

# The coevolutionary mosaic of betacoronavirus emergence 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>

Daniel J. Becker<sup>6</sup> Colin J. Carlson<sup>7,8,9,‡</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, School of Veterinary Science, Massey University, New Zealand <sup>4</sup> Paul G. Allen School for Global Health, Washington State University, Pullman, WA, United States <sup>5</sup> Department of Biology, Georgetown University, Washington, DC, USA <sup>6</sup> Department of Biology, University of Oklahoma, Norman, OK, USA <sup>7</sup> Department of Biology, Georgetown University, Washington, DC,

<sup>8</sup> Center for Global Health Science and Security, Georgetown University Medical Center, Washington, DC, USA <sup>9</sup> Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, DC, USA

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

## Correspondance to:

Timothée Poisot — [timothee.poisot@umontreal.ca](mailto:timothee.poisot@umontreal.ca)

Driven by the need to understand the ecological factors involved in the emergence of *Betacoronavirus* (the genus causing the SARS, MERS, and COVID-19 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 is best understood through the multi-faceted prism of ecological and evolutionary mechanisms. Our framework reveals that host 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 and phylogeographic regions. 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 interaction network may facilitate the emergence of novel viruses and their spillover into human populations.

1 Disease emergence is complex, and is driven not only by animal-human contact, but also by the  
2 underlying evolutionary dynamics in viral reservoirs.<sup>1</sup> Although host richness is often used as a superficial  
3 proxy for spillover risk,<sup>2,3</sup> these approaches oversimplify the relevant interspecific heterogeneity in  
4 immunology, behavior, and other traits, and therefore overlook unique host pools that allow for the rapid  
5 evolution of highly divergent viruses.<sup>4</sup> In the case of generalist pathogens like betacoronaviruses, there is  
6 conceptual and empirical support to the idea that these community-level mechanisms are even more  
7 important,<sup>5</sup> particularly given that cross-species transmission may, as a rule, structure viral evolution  
8 more than co-divergence with hosts.<sup>6</sup> This creates a disconnect between coevolutionary theory (including  
9 empirical evidence from virology) and most existing ecological frameworks for mapping spillover risk.

10 The Geographic Mosaic Theory of Coevolution (GMTC) attempts to explicitly connect microevolutionary  
11 dynamics to the macroecology and biogeography of symbiotic interactions.<sup>7</sup> The GMTC posits that  
12 coevolutionary processes among pairs<sup>8</sup> or complexes<sup>9</sup> of species are structured in space by the rippling  
13 effects of abiotic conditions onto evolutionary mechanism, giving rise to fragmented systems with  
14 different structure and ecologically dynamics over large spatial extents.<sup>10</sup> The GMTC predicts a spatial  
15 fragmentation of coevolutionary dynamics under the joint action of three processes:<sup>11</sup> coevolutionary hot-  
16 and coldspots, which appear when the intensity of *interaction* (in terms of reciprocal fitness consequences)  
17 varies spatially; selection mosaics, wherein the intensity of *selection* varies across space, driven by both the  
18 biotic complexity of the community (locally diverse hosts and viruses are more biotically complex) and the  
19 local favorability of the environment;<sup>12</sup> and trait remixing, which occurs when coevolutionary dynamics  
20 are driven by the arrival (or departure) of *functional traits*, through changes in community composition  
21 due to invasions, meta-community dynamics, and dispersal.

22 Here, we apply the GMTC to explore and explain the global biogeography of betacoronaviruses, the group  
23 that includes SARS-CoV, MERS-CoV, and SARS-CoV-2. In their bat reservoirs, coronaviruses evolve  
24 through a mix of host jumps, recombination among disparate lineages, and, to a lesser degree,  
25 co-divergence with their hosts—<sup>2</sup>a mix of mechanisms that creates a complex and nonlinear relationship  
26 between host diversity and viral emergence. Working from a recently published database of bat hosts of  
27 betacoronaviruses, we develop the first global maps of both host and virus evolutionary distinctiveness  
28 and biogeographic regions for this system. Aiming to explain these patterns, we develop a generalized  
29 framework for applying the GMTC to host-virus interactions, with a specific emphasis on the potential to  
30 create independent coevolutionary dynamics (and therefore spatial fragmentation in risk) through

31 heterogeneity. We develop a trivariate risk assessment system that connects each GMTC mechanism to a  
32 quantifiable aspect of host-virus interactions: (i) viral sharing rates in host communities, representing the  
33 strength of potential interaction between viruses and any one host (i.e., places where viruses undergo  
34 constant host switching may be coevolutionary coldspots); (ii) the phylogenetic diversity of hosts, as a  
35 proxy for variation in the immunological mechanisms that antagonize viruses (i.e., the selection mosaic);  
36 and (iii) the local uniqueness of the bat community, representing the potential for viruses to be exposed to  
37 novel host traits (e.g., variation in receptor sequences). Together, we argue that these can be used to  
38 identify and map the evolutionary drivers that—in conjunction with transmission processes (e.g., viral  
39 prevalence in reservoirs and animal-human contact rates)—determine disease emergence risk.

## 40 Results and Discussion

### 41 Hotspots of host richness and viral diversification are distinct

42 Bats, the second most diverse group of mammals, are found worldwide; gradients in their species  
43 richness generally track broader patterns of mammal diversity, with a striking Neotropical hotspot  
44 (especially in the Amazon basin) and a secondary hotspot centered in the southeast Asian peninsula.  
45 These hotspots of bat diversity are generally presumed to be hotspots of viral adaptive radiation, and  
46 therefore areas of concern for human health.<sup>2,13</sup> However, the hotspots of bat betacoronavirus reservoirs  
47 show a distinct pattern, with primary hotspots (both in terms of size and higher values) of host richness  
48 situated primarily southeast Asia, parts of southern Europe, and to a lesser extent parts of Africa in the  
49 -25-0 range of latitudes (fig. 1; top). Although hundreds of species likely host undiscovered  
50 betacoronaviruses, machine learning predictions have suggested that these undiscovered reservoirs should  
51 follow the same diversity gradient.<sup>14</sup> In principle, these hotspots of locally-diverse, virus-rich bat  
52 communities should drive more adaptive diversification in their viruses.

53 [Figure 1 about here.]

54 However, we find that the global pattern of betacoronavirus phylogenetic distinctiveness is quite distinct  
55 from both bat host richness and phylogenetic distinctiveness (fig. 1; bottom). In contrast to the sparsity of  
56 Neotropical betacoronavirus hosts, South America has the most evolutionary distinct hosts *and* viruses,

57 followed by secondary hotspots in southeast Asia and the Rift Valley region have mostly distinct viruses.  
58 Some degree of sampling bias may contribute to these patterns: for example, South-America is one of the  
59 places where the fewest bat betacoronavirus sequences have been generated,<sup>2,13,15</sup> resulting in a sparser  
60 phylogenetic tree, and artificially inflating distinctiveness; conversely, disproportionate research effort in  
61 eastern China<sup>16</sup> may have led to a more complete inventory of the local diversity of coronaviruses, again  
62 inflating these metrics relative to underlying patterns. Even accounting for these potential biases, though,  
63 there is obvious heterogeneity in betacoronavirus evolutionary distinctiveness that is distinct from overall  
64 bat diversity.

65 On closer inspection, these patterns recapitulate the evolutionary history of both the order Chiroptera and  
66 the genus *Betacoronavirus*. Horseshoe bats (Rhinolophidae) are both the reservoirs of the SARS-like  
67 viruses (subgenus *Sarbecovirus*) and the presumed ancestral host of *Betacoronavirus*.<sup>17</sup> The hotspot of host  
68 richness and viral diversity in southeast Asia—both of which are disproportionately high, considering the  
69 global landscape of bat species richness—is almost entirely driven by adaptive radiation in this clade<sup>3,14</sup>.  
70 In contrast, the Neotropical hotspot of viral distinctiveness is the result of isolation by vicariance for both  
71 bats and their viruses. Out of the four main groups of betacoronaviruses, only the subgenus *Merbecovirus*  
72 has been found in animals in the Americas—an introduction that is generally presumed to be ancient.  
73 While comparatively understudied, New World merbecoviruses have been found in the ghost-faced bats  
74 (Mormoopidae), New World leaf-nosed bats (Phyllostomidae), and free-tailed bats (Molossidae) (add cite:  
75 Olival 2020 PLoS Pathogens). The former two groups are endemic to the Neotropics, while the latter two  
76 and particularly the phyllostomids are responsible for the hotspot of bat diversity in the Amazon.  
77 Together, these groups play host to a distinct regime of bat-betacoronavirus coevolution in the Americas.

## 78 The phylogeographic regions of hosts and their viruses overlap

79 Despite differences in hotspots of evolutionary distinctiveness, there are reasons to expect that the  
80 phylogeography of bats and betacoronaviruses should show some degree of congruence.<sup>18,19</sup> In particular,  
81 this should be the case if viruses can circulate among hosts and co-evolve with local hosts communities,  
82 making their evolutionary process more than a byproduct of host evolution. High density of hosts sharing  
83 the same virus (albeit possibly different strains) can drive or result from evolution of the bat antiviral  
84 immune system, resulting in spatially distinct immunological responses, as evidenced in several bat  
85 species.<sup>20</sup> Immune characteristics that allow bats to be better adapted to infection by emerging viruses,<sup>21,22</sup>

86 in addition to being hardcoded in their genome,<sup>23</sup> may be related to a wide variety of diets,<sup>20,24</sup> themselves  
87 likely to be driven by spatial effects, especially at the local scale – bats, indeed, occupy a variety of  
88 environments, and therefore display a variety of adaptations to these environments.<sup>25</sup>

89 [Figure 2 about here.]

90 In fig. 2, we show a projection of the phylogeographic signal of bats (top) and viruses (bottom) in space;  
91 the distinct groupings (represented by different colors symbolizing positions in the subspace formed by  
92 the first two axes of the PCoA) are essentially equivalent between the two groups, and can be coarsely  
93 delineated as southeast Asia, Eurasia above a northing of 25, and Africa and south America. These results  
94 suggest that, although the evolutionary distinctiveness of the bat-betacoronaviruses complex varies  
95 spatially, the system shows an important degree of spatial consistency, with a reduced number of  
96 bioregions. Available information describing the spillover of zoonotic betacoronaviruses of bat origin  
97 where data was available before and up through the COVID-19 pandemic puts spillover events of  
98 SARS-CoV-2 in Wuhan, China; SARS-CoV in Guangdong, China based on the presence of closest known  
99 viruses circulating in nature, and a nearby location where serological (antibody) evidence has indicated  
100 human exposure to SARS-like viruses;<sup>26</sup> MERS-CoV in Saudi Arabia based on index cases available from a  
101 recently-published compendium of cases.<sup>27</sup> For the latest event, most if not all index cases are presumed to  
102 be camel-to-human transmission, and the precise origin point (if it exists) of MERS-CoV in bats is  
103 uncertain. Recent recombinant canine coronavirus spillover events in Haiti<sup>28</sup> and Europe<sup>29</sup> are not  
104 relevant here, as bats' involvement in these cycles of transmission have been supposed to be non-existent.  
105 These index cases fall within different phylogeographic bioregions (fig. 2), which further highlight the  
106 issue that different host-virus sub-systems may lead to widespread emergence.

## 107 **Coevolution-informed emergence risk is different in space**

108 As host richness, joint distinctiveness, or phylogeographic structure suggest that the bat-betacoronaviruses  
109 complex is globally fragmented enough to give rise to both different levels of risk (as evidenced by the  
110 spatial location of spillover events) and different types of co-evolutionary dynamics, we turn to the  
111 Geographic Mosaic Theory of Coevolution to provide a measure of risk accounting for multiple processes.  
112 In fig. 3, we overlapped three components of spillover risk: viral sharing, *i.e.* the chance that two bats will  
113 share viruses overall; Local Contribution to Beta Diversity, *i.e.* the fact that a bat community is

114 compositionally unique compared to the average compositional similarity across the entire system; finally,  
115 host phylogenetic diversity, *i.e.* how dispersed the bats in a location are within the tree of life. This  
116 approach leads to the definition of broad biogeographic regions of risk, where the same color represents  
117 the same type of risk. By way of contrast to figures fig. 1 and fig. 2, these regions do not necessarily  
118 overlap with previous spatial partitions of the bat-betacoronaviruses complex.

119 [Figure 3 about here.]

120 From the perspective of spillover risk, the most important combination of factors is a high phylogenetic  
121 diversity of hosts with low viral sharing; this, essentially, means that very different betacoronaviruses  
122 could co-exist within the same place. This is particularly the case given that betacoronaviruses often  
123 evolve and even achieve host shifts through recombination, which requires the co-occurrence of  
124 sufficiently distinct viruses to be a major driver of emergence. In fig. 3, this corresponds to yellow to pale  
125 green areas, which are essentially limited to South-Eastern Asia, and to some part of Sub-Saharan Africa.  
126 Adopting a geographic mosaic theory perspective on risk, other regions of the world are of lesser concern  
127 (fig. 4). Our risk decomposition does not account for viral diversity or distinctiveness. The simple  
128 rationale behind it is that the acquisition of viral data is rarely disconnected from the acquisition of host  
129 data. There are more sources of information on hosts than on viruses, allowing to develop a host-centric  
130 perspective on risk (although this estimate would be more accurate with viral traits related to *e.g.* ability to  
131 switch hosts or pathogenic potential). Areas with high bat diversity and high turnover *may* facilitate the  
132 evolutionary radiation of viruses, matching previous findings that the diversification of bat coronaviruses  
133 is driven largely by host shifts (inter-genus or higher levels of cross-species transmission) and, to a lesser  
134 degree, cospeciation and sharing, representing intra-genus cross-species transmission.<sup>2</sup> This  
135 diversification is not an actual risk factor for spillover itself, but acts downstream of a spillover event by  
136 increasing the random chance of the emergence of a virus with the raw genomic components required for  
137 the potential to infect humans.

138 [Figure 4 about here.]

139 From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide  
140 hotspots of *Betacoronavirus* risk through mixing of unique viruses (via codivergence) and in turn  
141 recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat

142 species are endemic following evolutionary divergence from sister species in both African and Asian  
143 continents.<sup>30</sup> Recent surveillance<sup>31</sup> has identified a novel *Betacoronavirus* (in the subgenus *Nobecovirus*) in  
144 Madagascar-endemic pteropid bat species (*Pteropus rufus*, *Rousettus madagascariensis*), emphasizing  
145 strong proof of principle in model predictions.

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

147 Based on the previous result, we extracted the risk component from the composite map (see Methods), to  
148 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. As this map  
149 represents the potential risk, it must be weighed by the potential for contacts with humans. As a proxy for  
150 this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a reasonable  
151 proxy for the density of humans per unit area, which increases the probability of pathogen spread more  
152 widely.<sup>32</sup> Since human activity is required to amplify the frequency of virus encounters and thus create  
153 areas of viral amplification, mapping the potential risk against measures of land use is required to generate  
154 a more actionable assessment of risk. This map is presented in fig. 5. Most of South America and Europe  
155 are at comparatively lower risk, as although densely populated, settlements tend to be in areas with lower  
156 potential risk. Regions like Malaysia and the North coast of Australia have a high risk component, but  
157 should represent a relatively lower effective risk due to low human density. However, this mapping reveals  
158 that South-East Asia, the Indian subcontinent, and parts of sub-Saharan Africa, are at high risk due to the  
159 overlap between built areas and bat communities representing more opportunities for cross-species  
160 transmission of betacoronaviruses. In looking for the origins of SARS in China,<sup>33</sup> present serological  
161 evidence that strongest human-animal contact results in higher risk of virus exposure, regardless of the  
162 animal species, but that different types of contact had different impacts. Ideally, finer-grained information  
163 about human activity (rather than human presence through anthropisation) could allow to partition this  
164 risk further, albeit at the cost of more hypotheses required to estimate the amount of risk represented by  
165 each activity. Our map of purported high risk/diversitifcation potential (Madagascar, South-America)  
166 overlay with sampling gaps for *Betacoronavirus*,<sup>16</sup> stressing the need for spatially targeted monitoring and  
167 discovery.

168

[Figure 5 about here.]

169 **Conclusion**

170 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to  
171 human health.<sup>18,19</sup> Chiropterans emerged around 64 million years ago and are one of the most diverse  
172 mammalian orders, with an estimated richness of more than 1400 species.<sup>34,35</sup> They exhibit a broad variety  
173 of habitat use, behaviour, and feeding strategies, putting them at key positions in the delivery and  
174 provisioning of several ecosystem services, tied to important ecosystem-derived benefits to human.<sup>36</sup> For  
175 example, bats are an essential component of many seed-dispersal networks.<sup>37</sup> Over two-thirds of bats are  
176 know to be either obligate or facultative insectivores, therefore actively contributing for agricultural pest  
177 control,<sup>38,39</sup> and vectors of pathogens that put a risk on human health.<sup>40,41</sup> Because bats are globally  
178 distributed and have a long evolutionary history, phylogeographic and biogeographic approaches are  
179 required to shed light on the contemporary distribution of coevolutionary processes between bats and the  
180 pathogens they host. Not all areas in which bats, viruses, and human are co-occurring are facing a risk of  
181 spillover towards human populations, and the areas in which this risk exist may not be facing risks of the  
182 same nature and magnitude.

183 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries  
184 like highly divergent nobecoviruses in Madagascar and the previously-neglected adaptive radiation of  
185 sarbecoviruses outside of southern China and throughout southeast Asia. In doing so, it advances  
186 ecological theory beyond the current state of the art for global maps of emergence risk. For example,  
187 previous studies that have used host richness as proxy have predicted a high diversity of unsampled bat  
188 viruses,<sup>13</sup> bat coronaviruses,<sup>2</sup> and even specifically betacoronaviruses<sup>14</sup> in both the Amazon and southeast  
189 Asia. While we find that both regions are characterized by highly divergent host and viral communities,  
190 our framework identifies key differences between the regions. We find that Latin America is a hotspot of  
191 both host and viral distinctiveness, suggesting that this branch of the bat-betacoronavirus complex may be  
192 undergoing independent evolutionary dynamics from the rest of the global pool, but with limited potential  
193 for viral diversification—a finding that is supported by previous work indicating a higher rate of  
194 codivergence in Latin America.<sup>2</sup> In contrast, in southeast Asia, host richness and viral distinctiveness are  
195 high but sharing is low; this suggests a different type of evolutionary dynamics that could generate high  
196 local diversity of viruses through host switching and viral recombination (see e.g.,<sup>17</sup> as well as the  
197 discovery of recombinant viruses that share genetic material from both the SARS-CoV and SARS-CoV-2

198 branches of the Sarbecovirus lineage).<sup>42</sup> Both of these regions are priority areas for sampling, especially  
199 given predictions that they contain many bat hosts of undiscovered betacoronaviruses.<sup>14,16</sup> However, both  
200 the evolutionary and ecological aspects of emergence risk are likely higher in southeast Asia—a fact that  
201 will only become more relevant, as bats track shifting climates and exchange viruses with other species,  
202 creating a hotspot of cross-species transmission unique to the region.<sup>43</sup>

203 The diversity and diversification potential of bats responds to anthropogenic factors others than shifting  
204 climates.<sup>44</sup> Land use changes could significantly decrease bat suitability, notably through effects on diet  
205 and availability of habitats.<sup>45</sup> As our results establish that the diversification of bats betacoronaviruses  
206 happens on top of processes affecting hosts, biogeographic variation in human population density and  
207 anthropogenic disturbances may feed into co-evolutionary dynamics. Increase in humans-hosts contacts  
208 also increase the risk of emergence of novel diseases,<sup>46</sup> so does the changes in landscape connectivity at  
209 local/regional scales.<sup>47</sup> This represents a challenge for both conservation strategies and disease ecology:  
210 some areas can a high emergence risk and more potential for the acquisition of zoonotic viruses through  
211 bat-human encounters.<sup>48</sup> In particular, the challenge ahead lies in the need to quantify actual exposure  
212 (and risk) accounting for several transmission scenarios, including both direct and indirect bat - human  
213 interactions, and feeding back into the provision of ecosystem services by bats.

214 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
215 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
216 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research  
217 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des  
218 Données (IVADO). This research was enabled in part by support provided by Calcul Québec  
219 ([www.calculquebec.ca](http://www.calculquebec.ca)) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)). NF is funded by the NSERC  
220 BIOS<sup>2</sup> CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by  
221 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation. DJB was  
222 supported by the National Institute of General Medical Sciences of the National Institutes of Health  
223 (P20GM134973).

224 **Methods**

225 **Known *Betacoronavirus* hosts**

226 We downloaded the data on bats hosts of *Betacoronavirus* from  
227 <https://www.viralemergence.org/betacov> on Apr. 2022,<sup>14</sup> and filtered it to “known” hosts (established  
228 before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling and competence  
229 assays since the initial data collection). The original database was assembled by a combination of data  
230 mining and literature surveys, including automated alerts on the “bats” and “coronavirus” keywords to  
231 identify novel empirical evidence of bats-betacoronaviruses associations; this yielded a total of 126 known  
232 hosts, 47 of which were novel hosts.

233 **Bat occurrences**

234 We downloaded the rangemap of every current bat species that was classified as an empirically  
235 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data.<sup>49</sup> The range  
236 maps were subsequently rasterized using the rasterize function from GDAL<sup>50</sup> at a resolution of  
237 approximately 100kmx100km. For every pixel in the resulting raster where at least one bat host of  
238 *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was used to  
239 calculate the following risk assessment components: bat phylogenetic diversity, bat compositional  
240 uniqueness, and predicted viral sharing risk.

241 **Bat phylogenetic diversity**

242 For every pixel, we measured Faith’s Phylogenetic Diversity<sup>51</sup> based on a recent synthetic tree with robust  
243 time calibration, covering about 6000 mammalian species.<sup>52</sup> Faith’s PD measures the sum of unique  
244 branches from an arbitrary root to a set of tips, and comparatively larger values indicate a more  
245 phylogenetic diverse species pool. We measured phylogenetic diversity starting from the root of the entire  
246 tree (and not from Chiroptera); this bears no consequences on the resulting values, since all branches  
247 leading up to Chiroptera are only counted one per species pool, and (as we explain when describing the  
248 assembly of the composite risk map), all individual risk components are ranged in [0,1]. This measure  
249 incorporates a richness component, which we chose not to correct for; the interpretation of the

250 phylogenetic diversity is therefore a weighted species richness, that accounts for phylogenetic  
251 over/under-dispersal in some places.

## 252 **Bat compositional uniqueness**

253 For every species pool, we measured its Local Contribution to Beta-Diversity;<sup>53</sup> LCBD works from a  
254 species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites are columns, and a value of  
255 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel represents a unique location,  
256 and following best practices<sup>54</sup> transformed it using Hellinger's distance to account for unequal bat  
257 richness at different pixels. The correction of raw community data is particularly important for two  
258 reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect  
259 of overall species richness, which is already incorporated in the phylogenetic diversity component. High  
260 values of LCBD indicate that the pixel has a community that is on average more dissimilar in species  
261 composition than what is expected knowing the entire matrix, i.e. a more unique community. Recent  
262 results by<sup>55</sup> shows that LCBD measures are robust with regards to spatial scale, and are therefore  
263 applicable at the global scale.

## 264 **Viral sharing between hosts**

265 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a  
266 previously published generalized additive mixed model of virus sharing by a tensor function of  
267 phylogenetic distance and geographic range overlap across mammals.<sup>56</sup> This network stores pairwise  
268 values of viral community similarity. To project viral sharing values into a single value for every pixel, we  
269 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant  
270 bat assemblage is likely to be proficient at exchanging viruses.

## 271 **Composite risk map**

272 To visualize the aggregated risk at the global scale, we combine the three individual risk components  
273 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model.<sup>57</sup> In  
274 this approach, every risk component gets assigned a component in the RGB color model (phylogenetic  
275 diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid

276 RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no  
277 phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for  
278 each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the  
279 risk as colors diverge towards combinations of values for three risk components. Out of the possible  
280 combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic  
281 diversity and low viral sharing,<sup>58</sup> in that this allows multiple independent host-virus coevolutionary  
282 dynamics to take place in the same location. In the colorimetric space, this corresponds to yellow – because  
283 the HSV space is more amenable to calculations for feature extraction,<sup>59</sup> we measured the risk level by  
284 calculating the angular distance of the hue of each pixel to a reference value of 60 (yellow), and weighted  
285 this risk level by the value component. Specifically, given a pixel with colorimetric coordinates  $(h, s, v)$ , its  
286 ranged weighted risk value is

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

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

## 288 Viral phyogeography and evolutionary diversification

289 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed  
290 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data  
291 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR  
292 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR  
293 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated  
294 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained  
295 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and  
296 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East  
297 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus  
298 RdRp sequences were then aligned using MAFFT<sup>60</sup> v1.4.0 (Algorithm FFT-NS-2, Scoring matrix 200PAM /  
299 k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree reconstructed in  
300 IQ-TREE<sup>61</sup> v1.6.12 with ModelFinder<sup>62</sup> ultrafast bootstrap approximation<sup>63</sup> with a general time reversible  
301 model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of nucleotide

302 substitution (GTR+F+R5).

303 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat  
304 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the  
305 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary  
306 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary  
307 distinctiveness<sup>64</sup> for each of the viruses in the tree, then averaged these at the bat species level, projected  
308 these values onto their geographic distributions, and averaged across every bat found in a given pixel. As  
309 such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral  
310 community believed to be associated with a particular subset of bats present.

### 311 **Co-distribution of hosts and viral hotspots**

312 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
313 biogeography of their hosts. To test this idea, we loosely adapted a method from,<sup>65,66</sup> who proposed a  
314 phylogenetic method for the delineation of animal biogeographic regions. In their original method, a  
315 distance matrix - where each row or column represents a geographic raster’s grid cell, and the dissimilarity  
316 values are the “beta diversity similarity” of their community assemble - undergoes non-metric  
317 multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a  
318 four-color bivariate map. Here, we build on this idea with an entirely novel methodology. First, we  
319 measure the phylogenetic distance between the different viruses in the betacoronaviruses tree by using the  
320 cophenetic function in ape;<sup>67</sup> subsequently, we take a principal components analysis of that distance  
321 matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional  
322 space. We then take the first two principal components and, as with the evolutionary distinctiveness  
323 analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

324 **References**

- 325 1.
- 326 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 327 2.
- 328 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 329 3.
- 330 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 331 4.
- 332 Agosta, S. J., Janz, N. & Brooks, D. R. How specialists can be generalists: Resolving the ‘parasite paradox’ and implications for emerging infectious disease. *Zoologia (Curitiba)* **27**, 151–162 (2010).
- 333 5.
- 334 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 335 6.
- 336 Geoghegan, J. L., Duchêne, S. & Holmes, E. C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLOS Pathogens* **13**, e1006215 (2017).
- 337 7.
- 338 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 339 8.
- 340 Thompson, J. N. *The Coevolutionary Process*. (University Of Chicago Press, 1994).
- 341 9.
- 342 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 343 10.
- 344 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).  
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 345 11.
- 346 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).

- 347 12.
- 348 Thrall, P. H., Hochberg, M. E., Burdon, J. J. & Bever, J. D. Coevolution of symbiotic mutualists and parasites  
in a community context. *Trends in Ecology & Evolution* **22**, 120–126 (2007).
- 349 13.
- 350 Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
- 351 14.
- 352 Becker, D. J. *et al.* Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The  
Lancet Microbe* (2022) doi:[10.1016/S2666-5247\(21\)00245-7](https://doi.org/10.1016/S2666-5247(21)00245-7).
- 353 15.
- 354 Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications* **8**,  
(2017).
- 355 16.
- 356 Cohen, L. E., Fagre, A. C., Chen, B., Carlson, C. J. & Becker, D. J. Sampling strategies and pre-pandemic  
surveillance gaps for bat coronaviruses. 2022.06.15.496296 (2022) doi:[10.1101/2022.06.15.496296](https://doi.org/10.1101/2022.06.15.496296).
- 357 17.
- 358 Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *bioRxiv: The Preprint  
Server for Biology* 2020.05.31.116061 (2020) doi:[10.1101/2020.05.31.116061](https://doi.org/10.1101/2020.05.31.116061).
- 359 18.
- 360 Van Brussel, K. & Holmes, E. C. Zoonotic disease and virome diversity in bats. *Current Opinion in Virology*  
**52**, 192–202 (2022).
- 361 19.
- 362 Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover  
and emergence. *Nature Reviews Microbiology* **18**, 461–471 (2020).
- 363 20.
- 364 Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Frontiers in Immunology* **11**, 26 (2020).
- 365 21.
- 366 Gorbunova, V., Seluanov, A. & Kennedy, B. K. The World Goes Bats: Living Longer and Tolerating Viruses.  
*Cell Metabolism* **32**, 31–43 (2020).
- 367 22.
- 368 Irving, A. T., Ahn, M., Goh, G., Anderson, D. E. & Wang, L.-F. Lessons from the host defences of bats, a  
unique viral reservoir. *Nature* **589**, 363–370 (2021).
- 369 23.
- 370 Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).

- 371 24.
- 372 Moreno Santillán, D. D. *et al.* Large-scale genome sampling reveals unique immunity and metabolic  
adaptations in bats. *Molecular Ecology* **mec.16027** (2021) doi:[10.1111/mec.16027](https://doi.org/10.1111/mec.16027).
- 373 25.
- 374 Muylaert, R. L. *et al.* Present and future distribution of bat hosts of sarbecoviruses: Implications for  
conservation and public health. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220397 (2022).
- 375 26.
- 376 Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.  
*Virologica Sinica* **33**, 104–107 (2018).
- 377 27.
- 378 Ramshaw, R. E. *et al.* A database of geopositioned Middle East Respiratory Syndrome Coronavirus  
occurrences. *Scientific Data* **6**, 318 (2019).
- 379 28.
- 380 Lednicky, J. A. *et al.* Isolation of a Novel Recombinant Canine Coronavirus from a Visitor to Haiti: Further  
Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. *Clinical Infectious Diseases: An  
Official Publication of the Infectious Diseases Society of America* **ciab924** (2021) doi:[10.1093/cid/ciab924](https://doi.org/10.1093/cid/ciab924).
- 381 29.
- 382 Vlasova, A. N. *et al.* Animal alphacoronaviruses found in human patients with acute respiratory illness in  
different countries. *Emerging Microbes & Infections* **11**, 699–702 (2022).
- 383 30.
- 384 Shi, J. J. *et al.* A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with Evidence for  
Widespread Panmixia. *Acta Chiropterologica* **16**, 279–292 (2014).
- 385 31.
- 386 Kettenburg, G. *et al.* Full Genome Nobcovirus Sequences From Malagasy Fruit Bats Define a Unique  
Evolutionary History for This Coronavirus Clade. *Frontiers in Public Health* **10**, (2022).
- 387 32.
- 388 Hazarie, S., Soriano-Paños, D., Arenas, A., Gómez-Gardeñes, J. & Ghoshal, G. Interplay between population  
density and mobility in determining the spread of epidemics in cities. *Communications Physics* **4**, 191 (2021).
- 389 33.
- 390 Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10**, 1030–1037  
(2004).
- 391 34.
- 392 Peixoto, F. P., Braga, P. H. P. & Mendes, P. A synthesis of ecological and evolutionary determinants of bat  
diversity across spatial scales. *BMC ecology* **18**, 18 (2018).

- 393 35.
- 394 Simmons, N. B. & Cirranello, A. L. Bat Species of the World: A taxonomic and geographic database. <https://batnames.org/> (2020).
- 395 36.
- 396 Kasso, M. & Balakrishnan, M. Ecological and Economic Importance of Bats (Order Chiroptera). *ISRN Biodiversity* **2013**, e187415 (2013).
- 397 37.
- 398 Mello, M. A. R. *et al.* The Missing Part of Seed Dispersal Networks: Structure and Robustness of Bat-Fruit Interactions. *PLOS ONE* **6**, e17395 (2011).
- 399 38.
- 400 *Bats in the Anthropocene: Conservation of Bats in a Changing World.* (Springer International Publishing, 2016). doi:[10.1007/978-3-319-25220-9](https://doi.org/10.1007/978-3-319-25220-9).
- 401 39.
- 402 Williams-Guillén, K., Perfecto, I. & Vandermeer, J. Bats Limit Insects in a Neotropical Agroforestry System. *Science* **320**, 70–70 (2008).
- 403 40.
- 404 Gonsalves, L., Bicknell, B., Law, B., Webb, C. & Monamy, V. Mosquito Consumption by Insectivorous Bats: Does Size Matter? *PLOS ONE* **8**, e77183 (2013).
- 405 41.
- 406 Gonsalves, L., Lamb, S., Webb, C., Law, B. & Monamy, V. Do mosquitoes influence bat activity in coastal habitats? *Wildlife Research* **40**, 10–24 (2013).
- 407 42.
- 408 Wu, Z. *et al.* A comprehensive survey of bat sarbecoviruses across China for the origin tracing of SARS-CoV and SARS-CoV-2. (2021) doi:[10.21203/rs.3.rs-885194/v1](https://doi.org/10.21203/rs.3.rs-885194/v1).
- 409 43.
- 410 Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk. *Nature* 1–1 (2022) doi:[10.1038/s41586-022-04788-w](https://doi.org/10.1038/s41586-022-04788-w).
- 411 44.
- 412 Alves, D. M. C. C., Diniz-Filho, J. A. F., da Silva e Souza, K., Gouveia, S. F. & Villalobos, F. Geographic variation in the relationship between large-scale environmental determinants and bat species richness. *Basic and Applied Ecology* **27**, 1–8 (2018).
- 413 45.
- 414 Treitler, J. T., Heim, O., Tschapka, M. & Jung, K. The effect of local land use and loss of forests on bats and nocturnal insects. *Ecology and Evolution* **6**, 4289–4297 (2016).

- 415 46.
- 416 Johnson, C. K. *et al.* Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20192736 (2020).
- 417 47.
- 418 Gryseels, S. *et al.* When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure in the Spatial Spread of Arenaviruses. *PLOS Pathogens* **13**, e1006073 (2017).
- 419 48.
- 420 Amman, B. R. *et al.* *Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and public health interest.* (FAO, 2011).
- 421 49.
- 422 IUCN. The IUCN Red List of Threatened Species. <https://www.iucnredlist.org> (2021).
- 423 50.
- 424 Rouault, E. *et al.* *GDAL/OGR Geospatial Data Abstraction software Library.* (Zenodo, 2022). doi:[10.5281/ZENODO.5884351](https://doi.org/10.5281/ZENODO.5884351).
- 425 51.
- 426 Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10 (1992).
- 427 52.
- 428 Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology* **17**, e3000494 (2019).
- 429 53.
- 430 Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters* **16**, 951–963 (2013).
- 431 54.
- 432 Legendre, P. & Condit, R. Spatial and temporal analysis of beta diversity in the Barro Colorado Island forest dynamics plot, Panama. *Forest Ecosystems* **6**, 7 (2019).
- 433 55.
- 434 Dansereau, G., Legendre, P. & Poisot, T. Evaluating ecological uniqueness over broad spatial extents using species distribution modelling. *Oikos* **n/a**, e09063 (2022).
- 435 56.
- 436 Albery, G. F., Eskew, E. A., Ross, N. & Olival, K. J. Predicting the global mammalian viral sharing network using phylogeography. *Nature Communications* **11**, 2260 (2020).
- 437 57.

- 438 Seekell, D. A., Lapierre, J.-F. & Cheruvelil, K. S. A geography of lake carbon cycling. *Limnology and Oceanography Letters* **3**, 49–56 (2018).
- 439 58.
- 440 Gomulkiewicz, R., Thompson, J. N., Holt, R. D., Nuismer, S. L. & Hochberg, M. E. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. *The American Naturalist* **156**, 156–174 (2000).
- 441 59.
- 442 Keke, S., Peng, Z. & Guohui, L. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. in *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* vol. 2 612–615 (2010).
- 443 60.
- 444 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
- 445 61.
- 446 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).
- 447 62.
- 448 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589 (2017).
- 449 63.
- 450 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
- 451 64.
- 452 Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. *PLOS ONE* **2**, e296 (2007).
- 453 65.
- 454 Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**, 5925–5930 (2007).
- 455 66.
- 456 Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* **37**, 2029–2053 (2010).
- 457 67.
- 458 Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

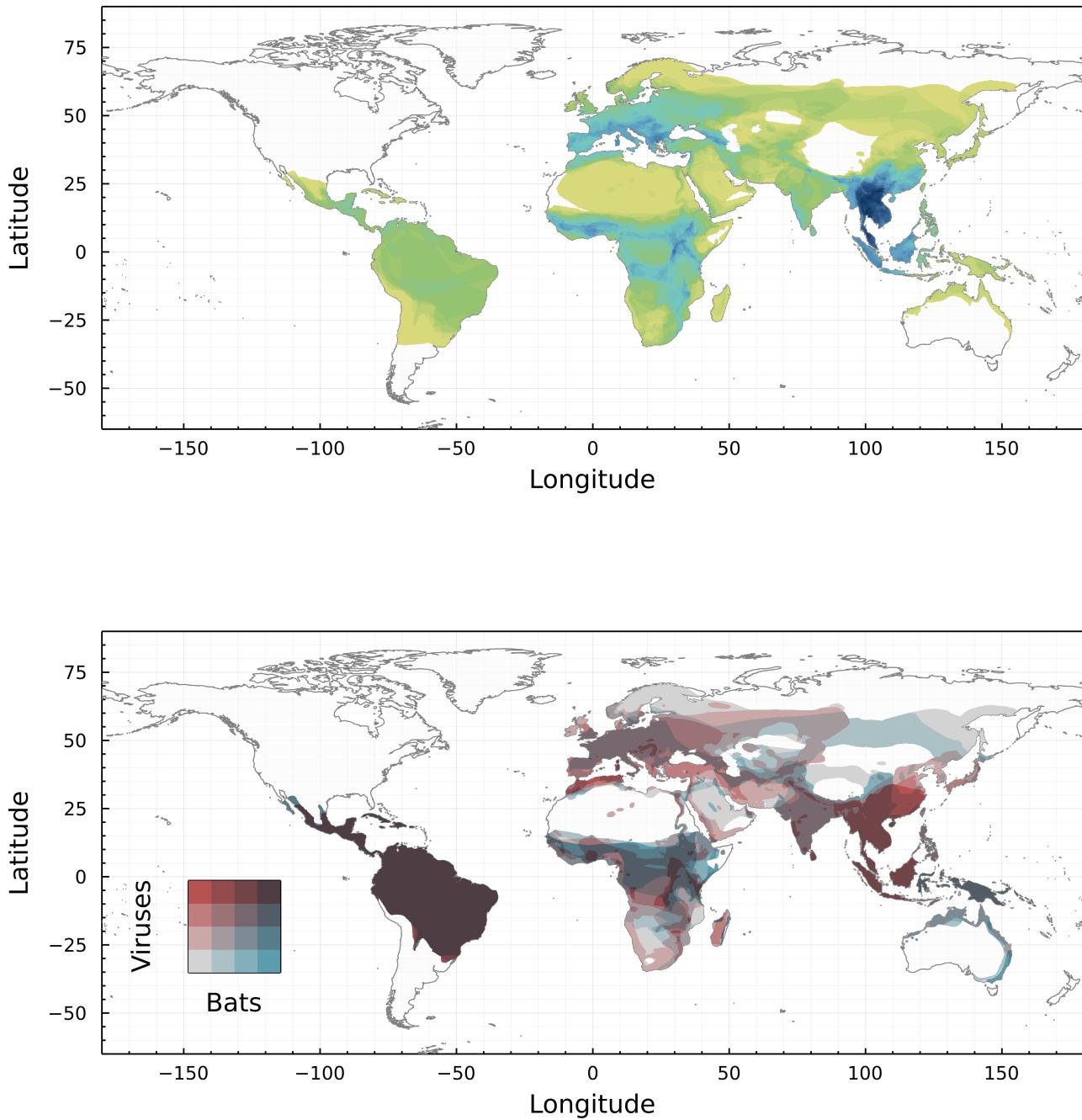


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).

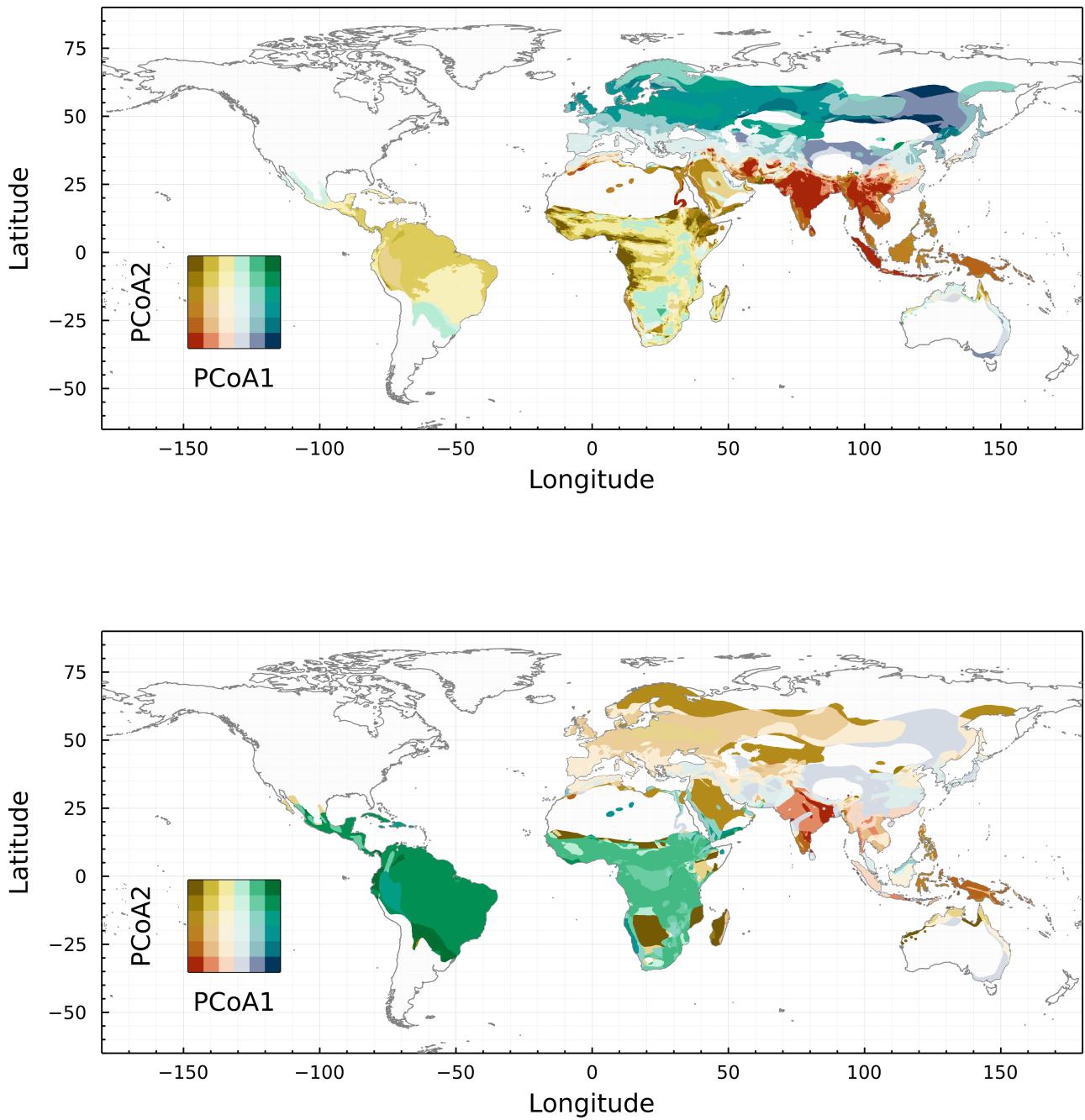


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

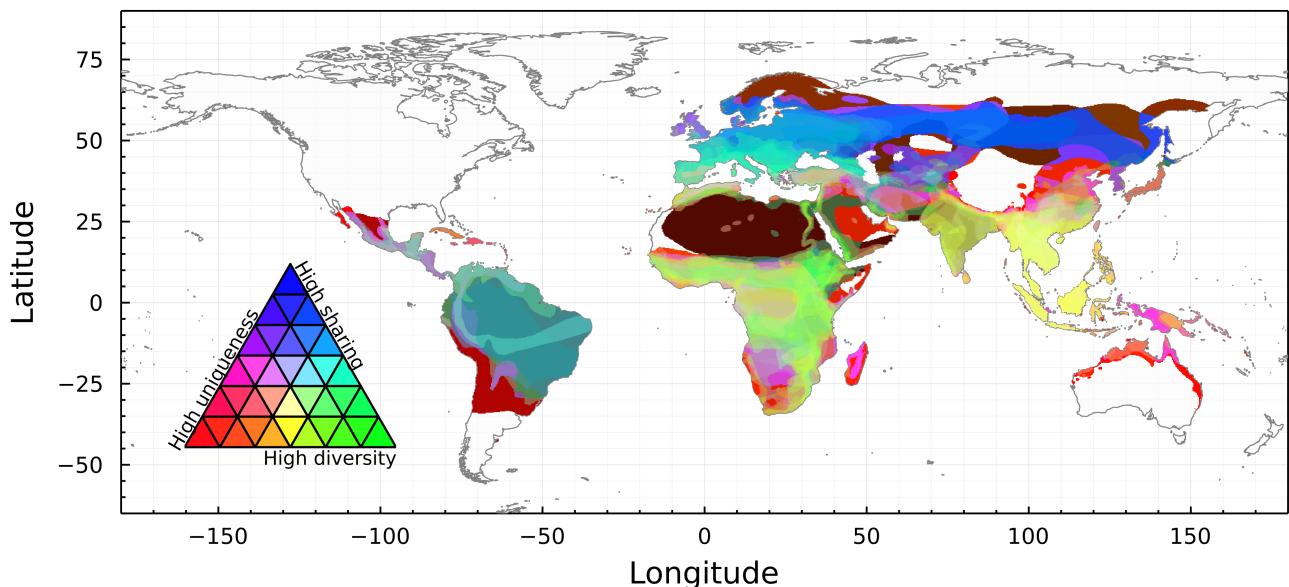


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible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

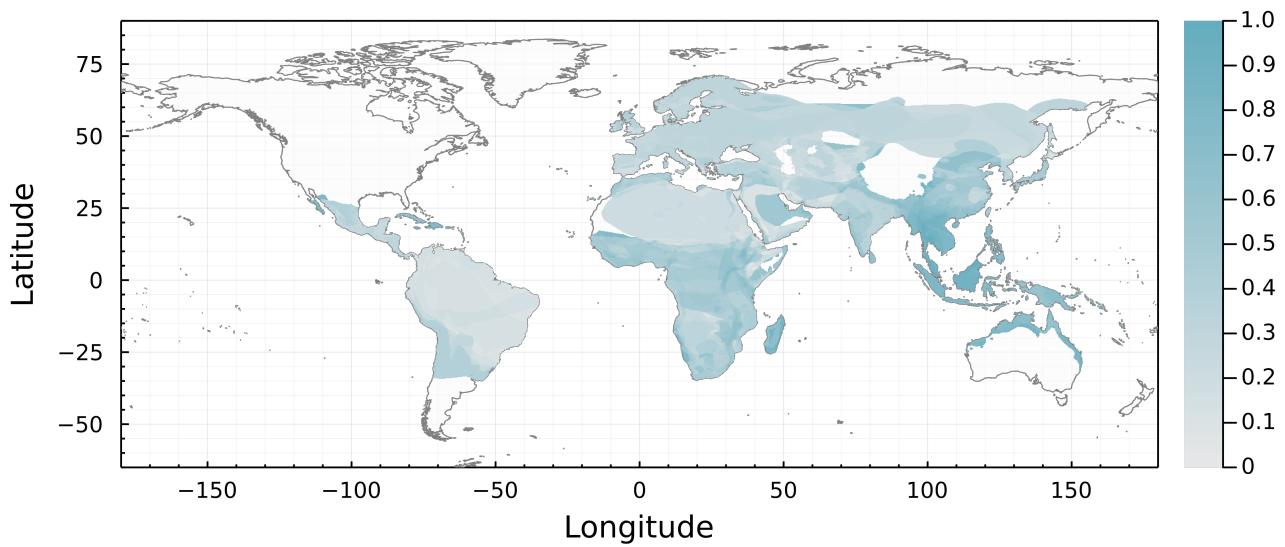


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

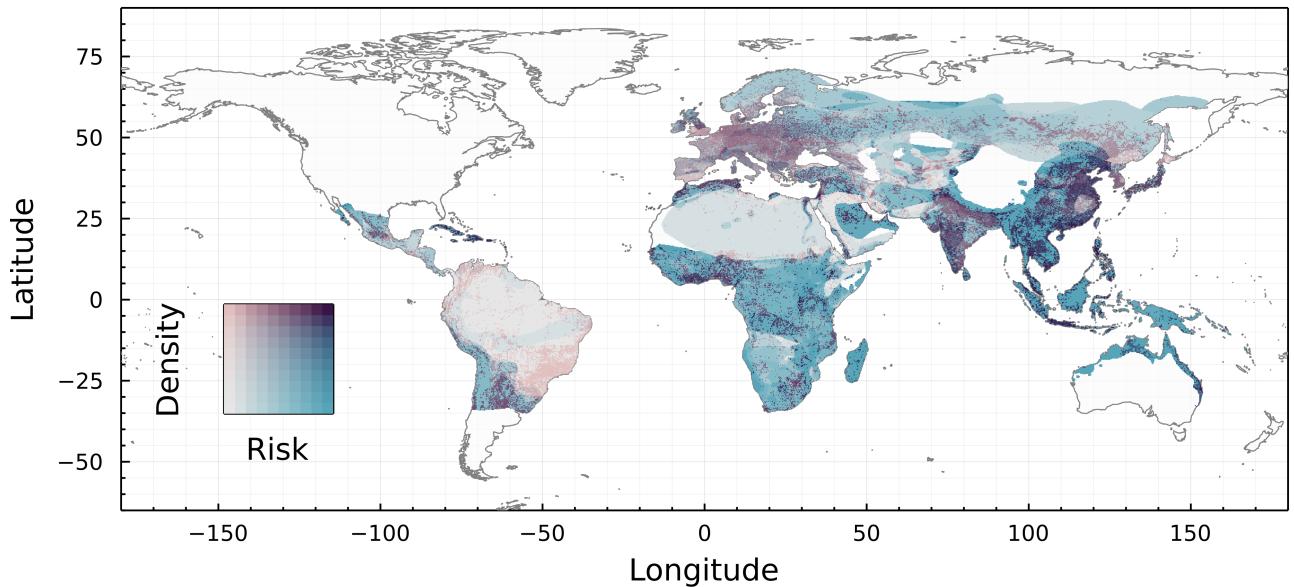


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk, 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 Arabian peninsula where areas of high to moderate risk overlap with areas of denser population.