

# The coevolutionary mosaic of bat 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

‡ 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, 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 and most  
9 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 test whether spatial structure in bat-betacoronavirus coevolution is identifiable at a  
28 global scale. Aiming to explain these patterns, we develop a generalized framework for applying the  
29 GMTC to host-virus interactions, with a specific emphasis on the potential to create independent  
30 coevolutionary dynamics (and therefore spatial fragmentation in risk) through heterogeneity. We develop

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

## 40 Results and Discussion

### 41 Bat and betacoronavirus biogeography are broadly consistent

42 Most previous work has assumed that the presence or richness of key groups of bat hosts are predictive of  
43 coronavirus diversity.<sup>2,3</sup> Projecting bat and betacoronavirus phylogeny over space (fig. 1), we find support  
44 for the idea that bat community assembly is directly responsible for a global mosaic of viral evolution. The  
45 distinct groupings (represented by different colors, symbolizing positions in a subspace formed by the first  
46 two phylogenetic principal components) are essentially equivalent between the two groups, and can be  
47 coarsely delineated as (1) south and southeast Asia, (2) east Asia (including northern China), west Asia,  
48 and the Mediterranean coast; (3) Eurasia above a northing of 40; and (4) Africa and south America. In  
49 some cases, this diverges from expectations about coronavirus biogeography: for example, previous work  
50 has rarely flagged India as a region of interest, but for both bats and betacoronaviruses, the subcontinent  
51 falls into the same regions as the southeast Asian peninsula (and indeed, the region is home to known bat  
52 hosts of nobecoviruses, sarbecoviruses, and merbecoviruses).<sup>3</sup>

53 [Figure 1 about here.]

54 Overall, these results suggest that the boundaries of bat and betacoronavirus biogeographic regions are  
55 largely consistent. This may be surprising, given that cospeciation plays a minor role in coronavirus  
56 diversification,<sup>2</sup> a property that would theoretically allow for substantial broad divergence in their

57 biogeography. However, host jumps at the family level or higher are relatively rare and significant events  
58 in coronavirus evolutionary history;<sup>2,13</sup> as a result, the mosaic of betacoronavirus phylogeography is  
59 assembled from a set of overlapping smaller coevolutionary systems, superimposed in space and filtered  
60 by the importance of different subgroups in local host communities. For example, the most speciose and  
61 cosmopolitan family of bats, the vesper bats (Vespertilionidae), are considered the primary reservoir of  
62 merbecoviruses;<sup>3,13</sup> but in the Americas, where merbecoviruses are the only lineage present, they have  
63 only been found in other bat taxa. At the coarsest scale, these heterogeneities are lost, and betacoronavirus  
64 biogeography tracks the deep rifts in bat evolutionary history—but within broad regions, the component  
65 coevolutionary systems may have very different dynamics.

66 **Hotspots of bat and betacoronavirus biodiversity are distinct**

67 Bats, the second most diverse groups of mammals, are found worldwide; gradients in their species  
68 richness generally track broader patterns of mammal diversity, with a striking Neotropical hotspot  
69 (especially in the Amazon basin) and a secondary hotspot centered in the southeast Asian peninsula.  
70 These hotspots of bat diversity are generally presumed to be hotspots of viral adaptive radiation, and  
71 therefore areas of concern for human health.<sup>2,14</sup> However, the hotspots of known bat betacoronavirus  
72 hosts show a distinct pattern, with primary hotspots (both in terms of size and higher values) of host  
73 richness situated in southeast Asia, parts of southern Europe, and to a lesser extent parts of Africa in the  
74 -25-0 range of latitudes (fig. 2; top). Although hundreds of species likely host undiscovered  
75 betacoronaviruses, machine learning predictions have suggested that these undiscovered reservoirs should  
76 follow the same diversity gradient.<sup>15</sup> In principle, these hotspots of locally-diverse, virus-rich bat  
77 communities should drive more adaptive diversification in their viruses.

78 [Figure 2 about here.]

79 However, we find that the global pattern of betacoronavirus phylogenetic distinctiveness is quite distinct  
80 from both bat host richness and phylogenetic distinctiveness (fig. 2; bottom). In contrast to the sparsity of  
81 Neotropical betacoronavirus hosts, South America has the most evolutionary distinct hosts *and* viruses,  
82 followed by secondary hotspots in southeast Asia and the Rift Valley region have mostly distinct viruses.  
83 Some degree of sampling bias may contribute to these patterns: for example, South-America is one of the  
84 places where the fewest bat betacoronavirus sequences have been generated,<sup>2,14,16</sup> resulting in a sparser

85 phylogenetic tree, and artificially inflating distinctiveness; conversely, disproportionate research effort in  
86 eastern China<sup>17</sup> may have led to a more complete inventory of the local diversity of coronaviruses, again  
87 inflating these metrics relative to underlying patterns. Even accounting for these potential biases, though,  
88 there is obvious heterogeneity in betacoronavirus evolutionary distinctiveness that is distinct from overall  
89 bat diversity.

90 Overall, these patterns recapitulate the evolutionary history of both the order Chiroptera and the genus  
91 *Betacoronavirus*. Horseshoe bats (Rhinolophidae) are both the reservoirs of the SARS-like viruses  
92 (subgenus *Sarbecovirus*), the group of pandemic threats that have been of the greatest interest to  
93 researchers<sup>13</sup> (and so have been sampled most intensively).<sup>17</sup> The hotspots of host richness and viral  
94 diversity in southeast Asia—both of which are disproportionately high, considering the global landscape  
95 of bat species richness—are almost entirely driven by viral adaptive radiation through host switching  
96 within this clade<sup>3,15</sup>. In contrast, the Neotropical hotspot of viral distinctiveness is driven by isolation by  
97 host vicariance. Out of the four main groups of betacoronaviruses, only the subgenus *Merbecovirus*  
98 (MERS-like viruses) has been found in animals in the Americas—an introduction that is generally  
99 presumed to be ancient.<sup>3</sup> While comparatively understudied, New World merbecoviruses have been found  
100 in the ghost-faced bats (Mormoopidae), New World leaf-nosed bats (Phyllostomidae), and free-tailed bats  
101 (Molossidae) (add cite: Olival 2020 PLoS Pathogens). The former two groups are endemic to the  
102 Neotropics, while the explosive adaptive radiations of the latter two (and particularly the phyllostomids)  
103 are responsible for the hotspot of bat diversity in the Amazon. Together, these clades of New World bats  
104 play host to a distinct regime of betacoronavirus coevolution.

## 105 Coevolution-informed emergence risk is different in space

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

114 approach leads to the definition of broad biogeographic regions of risk, where the same color represents  
115 the same type of risk. By way of contrast to figures fig. 2 and fig. 1, these regions do not necessarily  
116 overlap with previous spatial partitions of the bat-betacoronaviruses complex.

117 [Figure 3 about here.]

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

136 [Figure 4 about here.]

137 From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide  
138 hotspots of *Betacoronavirus* risk through mixing of unique viruses (via codivergence) and in turn  
139 recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat  
140 species are endemic following evolutionary divergence from sister species in both African and Asian  
141 continents.<sup>18</sup> Recent surveillance<sup>19</sup> has identified a novel *Betacoronavirus* (in the subgenus *Nobecovirus*) in

142 Madagascar-endemic pteropid bat species (*Pteropus rufus*, *Rousettus madagascariensis*), emphasizing  
143 strong proof of principle in model predictions.

144 **Human landscapes filter the geography of emergence risk**

145 The relationship between the underlying pathogen pool and emergence risk is mediated by  
146 human-wildlife contact and the risk that spillovers become epidemics. As a proxy for both, we finally  
147 overlaid the risk component from the composite map (see above) with the proportion of built land, as a  
148 proxy for a mix of habitat disturbance, potential for bat synanthropy, and human population density and  
149 connectivity<sup>1,20</sup> (add: Muylaert 2022 Proc B; Hassell 2017 TREE) (fig. 5).

150 Most of South America and Europe are at comparatively lower risk, as although densely populated,  
151 settlements tend to be in areas with lower potential risk. Regions like Malaysia and the North coast of  
152 Australia have a high risk component, but should represent a relatively lower effective risk due to low  
153 human density. However, this mapping reveals that South-East Asia, the Indian subcontinent, and parts of  
154 sub-Saharan Africa, are at high risk due to the overlap between built areas and bat communities  
155 representing more opportunities for cross-species transmission of betacoronaviruses. In looking for the  
156 origins of SARS in China,<sup>21</sup> present serological evidence that strongest human-animal contact results in  
157 higher risk of virus exposure, regardless of the animal species, but that different types of contact had  
158 different impacts. Ideally, finer-grained information about human activity (rather than human presence  
159 through anthropisation) could allow to partition this risk further, albeit at the cost of more hypotheses  
160 required to estimate the amount of risk represented by each activity. Our map of purported high  
161 risk/diversification potential (Madagascar, South-America) overlay with sampling gaps for  
162 *Betacoronavirus*,<sup>17</sup> stressing the need for spatially targeted monitoring and discovery.

163 [Figure 5 about here.]

164 PUT THIS SOMEWHERE: Available information describing the spillover of zoonotic betacoronaviruses of  
165 bat origin where data was available before and up through the COVID-19 pandemic puts spillover events  
166 of SARS-CoV-2 in Wuhan, China; SARS-CoV in Guangdong, China based on the presence of closest  
167 known viruses circulating in nature, and a nearby location where serological (antibody) evidence has  
168 indicated human exposure to SARS-like viruses;<sup>22</sup> MERS-CoV in Saudi Arabia based on index cases

available from a recently-published compendium of cases.<sup>23</sup> For the latest event, most if not all index cases are presumed to be camel-to-human transmission, and the precise origin point (if it exists) of MERS-CoV in bats is uncertain. Recent recombinant canine coronavirus spillover events in Haiti<sup>24</sup> and Europe<sup>25</sup> are not relevant here, as bats' involvement in these cycles of transmission have been supposed to be non-existent. These index cases fall within different phylogeographic bioregions (fig. 1), which further highlight the issue that different host-virus sub-systems may lead to widespread emergence.

## Conclusion

Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to human health.<sup>26,27</sup> Chiropterans emerged around 64 million years ago and are one of the most diverse mammalian orders, with an estimated richness of more than 1400 species.<sup>28,29</sup> They exhibit a broad variety of habitat use, behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of several ecosystem services, tied to important ecosystem-derived benefits to human.<sup>30</sup> For example, bats are an essential component of many seed-dispersal networks.<sup>31</sup> Over two-thirds of bats are known to be either obligate or facultative insectivores, therefore actively contributing for agricultural pest control,<sup>32,33</sup> and vectors of pathogens that put a risk on human health.<sup>34,35</sup> Because bats are globally distributed and have a long evolutionary history, phylogeographic and biogeographic approaches are required to shed light on the contemporary distribution of coevolutionary processes between bats and the pathogens they host. Not all areas in which bats, viruses, and human are co-occurring are facing a risk of spillover towards human populations, and the areas in which this risk exist may not be facing risks of the same nature and magnitude.

Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries like highly divergent nobecoviruses in Madagascar and the previously-neglected adaptive radiation of sarbecoviruses outside of southern China and throughout southeast Asia. In doing so, it advances ecological theory beyond the current state of the art for global maps of emergence risk. For example, previous studies that have used host richness as proxy have predicted a high diversity of unsampled bat viruses,<sup>14</sup> bat coronaviruses,<sup>2</sup> and even specifically betacoronaviruses<sup>15</sup> in both the Amazon and southeast Asia. While we find that both regions are characterized by highly divergent host and viral communities, our framework identifies key differences between the regions. We find that Latin America is a hotspot of

197 both host and viral distinctiveness, suggesting that this branch of the bat-betacoronavirus complex may be  
198 undergoing independent evolutionary dynamics from the rest of the global pool, but with limited potential  
199 for viral diversification—a finding that is supported by previous work indicating a higher rate of  
200 codivergence in Latin America.<sup>2</sup> In contrast, in southeast Asia, host richness and viral distinctiveness are  
201 high but sharing is low; this suggests a different type of evolutionary dynamics that could generate high  
202 local diversity of viruses through host switching and viral recombination (see e.g.,<sup>13</sup> as well as the  
203 discovery of recombinant viruses that share genetic material from both the SARS-CoV and SARS-CoV-2  
204 branches of the Sarbecovirus lineage).<sup>36</sup> Both of these regions are priority areas for sampling, especially  
205 given predictions that they contain many bat hosts of undiscovered betacoronaviruses.<sup>15,17</sup> However, both  
206 the evolutionary and ecological aspects of emergence risk are likely higher in southeast Asia—a fact that  
207 will only become more relevant, as bats track shifting climates and exchange viruses with other species,  
208 creating a hotspot of cross-species transmission unique to the region.<sup>37</sup>

209 The diversity and diversification potential of bats responds to anthropogenic factors others than shifting  
210 climates.<sup>38</sup> Land use changes could significantly decrease bat suitability, notably through effects on diet  
211 and availability of habitats.<sup>39</sup> As our results establish that the diversification of bats betacoronaviruses  
212 happens on top of processes affecting hosts, biogeographic variation in human population density and  
213 anthropogenic disturbances may feed into co-evolutionary dynamics. Increase in humans-hosts contacts  
214 also increase the risk of emergence of novel diseases,<sup>40</sup> so does the changes in landscape connectivity at  
215 local/regional scales.<sup>41</sup> This represents a challenge for both conservation strategies and disease ecology:  
216 some areas can a high emergence risk and more potential for the acquisition of zoonotic viruses through  
217 bat-human encounters.<sup>42</sup> In particular, the challenge ahead lies in the need to quantify actual exposure  
218 (and risk) accounting for several transmission scenarios, including both direct and indirect bat - human  
219 interactions, and feeding back into the provision of ecosystem services by bats.

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

227 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation. DJB was  
228 supported by the National Institute of General Medical Sciences of the National Institutes of Health  
229 (P20GM134973).

230 **Methods**

231 **Known *Betacoronavirus* hosts**

232 We downloaded the data on bats hosts of *Betacoronavirus* from  
233 <https://www.viralemergence.org/betacov> on Apr. 2022,<sup>15</sup> and filtered it to “known” hosts (established  
234 before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling and competence  
235 assays since the initial data collection). The original database was assembled by a combination of data  
236 mining and literature surveys, including automated alerts on the “bats” and “coronavirus” keywords to  
237 identify novel empirical evidence of bats-betacoronaviruses associations; this yielded a total of 126 known  
238 hosts, 47 of which were novel hosts.

239 **Bat occurrences**

240 We downloaded the rangemap of every current bat species that was classified as an empirically  
241 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data.<sup>43</sup> The range  
242 maps were subsequently rasterized using the rasterize function from GDAL<sup>44</sup> at a resolution of  
243 approximately 100kmx100km. For every pixel in the resulting raster where at least one bat host of  
244 *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was used to  
245 calculate the following risk assessment components: bat phylogenetic diversity, bat compositional  
246 uniqueness, and predicted viral sharing risk.

247 **Bat phylogenetic diversity**

248 For every pixel, we measured Faith’s Phylogenetic Diversity<sup>45</sup> based on a recent synthetic tree with robust  
249 time calibration, covering about 6000 mammalian species.<sup>46</sup> Faith’s PD measures the sum of unique  
250 branches from an arbitrary root to a set of tips, and comparatively larger values indicate a more  
251 phylogenetic diverse species pool. We measured phylogenetic diversity starting from the root of the entire  
252 tree (and not from Chiroptera); this bears no consequences on the resulting values, since all branches  
253 leading up to Chiroptera are only counted one per species pool, and (as we explain when describing the  
254 assembly of the composite risk map), all individual risk components are ranged in [0,1]. This measure  
255 incorporates a richness component, which we chose not to correct for; the interpretation of the

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

## 258 **Bat compositional uniqueness**

259 For every species pool, we measured its Local Contribution to Beta-Diversity;<sup>47</sup> LCBD works from a  
260 species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites are columns, and a value of  
261 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel represents a unique location,  
262 and following best practices<sup>48</sup> transformed it using Hellinger's distance to account for unequal bat  
263 richness at different pixels. The correction of raw community data is particularly important for two  
264 reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect  
265 of overall species richness, which is already incorporated in the phylogenetic diversity component. High  
266 values of LCBD indicate that the pixel has a community that is on average more dissimilar in species  
267 composition than what is expected knowing the entire matrix, i.e. a more unique community. Recent  
268 results by<sup>49</sup> shows that LCBD measures are robust with regards to spatial scale, and are therefore  
269 applicable at the global scale.

## 270 **Viral sharing between hosts**

271 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a  
272 previously published generalized additive mixed model of virus sharing by a tensor function of  
273 phylogenetic distance and geographic range overlap across mammals.<sup>50</sup> This network stores pairwise  
274 values of viral community similarity. To project viral sharing values into a single value for every pixel, we  
275 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant  
276 bat assemblage is likely to be proficient at exchanging viruses.

## 277 **Composite risk map**

278 To visualize the aggregated risk at the global scale, we combine the three individual risk components  
279 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model.<sup>51</sup> In  
280 this approach, every risk component gets assigned a component in the RGB color model (phylogenetic  
281 diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid

282 RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no  
283 phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for  
284 each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the  
285 risk as colors diverge towards combinations of values for three risk components. Out of the possible  
286 combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic  
287 diversity and low viral sharing,<sup>52</sup> in that this allows multiple independent host-virus coevolutionary  
288 dynamics to take place in the same location. In the colorimetric space, this corresponds to yellow – because  
289 the HSV space is more amenable to calculations for feature extraction,<sup>53</sup> we measured the risk level by  
290 calculating the angular distance of the hue of each pixel to a reference value of 60 (yellow), and weighted  
291 this risk level by the value component. Specifically, given a pixel with colorimetric coordinates  $(h, s, v)$ , its  
292 ranged weighted risk value is

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

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

## 294 Viral phyogeography and evolutionary diversification

295 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed  
296 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data  
297 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR  
298 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR  
299 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated  
300 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained  
301 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and  
302 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East  
303 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus  
304 RdRp sequences were then aligned using MAFFT<sup>54</sup> v1.4.0 (Algorithm FFT-NS-2, Scoring matrix 200PAM /  
305 k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree reconstructed in  
306 IQ-TREE<sup>55</sup> v1.6.12 with ModelFinder<sup>56</sup> ultrafast bootstrap approximation<sup>57</sup> with a general time reversible  
307 model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of nucleotide

308 substitution (GTR+F+R5).

309 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat  
310 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the  
311 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary  
312 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary  
313 distinctiveness<sup>58</sup> for each of the viruses in the tree, then averaged these at the bat species level, projected  
314 these values onto their geographic distributions, and averaged across every bat found in a given pixel. As  
315 such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral  
316 community believed to be associated with a particular subset of bats present.

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

318 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
319 biogeography of their hosts. To test this idea, we loosely adapted a method from,<sup>59,60</sup> who proposed a  
320 phylogenetic method for the delineation of animal biogeographic regions. In their original method, a  
321 distance matrix - where each row or column represents a geographic raster’s grid cell, and the dissimilarity  
322 values are the “beta diversity similarity” of their community assemble - undergoes non-metric  
323 multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a  
324 four-color bivariate map. Here, we build on this idea with an entirely novel methodology. First, we  
325 measure the phylogenetic distance between the different viruses in the betacoronaviruses tree by using the  
326 cophenetic function in ape;<sup>61</sup> subsequently, we take a principal components analysis of that distance  
327 matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional  
328 space. We then take the first two principal components and, as with the evolutionary distinctiveness  
329 analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

330 **References**

- 331 1.
- 332 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 333 2.
- 334 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 335 3.
- 336 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 337 4.
- 338 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).
- 339 5.
- 340 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 341 6.
- 342 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).
- 343 7.
- 344 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 345 8.
- 346 Thompson, J. N. *The Coevolutionary Process*. (University Of Chicago Press, 1994).
- 347 9.
- 348 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 349 10.
- 350 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).  
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 351 11.
- 352 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).

- 353 12.
- 354 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).
- 355 13.
- 356 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).
- 357 14.
- 358 Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
- 359 15.
- 360 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).
- 361 16.
- 362 Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications* **8**,  
(2017).
- 363 17.
- 364 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).
- 365 18.
- 366 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).
- 367 19.
- 368 Kettenburg, G. *et al.* Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a Unique  
Evolutionary History for This Coronavirus Clade. *Frontiers in Public Health* **10**, (2022).
- 369 20.
- 370 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).
- 371 21.
- 372 Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10**, 1030–1037  
(2004).
- 373 22.
- 374 Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.  
*Virologica Sinica* **33**, 104–107 (2018).

- 375 23.
- 376 Ramshaw, R. E. *et al.* A database of geopositioned Middle East Respiratory Syndrome Coronavirus occurrences. *Scientific Data* **6**, 318 (2019).
- 377 24.
- 378 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).
- 379 25.
- 380 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).
- 381 26.
- 382 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).
- 383 27.
- 384 Van Brussel, K. & Holmes, E. C. Zoonotic disease and virome diversity in bats. *Current Opinion in Virology* **52**, 192–202 (2022).
- 385 28.
- 386 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).
- 387 29.
- 388 Simmons, N. B. & Cirranello, A. L. Bat Species of the World: A taxonomic and geographic database. <https://batnames.org/> (2020).
- 389 30.
- 390 Kasso, M. & Balakrishnan, M. Ecological and Economic Importance of Bats (Order Chiroptera). *ISRN Biodiversity* **2013**, e187415 (2013).
- 391 31.
- 392 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).
- 393 32.
- 394 *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).
- 395 33.
- 396 Williams-Guillén, K., Perfecto, I. & Vandermeer, J. Bats Limit Insects in a Neotropical Agroforestry System. *Science* **320**, 70–70 (2008).

- 397 34.
- 398 Gonsalves, L., Bicknell, B., Law, B., Webb, C. & Monamy, V. Mosquito Consumption by Insectivorous Bats:  
Does Size Matter? *PLOS ONE* **8**, e77183 (2013).
- 399 35.
- 400 Gonsalves, L., Lamb, S., Webb, C., Law, B. & Monamy, V. Do mosquitoes influence bat activity in coastal  
habitats? *Wildlife Research* **40**, 10–24 (2013).
- 401 36.
- 402 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).
- 403 37.
- 404 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).
- 405 38.
- 406 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).
- 407 39.
- 408 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).
- 409 40.
- 410 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).
- 411 41.
- 412 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).
- 413 42.
- 414 Amman, B. R. *et al.* Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and  
public health interest. (FAO, 2011).
- 415 43.
- 416 IUCN. The IUCN Red List of Threatened Species. <https://www.iucnredlist.org> (2021).
- 417 44.
- 418 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).

- 419 45.
- 420 Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10 (1992).
- 421 46.
- 422 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).
- 423 47.
- 424 Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters* **16**, 951–963 (2013).
- 425 48.
- 426 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).
- 427 49.
- 428 Dansereau, G., Legendre, P. & Poisot, T. Evaluating ecological uniqueness over broad spatial extents using species distribution modelling. *Oikos* **n/a**, e09063 (2022).
- 429 50.
- 430 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).
- 431 51.
- 432 Seekell, D. A., Lapierre, J.-F. & Cheruvellil, K. S. A geography of lake carbon cycling. *Limnology and Oceanography Letters* **3**, 49–56 (2018).
- 433 52.
- 434 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).
- 435 53.
- 436 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).
- 437 54.
- 438 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).
- 439 55.
- 440 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).

- 441 56.
- 442 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).
- 443 57.
- 444 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).
- 445 58.
- 446 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).
- 447 59.
- 448 Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National  
Academy of Sciences* **104**, 5925–5930 (2007).
- 449 60.
- 450 Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions.  
*Journal of Biogeography* **37**, 2029–2053 (2010).
- 451 61.
- 452 Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in  
R. *Bioinformatics* **35**, 526–528 (2019).

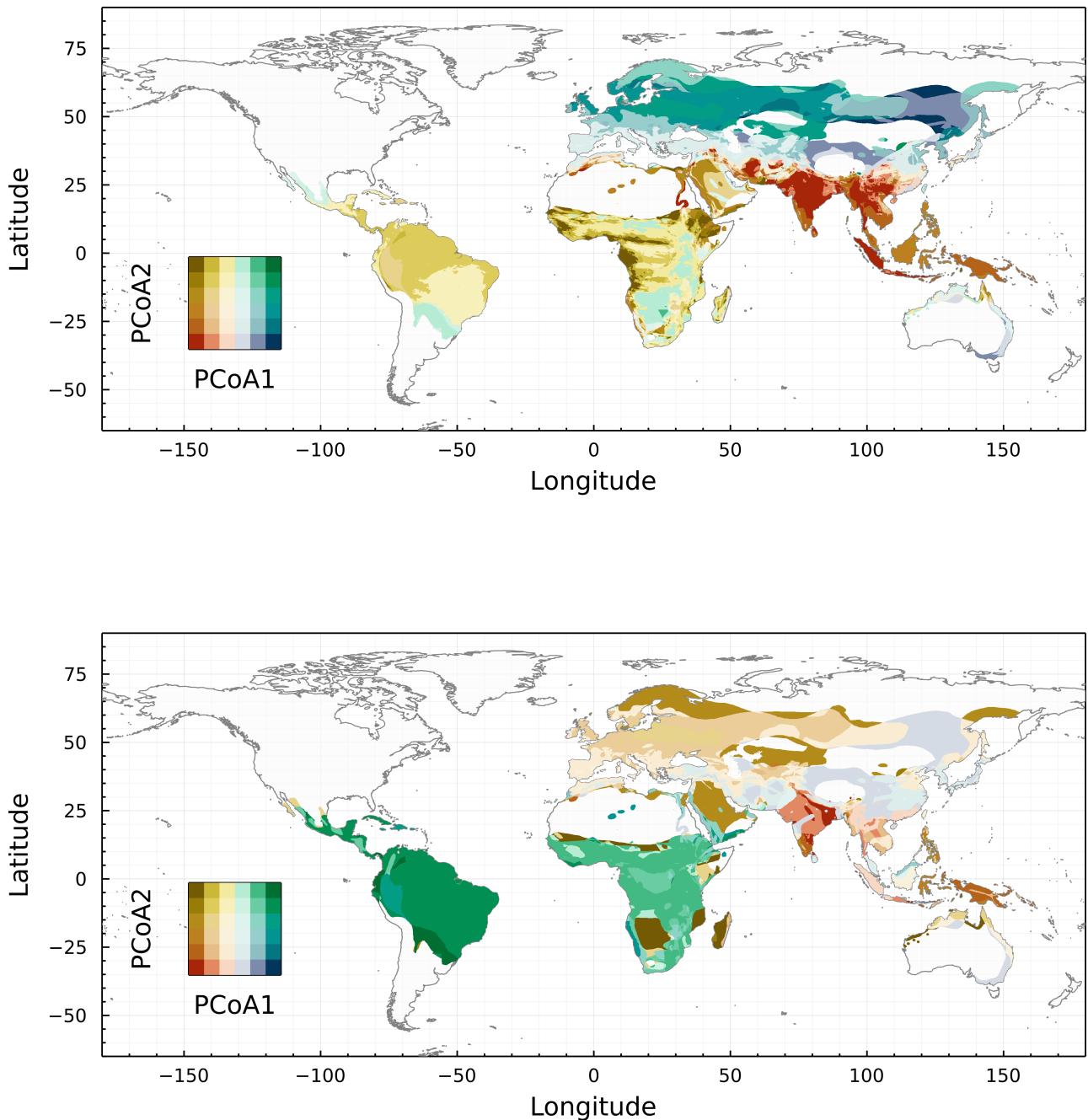


Figure 1: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent, and so cannot be compared directly – that being said, the regions can be compared across maps.

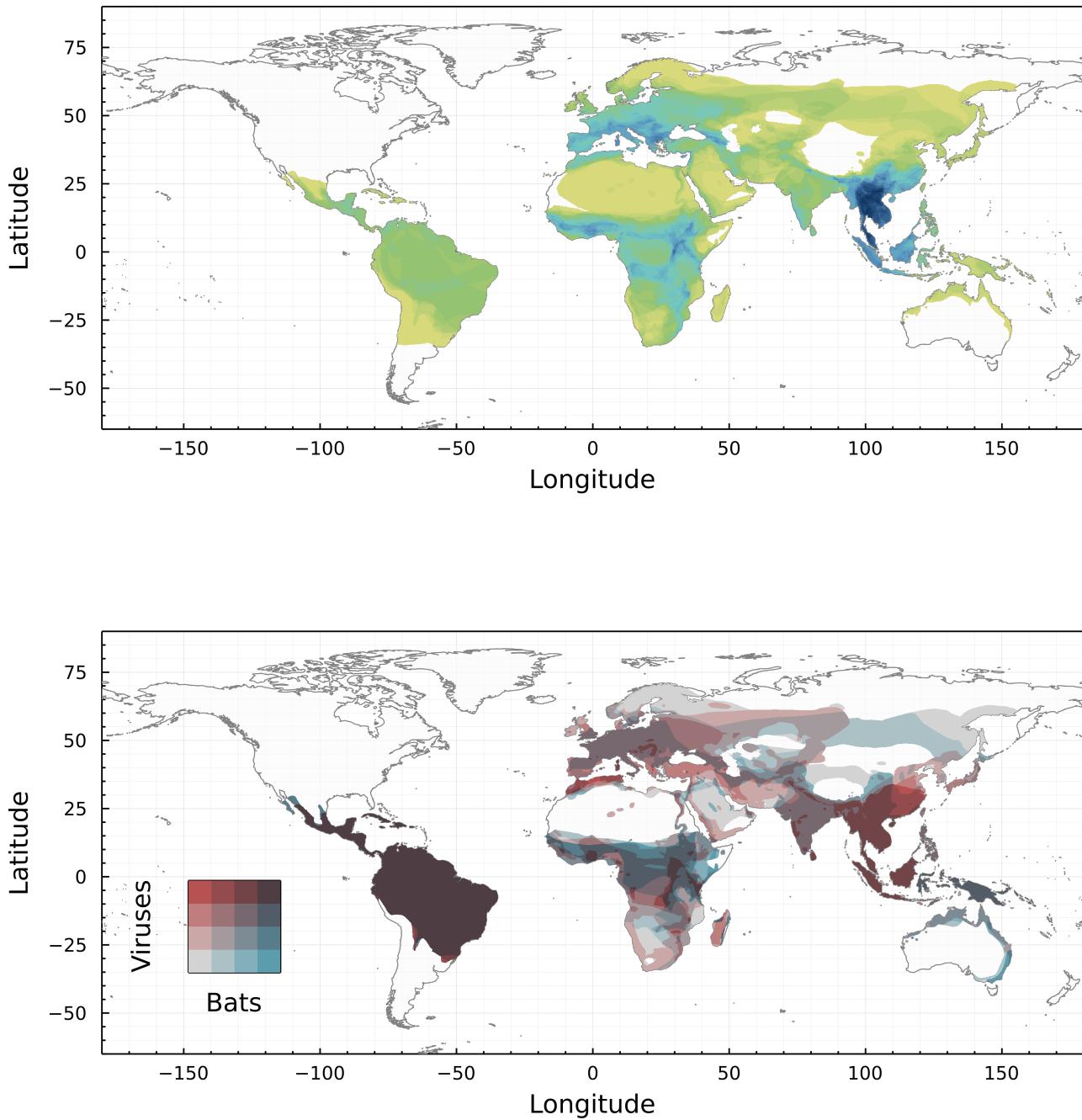


Figure 2: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).

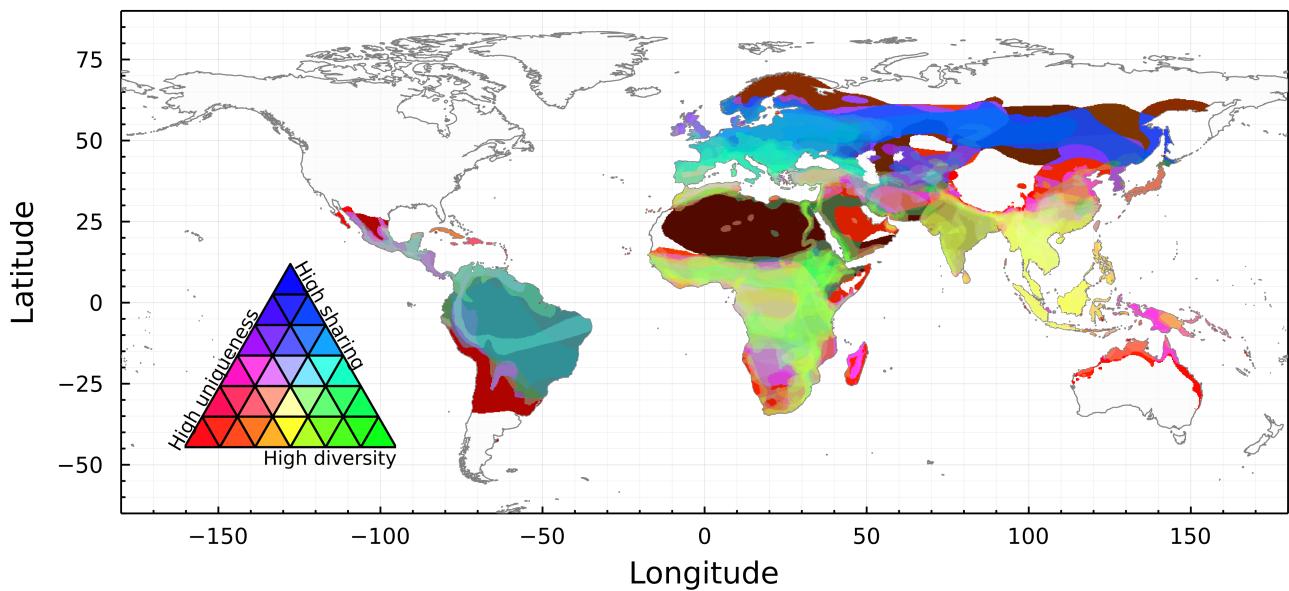


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible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

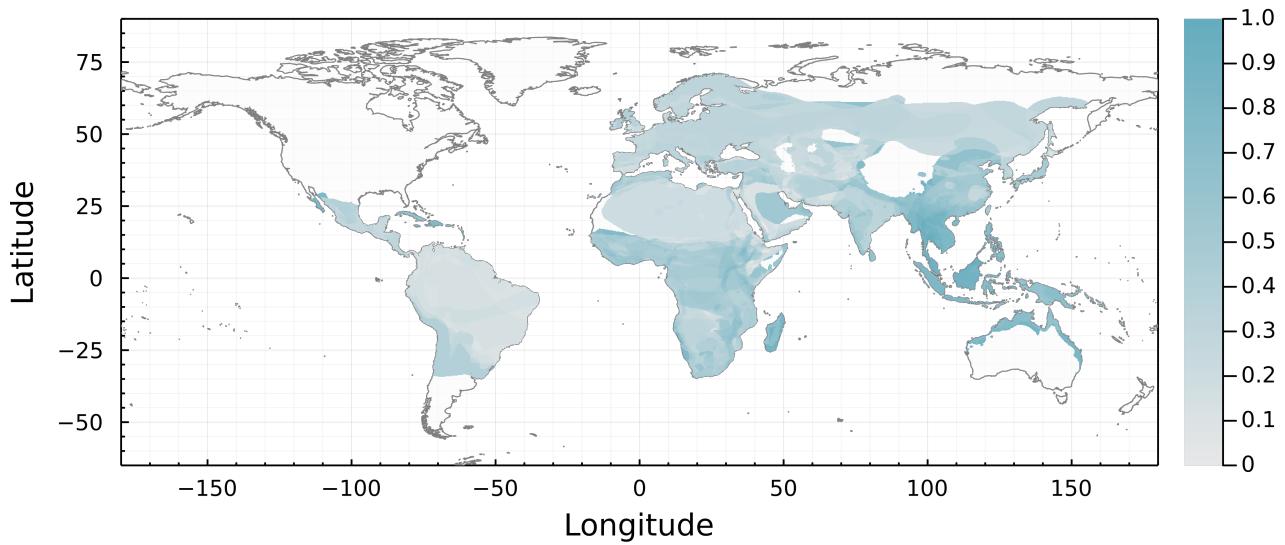


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

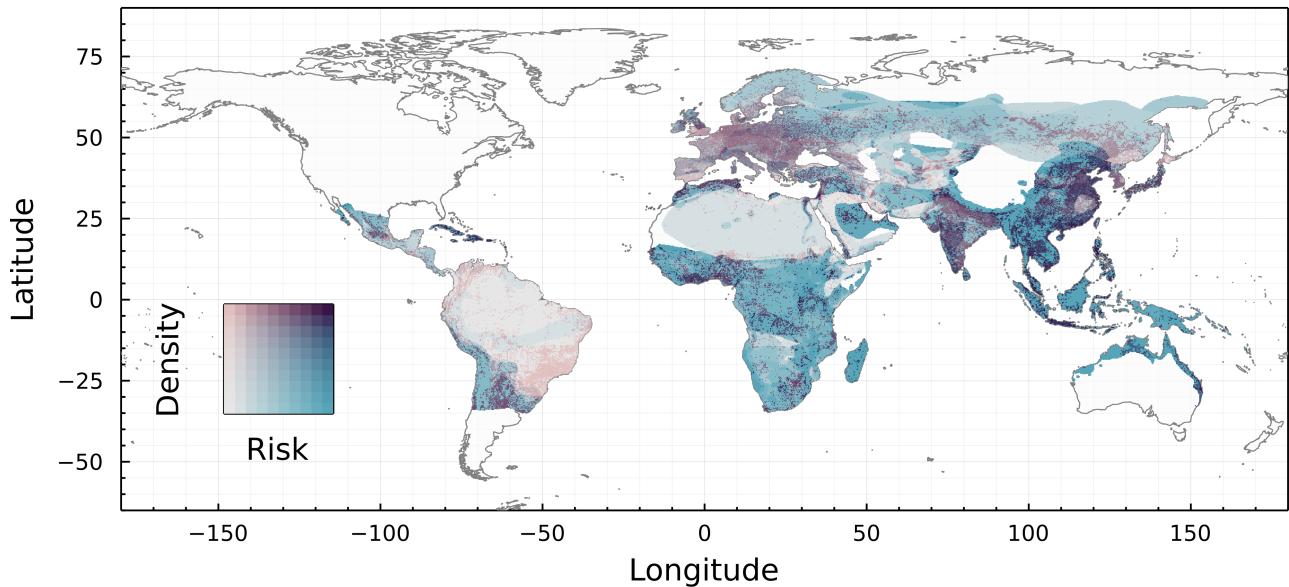


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk,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