

# 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

Pathogen evolution is one of the least predictable components of disease emergence, particularly in zoonotic reservoirs. Here, building on principles established by the geographic mosaic theory of coevolution, we develop a quantitative, spatially-explicit framework for mapping the evolutionary risk of viral emergence. Driven by interest in diseases like SARS, MERS, and COVID-19, we examine the global biogeography of bat-origin betacoronaviruses, and apply our coevolutionary framework to develop a map of emergence risk. We find that our framework helps explain patterns like a unique pool of merbecoviruses in the Neotropics, a recently-discovered lineage of divergent nobecoviruses in Madagascar, and—most importantly—hotspots of diversification in southeast Asia, sub-Saharan Africa, and the Middle East that correspond to the site of previous zoonotic emergence events. Our framework may help identify hotspots of future risk that have also been previously overlooked, like west Africa and the Indian subcontinent, and may more broadly help researchers understand how reservoir ecology shapes the evolution of pandemic threats.

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 mechanisms, giving rise to fragmented systems with  
14 different ecologies over large spatial extents.<sup>10</sup> The GMTC predicts a spatial fragmentation of  
15 coevolutionary dynamics under the joint action of three processes:<sup>11</sup> coevolutionary hot- and coldspots,  
16 which appear when the intensity of *interaction* (in terms of reciprocal fitness consequences) varies  
17 spatially; selection mosaics, wherein the intensity of *selection* varies across space, driven by both the biotic  
18 complexity of the community (locally diverse hosts and viruses are more biotically complex) and the local  
19 favorability of the environment;<sup>12</sup> and trait remixing, which occurs when coevolutionary dynamics change  
20 when community-level *functional traits* change through meta-community dynamics.

21 Here, we apply the GMTC to explore and explain the global biogeography of betacoronaviruses, the group  
22 that includes SARS-CoV, MERS-CoV, and SARS-CoV-2. In their bat reservoirs, coronaviruses evolve  
23 through a mix of host jumps, recombination among disparate lineages, and, to a lesser degree,  
24 co-divergence with their hosts—<sup>2</sup>a mix of mechanisms that creates a complex and nonlinear relationship  
25 between host diversity and viral emergence. Working from a recently published database of bat hosts of  
26 betacoronaviruses, we test whether spatial structure in bat-betacoronavirus coevolution is identifiable at a  
27 global scale. Aiming to explain these patterns, we develop a generalized framework for applying the  
28 GMTC to host-virus interactions, with a specific emphasis on the potential to create independent  
29 coevolutionary dynamics (and therefore spatial fragmentation in risk) through heterogeneity. We develop  
30 a trivariate risk assessment system that connects each GMTC mechanism to a quantifiable aspect of

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

## 39 Results and Discussion

### 40 Bat and betacoronavirus biogeography are broadly consistent

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

57 substantial broad divergence in their biogeography. However, host jumps at the family level or higher are  
58 relatively rare and significant events in coronavirus evolutionary history;<sup>2,13</sup> as a result, the mosaic of  
59 betacoronavirus phylogeography is assembled from a set of overlapping smaller coevolutionary systems,  
60 superimposed in space and filtered by the importance of different subgroups in local host communities.  
61 For example, the most speciose and cosmopolitan family of bats, the vesper bats (Vespertilionidae), are  
62 considered the primary hosts of the subgenus *Merbecovirus* (MERS-like viruses);<sup>3,13</sup> but in the Americas,  
63 where merbecoviruses are the only lineage present, they have only been found in other bat taxa (e.g.,  
64 Molossidae, Phyllostomidae).<sup>14–17</sup> At the coarsest scale, these heterogeneities are lost, and betacoronavirus  
65 biogeography tracks the deep rifts in bat evolutionary history—but within broad regions, the component  
66 coevolutionary systems may have very different dynamics.

## 67 **Hotspots of bat and betacoronavirus biodiversity are distinct**

68 Bats, the second most diverse groups of mammals, are found worldwide; gradients in their species  
69 richness generally track broader patterns of mammal diversity, with a striking Neotropical hotspot  
70 (especially in the Amazon basin) and a secondary hotspot centered in the southeast Asian peninsula.  
71 These hotspots of bat diversity are generally presumed to be hotspots of viral adaptive radiation, and  
72 therefore areas of concern for human health.<sup>2,18</sup> However, the hotspots of known bat betacoronavirus  
73 hosts show a distinct pattern, with primary hotspots (both in terms of size and higher values) of host  
74 richness situated in southeast Asia, parts of southern Europe, and to a lesser extent parts of Africa in the  
75 -25-0 range of latitudes (fig. 2; top). Although hundreds of species likely host undiscovered  
76 betacoronaviruses, machine learning predictions have suggested that these undiscovered reservoirs should  
77 follow the same diversity gradient.<sup>19</sup> In principle, these hotspots of locally-diverse, virus-rich bat  
78 communities should drive more adaptive diversification in their viruses.

79 [Figure 2 about here.]

80 However, we find that the global pattern of betacoronavirus phylogenetic distinctiveness is quite distinct  
81 from both bat host richness and phylogenetic distinctiveness (fig. 2; bottom). In contrast to the sparsity of  
82 Neotropical betacoronavirus hosts, South and Central America have the most evolutionary distinct hosts  
83 and viruses, followed by secondary hotspots in southeast Asia and the Rift Valley region have mostly  
84 distinct viruses. Some degree of sampling bias may contribute to these patterns: for example, the

85 Neotropics are one of the places where the fewest bat betacoronavirus sequences have been generated  
86 (cite2), resulting in a sparser phylogenetic tree, and artificially inflating distinctiveness; conversely,  
87 disproportionate research effort in eastern China<sup>20</sup> may have led to a more complete inventory of the local  
88 diversity of coronaviruses, again inflating these metrics relative to underlying patterns. Even accounting  
89 for these potential biases, though, there is obvious heterogeneity in betacoronavirus evolutionary  
90 distinctiveness that is distinct from overall bat diversity.

91 Overall, these patterns recapitulate the evolutionary history of both the order Chiroptera and the genus  
92 *Betacoronavirus*. Horseshoe bats (Rhinolophidae) include the reservoirs of the SARS-like viruses  
93 (subgenus *Sarbecovirus*), the group of pandemic threats that have been of the greatest interest to  
94 researchers<sup>13</sup> (and so have been sampled most intensively).<sup>20</sup> The hotspots of host richness and viral  
95 diversity in southeast Asia—both of which are disproportionately high, considering the global landscape  
96 of bat species richness—are almost entirely driven by viral adaptive radiation through host switching  
97 within this clade<sup>3,19</sup>. In contrast, the Neotropical hotspot of viral distinctiveness is driven by isolation by  
98 host vicariance. Out of the four main groups of betacoronaviruses, only merbecoviruses have been found  
99 in animals in the Americas—an introduction that is generally presumed to be ancient.<sup>3,21</sup> While  
100 comparatively understudied, New World merbecoviruses have been found in the ghost-faced bats  
101 (Mormoopidae), Neotropical leaf-nosed bats (Phyllostomidae), and free-tailed bats (Molossidae).<sup>14–17</sup> The  
102 former two groups and a clade of the latter are endemic to the Neotropics, while the explosive adaptive  
103 radiations of the phyllostomids are responsible for the hotspot of bat diversity in the Amazon.<sup>22</sup> Together,  
104 these clades of New World bats play host to a distinct regime of betacoronavirus coevolution.

## 105 Coevolutionary regimes structure evolutionary potential for zoonotic emergence

106 The existence of well-defined cophylogenetic regions suggests that the bat-betacoronavirus system is  
107 spatially fragmented enough to create divergent coevolutionary trajectories; in turn, this coevolutionary  
108 mosaic may alter the risk of zoonotic emergence. These ideas are, respectively, supported by the existence  
109 of hotspots of viral uniqueness and the diverse origins of human betacoronaviruses. Together, this  
110 framework points to a predictable relationship between host community structure and coevolutionary  
111 pressure: phylogeographic structure in bat hosts (and their diverse immune strategies)<sup>23</sup> creates a  
112 landscape of selective pressure; the trajectory of viruses' coevolutionary response is, in turn, constrained  
113 by their opportunities for either specialization or diversification through host jumps and recombination.

114 Based on the geographic mosaic theory of coevolution, we developed a trivariate map of three facets of  
115 coevolutionary pressure (fig. 3): (1) *host phylogenetic diversity*: a high diversity of evolutionary histories  
116 should expose viruses to more variation in host immune traits; (2) *host community uniqueness*: exposure to  
117 greater host trait heterogeneity can drive viral diversification, and coevolving with more unique host  
118 communities should create more unique branches of viral evolution; and (3) propensity for *viral sharing*:  
119 frequent cross-species transmission may act as a buffer on selective pressure, while lower rates of  
120 exchange may enable more simultaneous trajectories of viral specialization to coexist within a given  
121 community. We combine global maps of all three to generate a map of coevolutionary regimes, where  
122 close colors represent similar risks, and paler pixels represent overall higher risk (see Methods). We find  
123 that these regions do not neatly overlap with those defined in fig. 1 or fig. 2, reinforcing the notion that  
124 local-scale coevolutionary mosaics can form within cophylogenetic regions.

125 [Figure 3 about here.]

126 The greatest evolutionary potential for zoonotic emergence exists where pathogen pools have a high  
127 genetic diversity and high propensity for cross-species transmission. In our framework, emergence risk is  
128 therefore maximized under higher phylogenetic diversity (viruses are exposed to different host clades),  
129 higher host uniqueness (viruses are experiencing novel, heterogeneous host traits combinations), and low  
130 to medium viral sharing (host-virus pairs can coevolve independently, but divergent viruses may still have  
131 opportunities for recombination). In fig. 3, this corresponds to yellow areas (dynamics dominated by low  
132 viral sharing, with equal contributions of selection mosaics and trait remixing; southeast Asia, and the  
133 Indian sub-continent), green-yellow areas (dynamics with low viral sharing but dominated by the  
134 selection mosaic effect of host diversity; sub-Saharan Africa), and red-yellow areas (dynamics with low  
135 viral sharing but dominated by trait remixing in host communities; the Middle East). Translating this axis  
136 of variation back into a univariate risk map (fig. 4) highlights that this evolutionary landscape has a  
137 striking correspondence to regions where zoonotic betacoronaviruses have previously emerged.

138 Compared to approaches that map emergence risk based only on the number of known bat hosts of  
139 betacoronaviruses, our framework suggests regions where high viral sharing dominates coevolutionary  
140 dynamics—such as Latin America, or Eurasia above a northing of 30—would pose less of a relative risk of  
141 zoonotic emergence. Nevertheless, areas of high host uniqueness coupled with high viral sharing  
142 (red-to-pink in fig. 3) could create hotspots facilitated by viral codivergence. Our framework identifies

143 Madagascar, where most bat species are endemic following evolutionary divergence from sister species in  
144 both African and Asian continents,<sup>24</sup> as one such hotspot; interestingly, a recent study<sup>25</sup> reported a novel  
145 and highly divergent lineage of nobecoviruses from Madagascar-endemic pteropid bat species (*Pteropus*  
146 *rufus* and *Rousettus madagascariensis*), again supporting the predictive power of the coevolutionary  
147 framework.

148 [Figure 4 about here.]

#### 149 **Human landscapes filter the geography of emergence risk**

150 The relationship between the underlying pathogen pool and emergence risk is mediated by both  
151 human-wildlife interfaces (the probability of spillover) and opportunities for onward transmission (the  
152 probability that spillovers become epidemics)<sup>1</sup>. As a proxy for both, we finally overlaid the risk component  
153 from the composite map (see above) with the proportion of built land, as a proxy for a mix of habitat  
154 disturbance, potential for bat synanthropy or contact with bridge hosts like livestock,<sup>26,27</sup> and human  
155 population density and connectivity<sup>1,28,29</sup> (fig. 5). Accounting for these factors, most of South America and  
156 Europe are at comparatively lower risk, as—although densely populated—settlements tend to be in areas  
157 with lower potential risk. Conversely, regions like Malaysia and the northern coast of Australia have a  
158 high evolutionary risk component, but should represent a relatively lower effective risk due to low human  
159 density. However, southeast Asia, the Indian subcontinent, and scattered hotspots in sub-Saharan Africa  
160 are at high risk due to the overlap between human populations and natural opportunities for cross-species  
161 transmission of betacoronaviruses.

162 [Figure 5 about here.]

163 Reassuringly, these predictions correspond to the geographic origins of the three bat-origin coronaviruses  
164 that have recently emerged in human populations. While available information puts the spillover of  
165 SARS-CoV-2 in a live animal market in Wuhan, China, the ultimate origin of the virus is almost certainly  
166 in a divergent lineage of sarbecoviruses from the Indochinese peninsula that was poorly characterized  
167 prior to the pandemic.<sup>30–32</sup> Similarly, the SARS-CoV outbreak began in Guangdong province in 2002,  
168 reaching humans through small carnivore bridge hosts, but was eventually traced back to a set of likely  
169 progenitor viruses found in cave-dwelling horseshoe bats in Yunnan province;<sup>33</sup> nearby, antibody

170 evidence has indicated human exposure to SARS-like viruses.<sup>34</sup> MERS-CoV was originally detected in  
171 Saudi Arabia, accompanied by a nearly identical virus sequenced from an Egyptian tomb bat (*Taphozous*  
172 *perforatus*),<sup>35</sup> but is widespread in camels in East Africa and the Middle East, and may have reached its  
173 bridge host decades earlier than originally supposed;<sup>36</sup> as a result, the geography of the original  
174 bat-to-camel transmission is still widely regarded as uncertain. All of these are broadly consistent with the  
175 risk factors we identify. Notably, India and west Africa are additional hotspots that have yet to experience  
176 the emergence of a bat coronavirus into human populations, but may still be at risk—particularly given  
177 known gaps in bat surveillance,<sup>20</sup> and a dense population in both regions with global connectivity. In any  
178 of these regions, surveillance on viral reservoirs can be paired with targeted monitoring of high-risk  
179 human populations (i.e., those with regular wildlife contact)<sup>37</sup> for maximum impact.

## 180 Conclusion

181 Bats emerged around 64 million years ago, and are one of the most diverse mammalian orders, with more  
182 than 1,400 estimated species.<sup>38,39</sup> They exhibit a broad variety of habitat use, behaviour, and feeding  
183 strategies, putting them at key positions in the delivery and provisioning of several ecosystem services, tied  
184 to important ecosystem-derived benefits to humans.<sup>40</sup> Over two-thirds of bats are known to be either  
185 obligate or facultative insectivores, therefore actively contributing for agricultural pest control,<sup>41,42</sup> and  
186 vectors of pathogens that put a risk on human health;<sup>43,44</sup> some other species are essential links in many  
187 seed-dispersal networks.<sup>45</sup> However, many of these species face a high risk of extinction, particularly given  
188 persecution and killings that sometimes follows from messaging about their role in disease emergence.

189 Areas where bats, viruses, and humans co-occur are not always hotspots of risk for human health; as such,  
190 developing more precise ways to map zoonotic hazards can help bats and humans coexist safely, and  
191 support the conservation of these important and unique animals.

192 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries  
193 like highly divergent nobecoviruses in Madagascar and the once-neglected adaptive radiation of  
194 sarbecoviruses in the Indochinese peninsula. In doing so, it advances ecological theory beyond the current  
195 state of the art for global maps of emergence risk. For example, previous studies that have used host  
196 richness as a proxy have predicted a high diversity of unsampled bat viruses,<sup>18</sup> bat coronaviruses,<sup>2</sup> and  
197 even specifically betacoronaviruses<sup>19</sup> in both the Amazon and southeast Asia. While we find that both

198 regions are characterized by unique and diverse communities of both hosts and viruses, our framework is  
199 able to identify key differences between the two systems. We find that the merbecovirus complex in Latin  
200 America has been a unique branch of evolution separate from the rest of the global pool, but with limited  
201 potential for viral diversification—a finding that is supported by previous work indicating a higher rate of  
202 codivergence in Latin America.<sup>2</sup> In contrast, in southeast Asia, host richness and viral distinctiveness are  
203 high but sharing is low; this suggests a different type of evolutionary dynamics that could generate high  
204 local diversity of viruses through host switching and viral recombination (see e.g.,<sup>13</sup> as well as the  
205 discovery of recombinant viruses with genetic material from both the SARS-CoV and SARS-CoV-2  
206 branches of the Sarbecovirus lineage).<sup>46</sup>

207 Both of these regions are priority areas for sampling, especially given predictions that they contain many  
208 bat hosts of undiscovered betacoronaviruses.<sup>19,20</sup> However, both the evolutionary and ecological aspects of  
209 emergence risk are higher in southeast Asia—a fact that will only become more relevant, as bats track  
210 shifting climates and exchange viruses with other species, creating a hotspot of elevated cross-species  
211 transmission unique to the region.<sup>28,47</sup> Bats—and the spillover of their viruses—are also sensitive to  
212 anthropogenic factors others than climate change, including deforestation and other kinds of habitat loss,  
213 increased stress, and greater contact with potential bridge hosts like domesticated species<sup>26,48,49</sup> (add cite  
214 for 10.1038/nature13139). This represents a challenge for both conservation strategies and pandemic  
215 prevention,<sup>50</sup> but identifying areas at risk, and protecting the health of bats and ecosystems within those  
216 zones, can be a win-win intervention for both (add cites for DOIs: 10.1038/s41893-020-00640-z;  
217 10.1016/S2542-5196(21)00031-0).

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

228 **Methods**

229 **Known *Betacoronavirus* hosts**

230 We downloaded the data on bats hosts of *Betacoronavirus* from  
231 <https://www.viralemergence.org/betacov> on Apr. 2022,<sup>19</sup> and filtered it to “known” hosts (established  
232 before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling and competence  
233 assays since the initial data collection). The original database was assembled by a combination of data  
234 mining and literature surveys, including automated alerts on the “bats” and “coronavirus” keywords to  
235 identify novel empirical evidence of bats-betacoronaviruses associations; this yielded a total of 126 known  
236 hosts, 47 of which were novel hosts.

237 **Bat occurrences**

238 We downloaded the rangemap of every current bat species that was classified as an empirically  
239 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data.<sup>51</sup> The range  
240 maps were subsequently rasterized using the rasterize function from GDAL<sup>52</sup> at a resolution of  
241 approximately 100kmx100km. For every pixel in the resulting raster where at least one bat host of  
242 *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was used to  
243 calculate the following risk assessment components: bat phylogenetic diversity, bat compositional  
244 uniqueness, and predicted viral sharing risk.

245 **Bat phylogenetic diversity**

246 For every pixel, we measured Faith’s Phylogenetic Diversity<sup>53</sup> based on a recent synthetic tree with robust  
247 time calibration, covering about 6000 mammalian species.<sup>54</sup> Faith’s PD measures the sum of unique  
248 branches from an arbitrary root to a set of tips, and comparatively larger values indicate a more  
249 phylogenetic diverse species pool. We measured phylogenetic diversity starting from the root of the entire  
250 tree (and not from Chiroptera); this bears no consequences on the resulting values, since all branches  
251 leading up to Chiroptera are only counted one per species pool, and (as we explain when describing the  
252 assembly of the composite risk map), all individual risk components are ranged in [0,1]. This measure  
253 incorporates a richness component, which we chose not to correct for; the interpretation of the

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

## 256 **Bat compositional uniqueness**

257 For every species pool, we measured its Local Contribution to Beta-Diversity;<sup>55</sup> LCBD works from a  
258 species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites are columns, and a value of  
259 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel represents a unique location,  
260 and following best practices<sup>56</sup> transformed it using Hellinger's distance to account for unequal bat  
261 richness at different pixels. The correction of raw community data is particularly important for two  
262 reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect  
263 of overall species richness, which is already incorporated in the phylogenetic diversity component. High  
264 values of LCBD indicate that the pixel has a community that is on average more dissimilar in species  
265 composition than what is expected knowing the entire matrix, i.e. a more unique community. Recent  
266 results by<sup>57</sup> shows that LCBD measures are robust with regards to spatial scale, and are therefore  
267 applicable at the global scale.

## 268 **Viral sharing between hosts**

269 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a  
270 previously published generalized additive mixed model of virus sharing by a tensor function of  
271 phylogenetic distance and geographic range overlap across mammals.<sup>58</sup> This network stores pairwise  
272 values of viral community similarity. To project viral sharing values into a single value for every pixel, we  
273 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant  
274 bat assemblage is likely to be proficient at exchanging viruses.

## 275 **Composite risk map**

276 To visualize the aggregated risk at the global scale, we combine the three individual risk components  
277 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model.<sup>59</sup> In  
278 this approach, every risk component gets assigned a component in the RGB color model (phylogenetic  
279 diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid

280 RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no  
281 phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for  
282 each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the  
283 risk as colors diverge towards combinations of values for three risk components. Out of the possible  
284 combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic  
285 diversity and low viral sharing,<sup>60</sup> in that this allows multiple independent host-virus coevolutionary  
286 dynamics to take place in the same location. In the colorimetric space, this corresponds to yellow – because  
287 the HSV space is more amenable to calculations for feature extraction,<sup>61</sup> we measured the risk level by  
288 calculating the angular distance of the hue of each pixel to a reference value of 60 (yellow), and weighted  
289 this risk level by the value component. Specifically, given a pixel with colorimetric coordinates  $(h, s, v)$ , its  
290 ranged weighted risk value is

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

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

## 292 **Viral phyogeography and evolutionary diversification**

293 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed  
294 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data  
295 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR  
296 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR  
297 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated  
298 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained  
299 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and  
300 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East  
301 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus  
302 RdRp sequences were then aligned using MAFFT<sup>62</sup> v1.4.0 (Algorithm FFT-NS-2, Scoring matrix 200PAM /  
303 k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree reconstructed in  
304 IQ-TREE<sup>63</sup> v1.6.12 with ModelFinder<sup>64</sup> ultrafast bootstrap approximation<sup>65</sup> with a general time reversible  
305 model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of nucleotide

306 substitution (GTR+F+R5).

307 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat  
308 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the  
309 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary  
310 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary  
311 distinctiveness<sup>66</sup> for each of the viruses in the tree, then averaged these at the bat species level, projected  
312 these values onto their geographic distributions, and averaged across every bat found in a given pixel. As  
313 such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral  
314 community believed to be associated with a particular subset of bats present.

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

316 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
317 biogeography of their hosts. To test this idea, we loosely adapted a method from,<sup>67,68</sup> who proposed a  
318 phylogenetic method for the delineation of animal biogeographic regions. In their original method, a  
319 distance matrix - where each row or column represents a geographic raster’s grid cell, and the dissimilarity  
320 values are the “beta diversity similarity” of their community assemble - undergoes non-metric  
321 multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a  
322 four-color bivariate map. Here, we build on this idea with an entirely novel methodology. First, we  
323 measure the phylogenetic distance between the different viruses in the betacoronaviruses tree by using the  
324 cophenetic function in ape;<sup>69</sup> subsequently, we take a principal components analysis of that distance  
325 matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional  
326 space. We then take the first two principal components and, as with the evolutionary distinctiveness  
327 analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

328 **References**

- 329 1.
- 330 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 331 2.
- 332 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 333 3.
- 334 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 335 4.
- 336 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).
- 337 5.
- 338 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 339 6.
- 340 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).
- 341 7.
- 342 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 343 8.
- 344 Thompson, J. N. *The Coevolutionary Process*. (University Of Chicago Press, 1994).
- 345 9.
- 346 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 347 10.
- 348 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).  
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 349 11.
- 350 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).

- 351 12.
- 352 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).
- 353 13.
- 354 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).
- 355 14.
- 356 Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *The Journal of General Virology* **94**, 1028–1038  
(2013).
- 357 15.
- 358 Góes, L. G. B. *et al.* Novel Bat Coronaviruses, Brazil and Mexico. *Emerging Infectious Diseases* **19**, 1711–1713  
(2013).
- 359 16.
- 360 Góes, L. G. B. *et al.* Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil.  
*Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious  
Diseases* **44**, 510–513 (2016).
- 361 17.
- 362 Brandão, P. E. *et al.* A coronavirus detected in the vampire bat Desmodus rotundus. *Brazilian Journal of  
Infectious Diseases* **12**, 466–468 (2008).
- 363 18.
- 364 Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
- 365 19.
- 366 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).
- 367 20.
- 368 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).
- 369 21.
- 370 Olival, K. J. *et al.* Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A  
case study of bats. *PLoS pathogens* **16**, e1008758 (2020).
- 371 22.
- 372 Ammerman, L. K., Lee, D. N. & Tipps, T. M. First molecular phylogenetic insights into the evolution of free-  
tailed bats in the subfamily Molossinae (Molossidae, Chiroptera). *Journal of Mammalogy* **93**, 12–28 (2012).

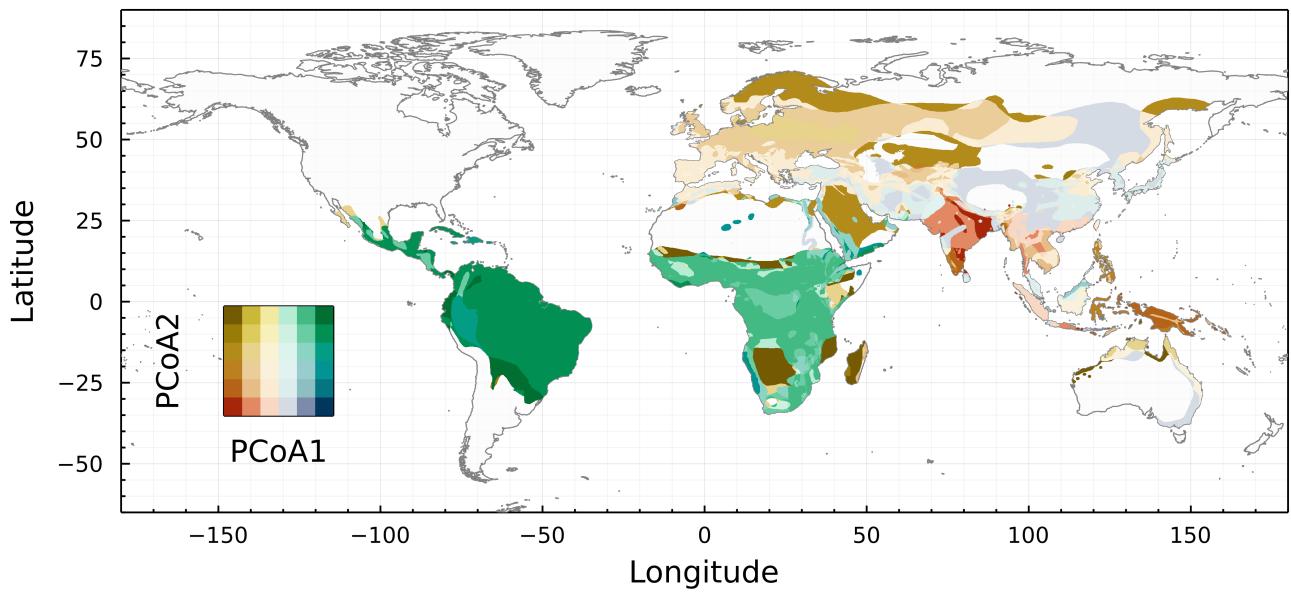
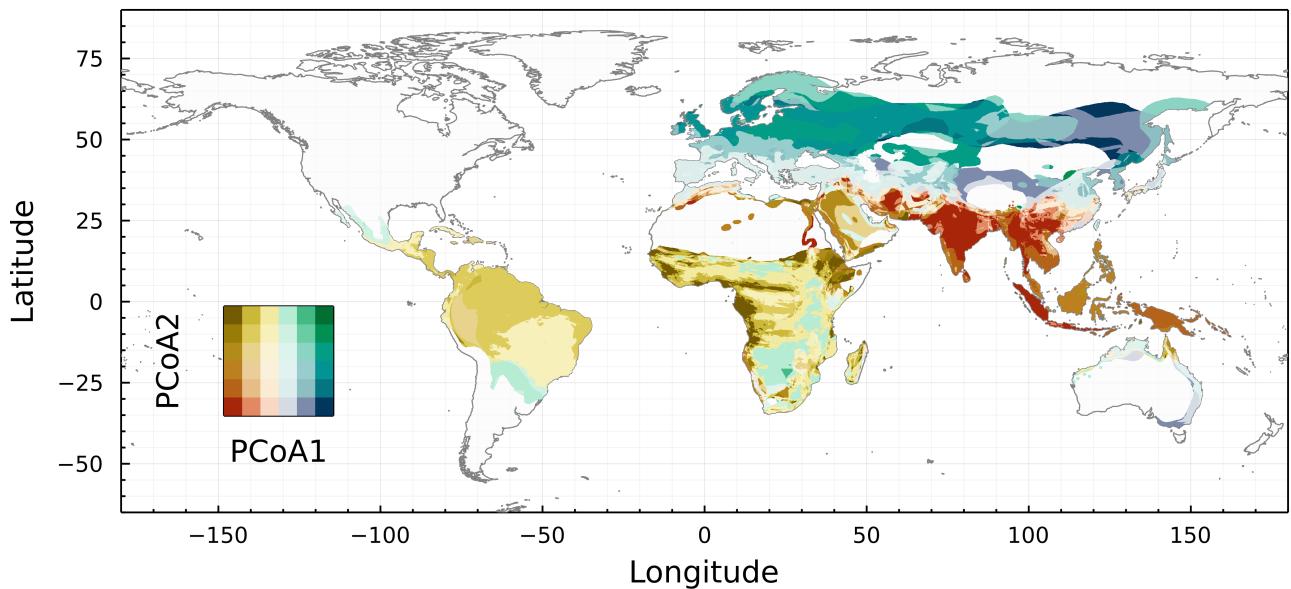
- 373 23.
- 374 Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Frontiers in Immunology* **11**, 26 (2020).
- 375 24.
- 376 Shi, J. J. *et al.* A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with Evidence for Widespread Panmixia. *Acta Chiropterologica* **16**, 279–292 (2014).
- 377 25.
- 378 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).
- 379 26.
- 380 Rulli, M. C., D'Odorico, P., Galli, N. & Hayman, D. T. Land-use change and the livestock revolution increase the risk of zoonotic coronavirus transmission from rhinolophid bats. *Nature Food* **2**, 409–416 (2021).
- 381 27.
- 382 Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* **17**, 181–192 (2019).
- 383 28.
- 384 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).
- 385 29.
- 386 Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. *Trends in ecology & evolution* **32**, 55–67 (2017).
- 387 30.
- 388 Worobey, M. *et al.* The Huanan market was the epicenter of SARS-CoV-2 emergence. *Zenodo* (2022)  
doi:[10.5281/zenodo.6299600](https://doi.org/10.5281/zenodo.6299600).
- 389 31.
- 390 Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
- 391 32.
- 392 Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature microbiology* **5**, 1408–1417 (2020).
- 393 33.
- 394 Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS pathogens* **13**, e1006698 (2017).

- 395 34.
- 396 Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica* **33**, 104–107 (2018).
- 397 35.
- 398 Memish, Z. A. *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerging infectious diseases* **19**, 1819 (2013).
- 399 36.
- 400 Müller, M. A. *et al.* MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983–1997. *Emerging infectious diseases* **20**, 2093 (2014).
- 401 37.
- 402 Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10**, 1030–1037 (2004).
- 403 38.
- 404 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).
- 405 39.
- 406 Simmons, N. B. & Cirranello, A. L. Bat Species of the World: A taxonomic and geographic database. <https://batnames.org/> (2020).
- 407 40.
- 408 Kasso, M. & Balakrishnan, M. Ecological and Economic Importance of Bats (Order Chiroptera). *ISRN Biodiversity* **2013**, e187415 (2013).
- 409 41.
- 410 *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).
- 411 42.
- 412 Williams-Guillén, K., Perfecto, I. & Vandermeer, J. Bats Limit Insects in a Neotropical Agroforestry System. *Science* **320**, 70–70 (2008).
- 413 43.
- 414 Gonsalves, L., Bicknell, B., Law, B., Webb, C. & Monamy, V. Mosquito Consumption by Insectivorous Bats: Does Size Matter? *PLOS ONE* **8**, e77183 (2013).
- 415 44.
- 416 Gonsalves, L., Lamb, S., Webb, C., Law, B. & Monamy, V. Do mosquitoes influence bat activity in coastal habitats? *Wildlife Research* **40**, 10–24 (2013).

- 417 45.
- 418 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).
- 419 46.
- 420 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).
- 421 47.
- 422 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).
- 423 48.
- 424 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).
- 425 49.
- 426 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).
- 427 50.
- 428 Amman, B. R. *et al.* *Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and*  
*public health interest.* (FAO, 2011).
- 429 51.
- 430 IUCN. The IUCN Red List of Threatened Species. <https://www.iucnredlist.org> (2021).
- 431 52.
- 432 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).
- 433 53.
- 434 Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10 (1992).
- 435 54.
- 436 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).
- 437 55.
- 438 Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients  
and partitioning. *Ecology Letters* **16**, 951–963 (2013).

- 439 56.
- 440 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).
- 441 57.
- 442 Dansereau, G., Legendre, P. & Poisot, T. Evaluating ecological uniqueness over broad spatial extents using  
species distribution modelling. *Oikos* **n/a**, e09063 (2022).
- 443 58.
- 444 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).
- 445 59.
- 446 Seekell, D. A., Lapierre, J.-F. & Cheruvelil, K. S. A geography of lake carbon cycling. *Limnology and  
Oceanography Letters* **3**, 49–56 (2018).
- 447 60.
- 448 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).
- 449 61.
- 450 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).
- 451 62.
- 452 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).
- 453 63.
- 454 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).
- 455 64.
- 456 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).
- 457 65.
- 458 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).
- 459 66.
- 460 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).

- 461 67.
- 462 Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National  
Academy of Sciences* **104**, 5925–5930 (2007).
- 463 68.
- 464 Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions.  
*Journal of Biogeography* **37**, 2029–2053 (2010).
- 465 69.
- 466 Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in  
R. *Bioinformatics* **35**, 526–528 (2019).



**Figure 1: Bat and betacoronavirus biogeographic regions.** Phylogeography of bats (top) and viruses (bottom) is categorized based on analysis of bat distributions paired with bat or virus phylogeny. 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.

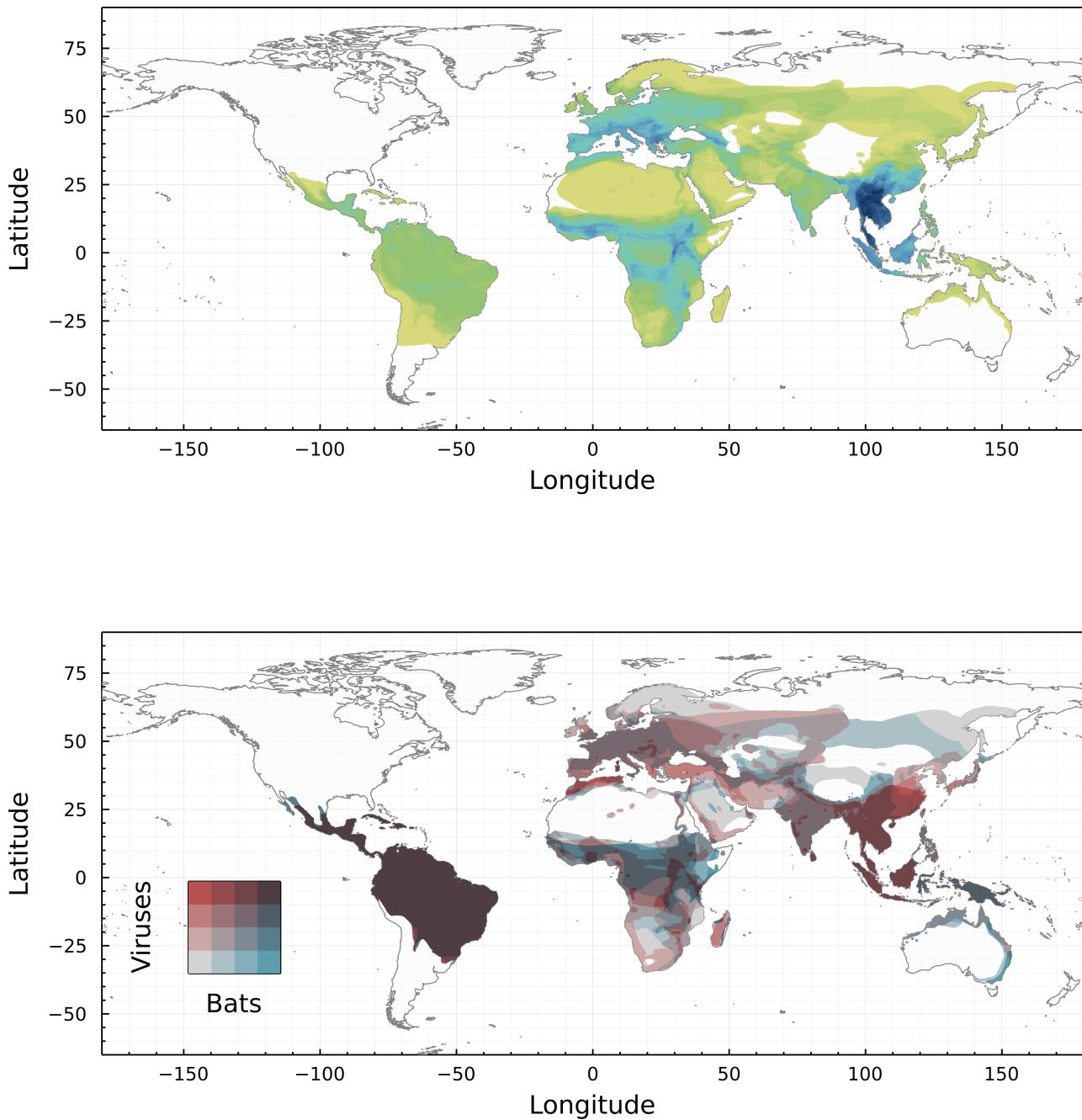
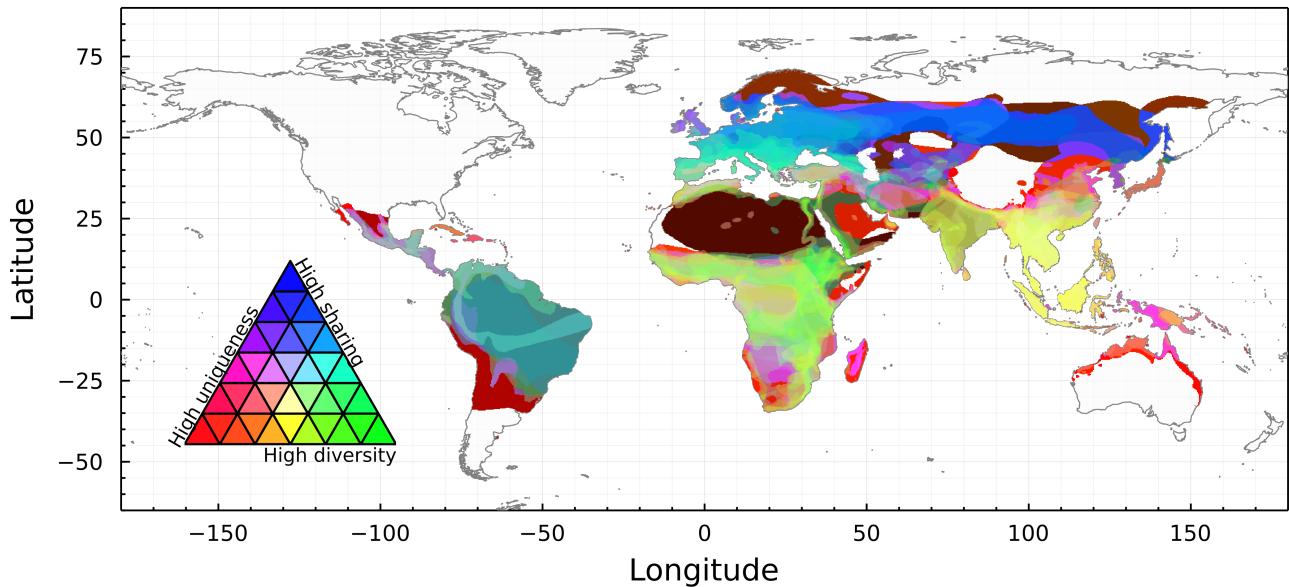


Figure 2: **Bat and betacoronavirus diversity.** Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey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



**Figure 3: Trivariate additive mapping of the components of risk.** Viral sharing runs from yellow (low) to blue (high); host phylogenetic diversity runs from pink (low) to high (green); and host compositional uniqueness runs from cyan (low) to red (high). The GMTC suggests that the highest evolutionary potential for emergence exists in unique and diverse host communities with low viral sharing, *i.e.* pixels around yellow. All components are scaled in brightness so that a pixel with no sharing, no phylogenetic diversity, and no compositional uniqueness would be black, and a pixel with maximal values for each would be white.

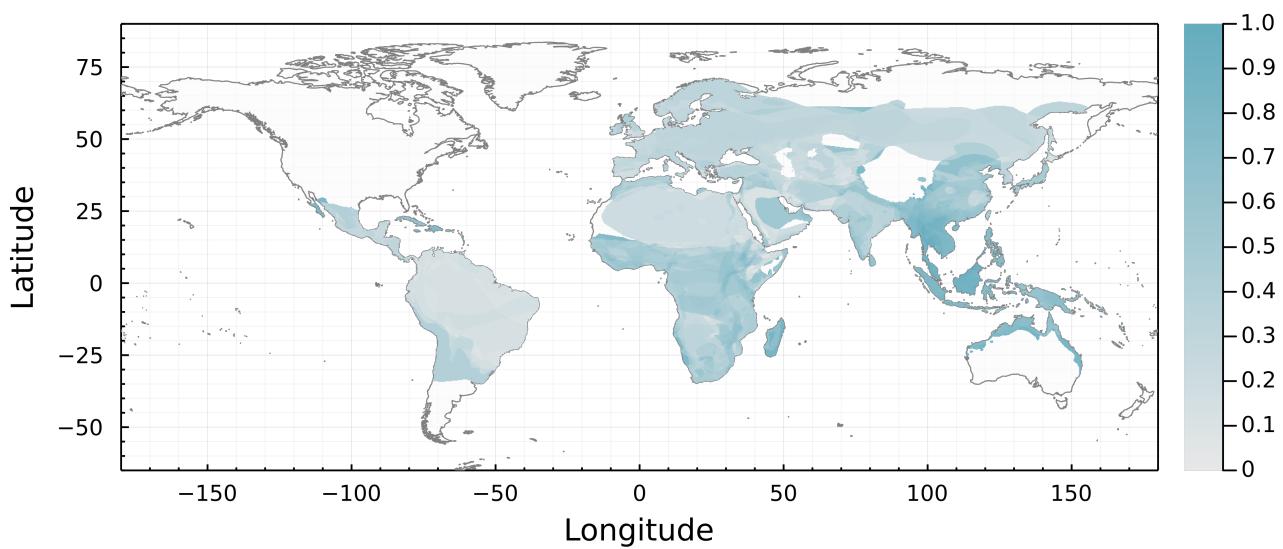


Figure 4: **Evolutionary potential for zoonotic emergence of bat-origin betacoronaviruses.** Risk is a composite measure of the color value and angular distance to the yellow hue in fig. 3 (see Methods).

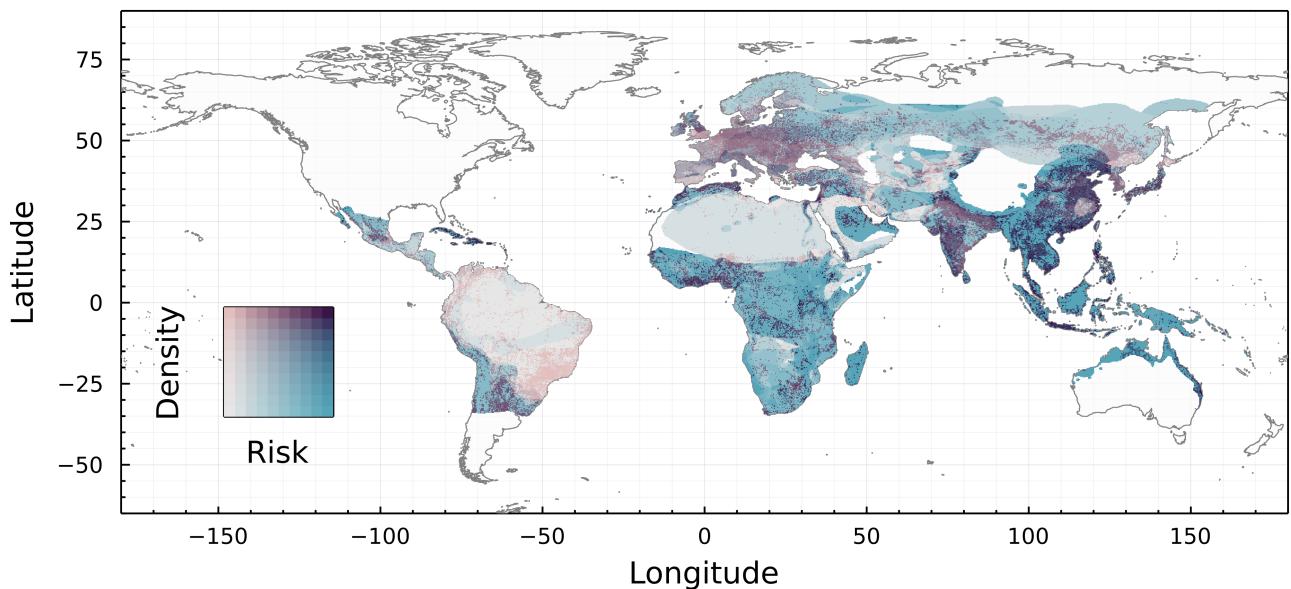


Figure 5: **Overlap between evolutionary potential and ecological opportunity for zoonotic emergence.** Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts 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.