. et al. 2018. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.  
440 - Virologica Sinica 33: 104–107.
- 441 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - Science 320:  
442 70–70.
- 443 Xu, R.-H. et al. 2004. Epidemiologic Clues to SARS Origin in China. - Emerging Infectious Diseases 10:  
444 1030–1037.

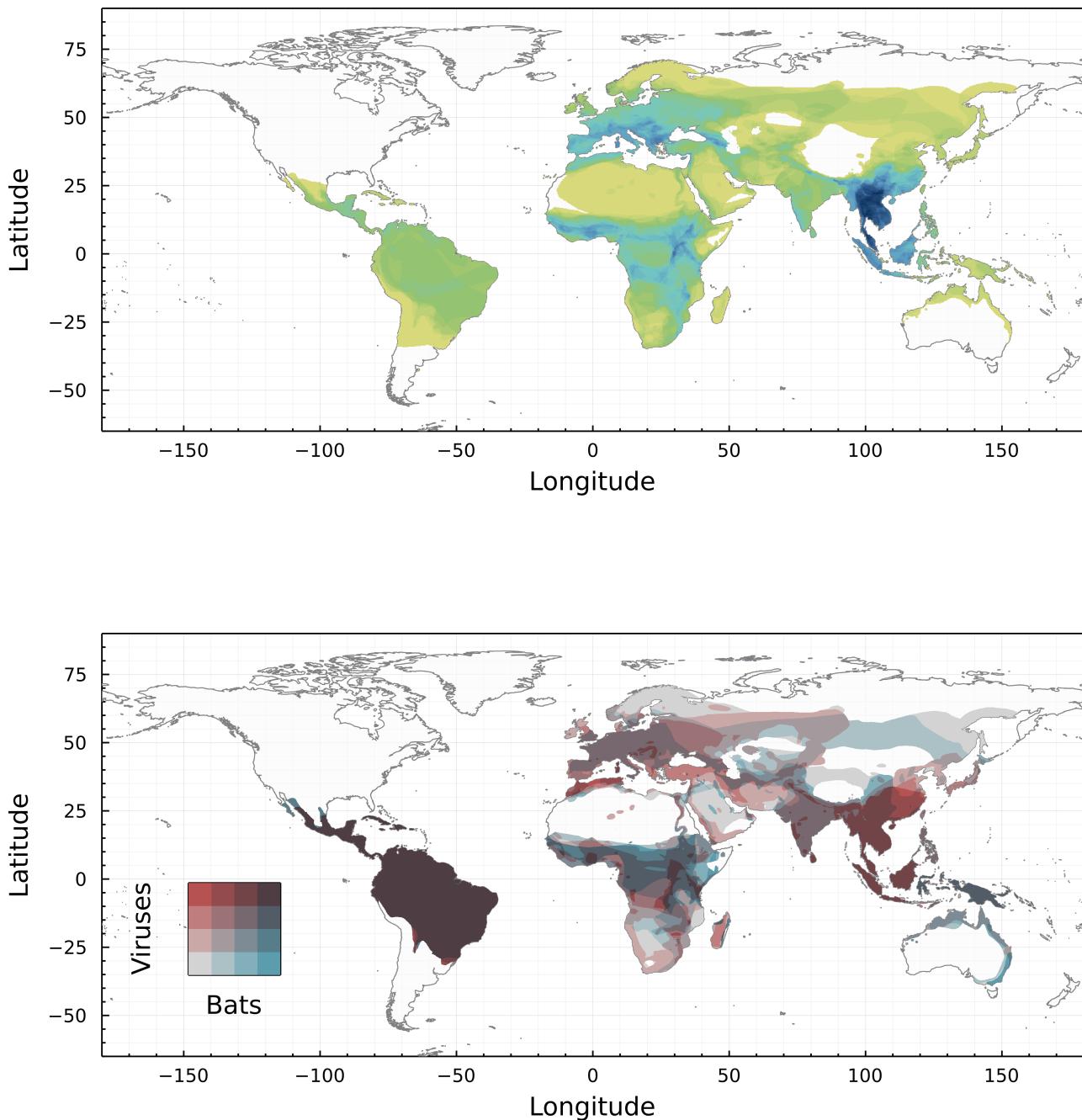


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red). 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.

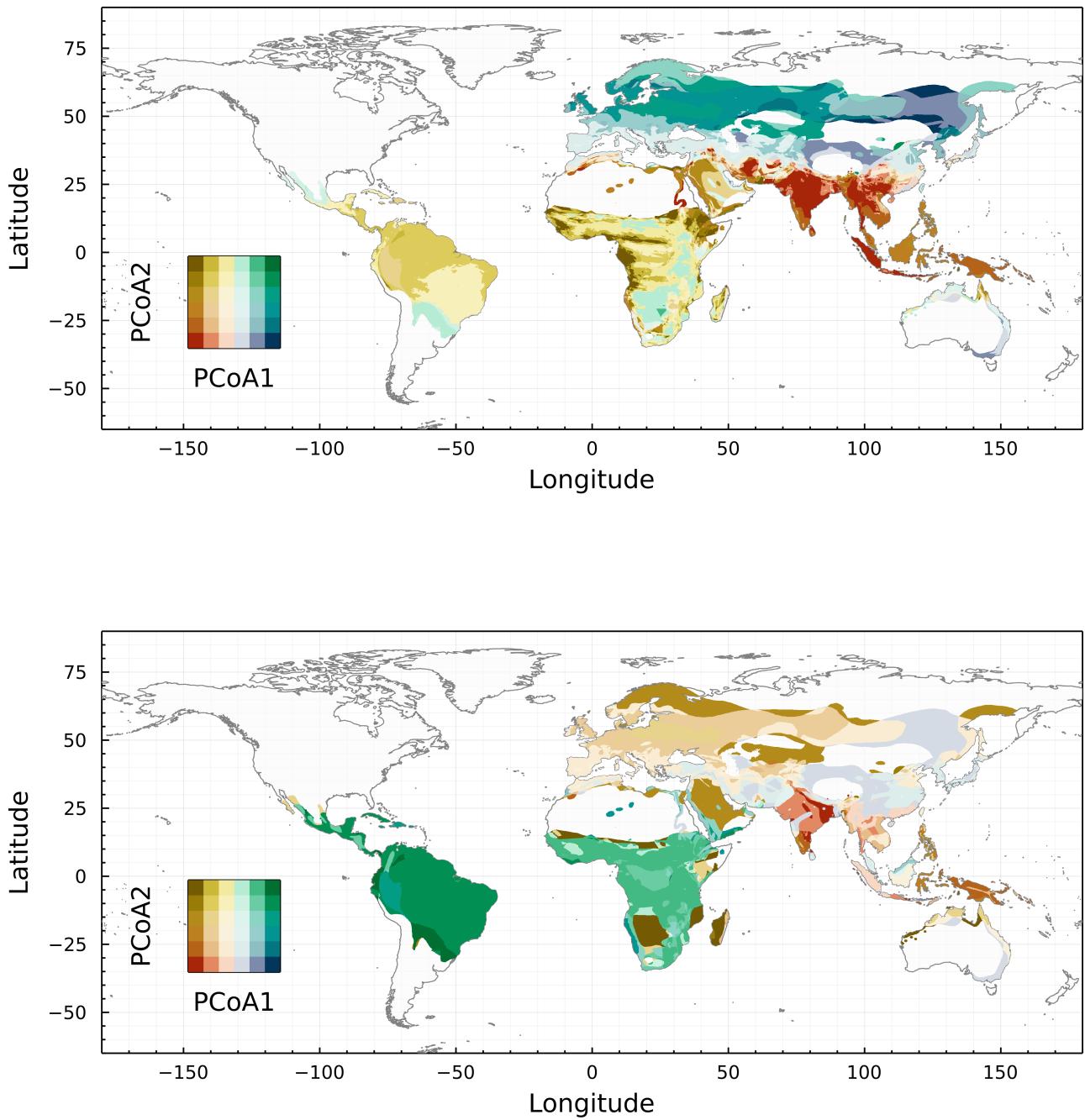


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

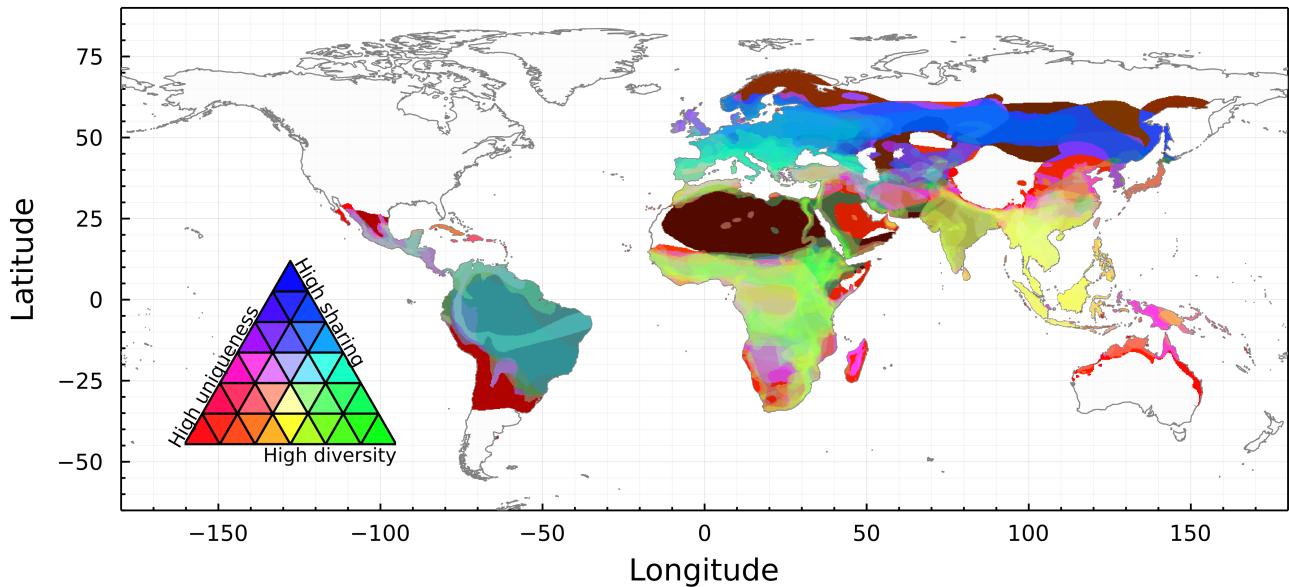


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures would be a pure white (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.

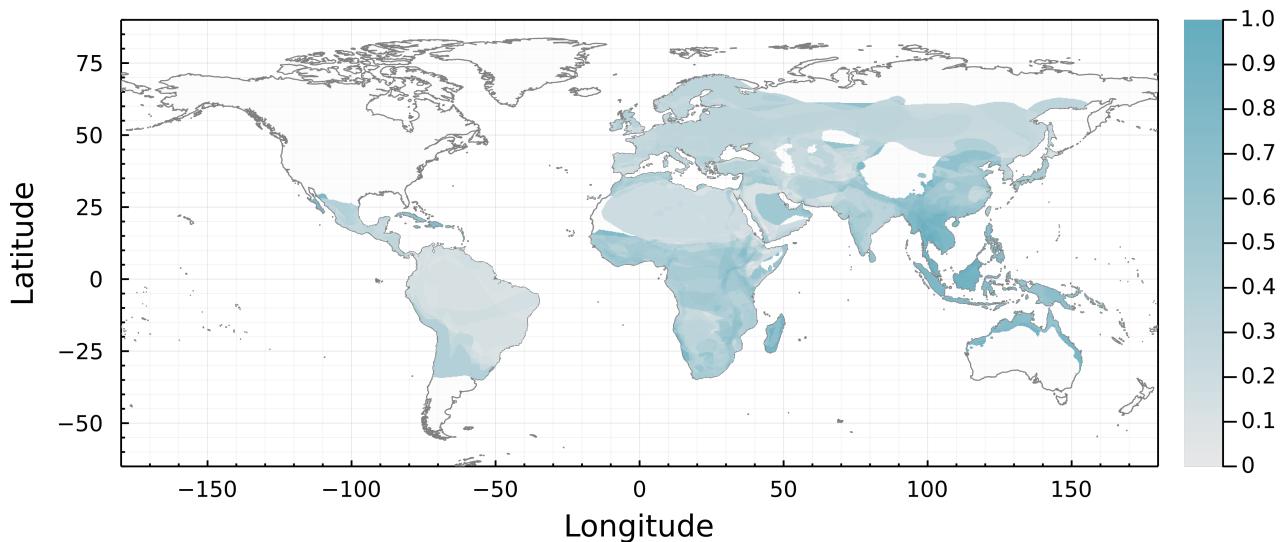


Figure 4: Extraction of a measure of risk based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions at high risk of spillover are southeast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

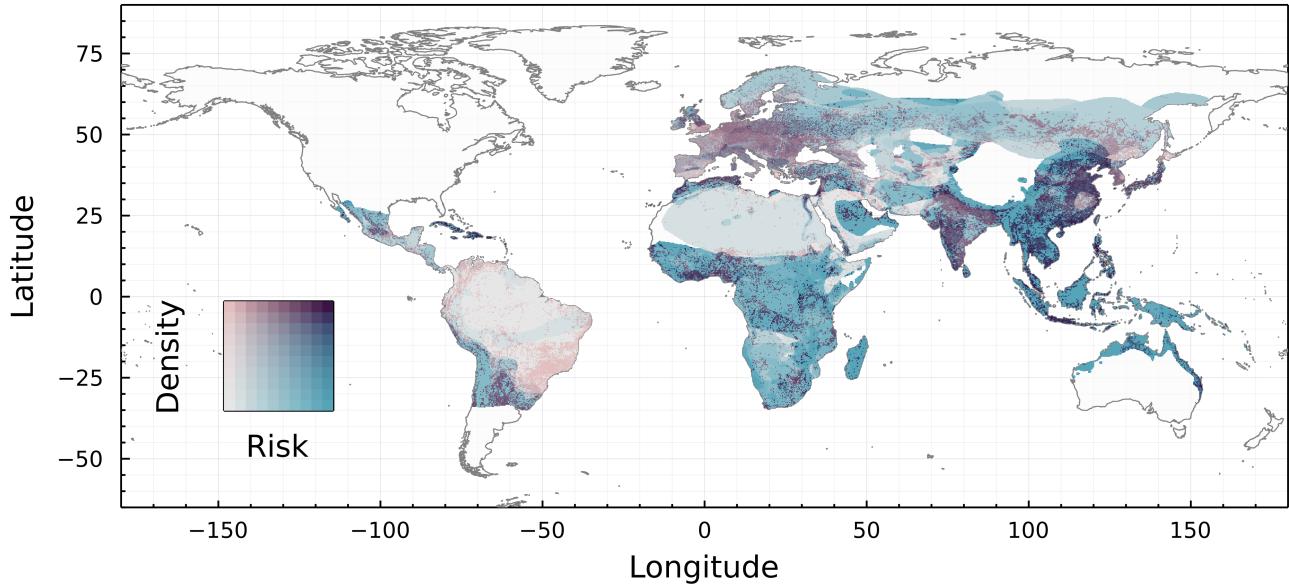


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the risk map. Darker pixels correspond to more risk, in that the GMTC-derived risk of fig. 4 is high *and* the pixel is densely occupied by human populations. This approach increases 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.