

# The coevolutionary mosaic of *Betacoronavirus* spillover risk from bat hosts

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 reservoir richness alone, although a component of viral hazard, is not a sufficiently integrative predictor of risk. We offer alternative insights based on viral sharing, host compositional uniqueness, and host phylogenetic diversity 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 system may originate novel viruses in humans.

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 (Plowright et al. 2017). Although host richness is  
3 often used as a superficial proxy for spillover risk (Anthony et al. 2017, Ruiz-Aravena et al. 2022), these  
4 approaches oversimplify the relevant interspecific heterogeneity in immunology, behavior, and other  
5 traits, and therefore overlook unique host pools that allow for the rapid evolution of highly divergent  
6 viruses (Agosta et al. 2010). In the case of generalist pathogens like betacoronaviruses, there is conceptual  
7 and empirical support to the idea that these community-level mechanisms are even more important  
8 (Power and Mitchell 2004), particularly given that cross-species transmission may, as a rule, structure viral  
9 evolution more than co-divergence with hosts [Geoghegan]. This creates a disconnect between  
10 coevolutionary theory (including empirical evidence from virology) and most existing ecological  
11 frameworks for mapping spillover risk.

12 The Geographic Mosaic Theory of Coevolution (GMTC; Thompson 2005) attempts to explicitly connect  
13 microevolutionary dynamics to the macroecology and biogeography of symbiotic interactions. The GMTC  
14 posits that coevolutionary processes among pairs (Thompson 1994) or complexes (Janzen 1980) of species  
15 are structured in space by the rippling effects of abiotic conditions onto evolutionary mechanism, giving  
16 rise to fragmented systems with different structure and ecologically dynamics over large spatial extents  
17 (see e.g. Price 2002). The GMTC predicts a spatial fragmentation of coevolutionary dynamics under the  
18 joint action of three processes (see notably Gomulkiewicz et al. 2007): coevolutionary hot- and coldspots,  
19 which appear when the intensity of *interaction* (in terms of reciprocal fitness consequences) varies  
20 spatially; selection mosaics, wherein the intensity of *selection* varies across space, driven by both the biotic  
21 complexity of the community (locally diverse hosts and viruses are more biotically complex) and the local  
22 favorability of the environment (**Thrall2007?**); and trait remixing, which occurs when coevolutionary  
23 dynamics are driven by the arrival (or departure) of *functional traits*, through changes in community  
24 composition due to invasions, meta-community dynamics, and dispersal.

25 Here, we apply the GMTC to explore and explain the global biogeography of betacoronaviruses, the group  
26 that includes SARS-CoV, MERS-CoV, and SARS-CoV-2. In their bat reservoirs, coronaviruses evolve  
27 through a mix of host jumps, recombination among disparate lineages, and, to a lesser degree,  
28 co-divergence with their hosts [Anthony]—a mix of mechanisms that creates a complex and nonlinear  
29 relationship between host diversity and viral emergence. Working from a recently published database of  
30 bat hosts of betacoronaviruses, we develop the first global maps of both host and virus evolutionary

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

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

45 Bats, the second most diverse group of mammals, are found worldwide and serve as the main animal  
46 reservoir for different strains of betacoronaviruses (Drexler et al. 2014). This has attracted attention to  
47 areas where high diversity of bats, and therefore presumably high diversity of betacoronaviruses, can be an  
48 important issue for human health (Calisher et al. 2006, Moratelli and Calisher 2015). By overlaying the  
49 IUCN range maps for confirmed bat hosts of betacoronaviruses [fig. 1; top], we see that the the main  
50 hotspots (both in terms of size and higher values) of host richness are primarily South-Eastern Asia, parts  
51 of Southern Europe, and to a lesser extent parts of Africa in the -25-0 range of latitudes. The description of  
52 host richness is an important first step towards understanding risk, as previous research (Anthony et al.  
53 2017, Mollentze and Streicker 2020) states that locally diverse bat communities could maintain more  
54 viruses and hence, a higher probability of having a pathogen that could represent a risk for human health.

55 [Figure 1 about here.]

56 Nevertheless, locally diverse and virus-rich bat communities could represent an increased risk of spillover  
57 under climate change through the creation of novel interactions (Carlson et al. 2022), and therefore the  
58 diversity of *Betacoronavirus* strains should similarly be accounted for. In fig. 1 (bottom), we contrast the

59 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness  
60 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as Yangochiroptera and  
61 Yinpterochiroptera elsewhere (Teeling et al. 2005, Springer 2013). Specifically, we would expect that the  
62 so-called “New World” group of bats, being more evolutionary distinct, would also have evolutionary  
63 distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a distinct  
64 bat-betacoronaviruses complex. This can be explained by the fact that Yangochiroptera, although not  
65 limited to the western hemisphere, contain the highly diverse adaptive radiation in the Phyllostomidae  
66 (Villalobos and Arita 2010), which is restricted to the western hemisphere. By contrast, South-Eastern  
67 Asia has a lot of non-evolutionary distinct bats, who nevertheless hosted evolutionary-distinct viruses.  
  
68 It is noteworthy that outside of South America, viral evolutionary distinctiveness does not accurately track  
69 host diversity, with some areas having over-distinct viruses (eastern China but, oddly, not the rest of  
70 southeast Asia). There are a number of likely explanations. First, given the richness of bats in southeast  
71 Asia, many betacoronaviruses likely remain to be discovered in this region. Indeed, global predictions by  
72 Becker et al. (2022) highlight that southeast Asia is a likely hotspot of unconfirmed hosts of  
73 betacoronaviruses, which would likely result in additional viral discoveries. This idea is unsurprising  
74 given the growing realization, especially since the emergence of SARS-CoV-2, that unique lineages of  
75 similar viruses are widespread in bats but still mostly undescribed. The most distinct  
76 bats-betacoronaviruses complex is found in South America, a region with a comparatively lower number  
77 of hosts; this matches with the isolation through variance of the host group, and may highlight a different  
78 co-evolutionary dynamic. Alternatively, this distinctiveness hotspot may be a product of under-sampling:  
79 South-America is one of the places where the fewest *Betacoronavirus* sequences have been discovered  
80 (Anthony et al. 2017, Olival et al. 2017, Allen et al. 2017), resulting in sparser phylogenetic tree, thereby  
81 artificially inflating distinctiveness. Adding more viruses would bring the distinctiveness of known  
82 sequences down. Finally, South-America is the range of Phyllostomidae, a group of bats that underwent  
83 explosive diversification events (Villalobos and Arita 2010), which may drive the emergence of multiple  
84 viral lineages.

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

86 Despite the difference in evolutionary distinctiveness globally, there are reasons to expect that the  
87 phylogeography of bats and betacoronaviruses should show some degree of congruence (Letko et al. 2020,

88 Van Brussel and Holmes 2022). In particular, this should be the case if viruses can circulate among hosts  
89 and co-evolve with local hosts communities, making their evolutionary process more than a byproduct of  
90 host evolution. High density of hosts sharing the same virus (albeit possibly different strains) can drive or  
91 result from evolution of the bat antiviral immune system, resulting in spatially distinct immunological  
92 responses, as evidenced in several bat species (Banerjee et al. 2020). Immune characteristics that allow  
93 bats to be better adapted to infection by emerging viruses (Gorbunova et al. 2020, Irving et al. 2021), in  
94 addition to being hardcoded in their genome (Jebb et al. 2020), may be related to a wide variety of diets  
95 (Banerjee et al. 2020, Moreno Santillán et al. 2021), themselves likely to be driven by spatial effects,  
96 especially at the local scale – bats, indeed, occupy a variety of environments, and therefore display a  
97 variety of adaptations to these environments (Muylaert et al. 2022).

98 [Figure 2 about here.]

99 In fig. 2, we show a projection of the phylogeographic signal of bats (top) and viruses (bottom) in space;  
100 the distinct groupings (represented by different colors symbolizing positions in the subspace formed by  
101 the first two axes of the PCoA) are essentially equivalent between the two groups, and can be coarsely  
102 delineated as southeast Asia, Eurasia above a northing of 25, and Africa and south America. These results  
103 suggest that, although the evolutionary distinctiveness of the bat-beta冠状病毒 complex varies  
104 spatially, the system shows an important degree of spatial consistency, with a reduced number of  
105 bioregions. Available information describing the spillover of zoonotic betacoronaviruses of bat origin  
106 where data was available before and up through the COVID-19 pandemic puts spillover events of  
107 SARS-CoV-2 in Wuhan, China; SARS-CoV in Guangdong, China based on the presence of closest known  
108 viruses circulating in nature, and a nearby location where serological (antibody) evidence has indicated  
109 human exposure to SARS-like viruses (Wang et al. 2018); MERS-CoV in Saudi Arabia based on index cases  
110 available from a recently-published compendium of cases (Ramshaw et al. 2019). For the latest event,  
111 most if not all index cases are presumed to be camel-to-human transmission, and the precise origin point  
112 (if it exists) of MERS-CoV in bats is uncertain. Recent recombinant canine coronavirus spillover events in  
113 Haiti (Lednicky et al. 2021) and Europe (Vlasova et al. 2022) are not relevant here, as bats' involvement in  
114 these cycles of transmission have been supposed to be non-existent. These index cases fall within different  
115 phylogeographic bioregions (fig. 2), which further highlight the issue that different host-virus sub-systems  
116 may lead to widespread emergence.

117 **Coevolution-informed spillover risk is different in space**

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

129 [Figure 3 about here.]

130 From the perspective of spillover risk, the most important combination of factors is a high phylogenetic  
131 diversity of hosts with low viral sharing; this, essentially, means that very different betacoronaviruses  
132 could co-exist within the same place. This is particularly the case given that betacoronaviruses often  
133 evolve and even achieve host shifts through recombination, which requires the co-occurrence of  
134 sufficiently distinct viruses to be a major driver of emergence. In fig. 3, this corresponds to yellow to pale  
135 green areas, which are essentially limited to South-Eastern Asia, and to some part of Sub-Saharan Africa.  
136 Adopting a geographic mosaic theory perspective on risk, other regions of the world are of lesser concern  
137 fig. 4. Our risk decomposition does not account for viral diversity or distinctiveness. The simple rationale  
138 behind it is that the acquisition of viral data is rarely disconnected from the acquisition of host data. There  
139 are more sources of information on hosts than on viruses, allowing to develop a host-centric perspective  
140 on risk (although this estimate would be more accurate with viral traits related to *e.g.* ability to switch hosts  
141 or pathogenic potential). Areas with high bat diversity and high turnover *may* facilitate the evolutionary  
142 radiation of viruses, matching previous findings that the diversification of bat coronaviruses is driven  
143 largely by host shifts (inter-genus or higher levels of cross-species transmission) and, to a lesser degree,  
144 cospeciation and sharing (intra-genus cross-species transmission; Anthony et al. 2017). This  
145 diversification is not an actual risk factor for spillover itself, but acts downstream of a spillover event by

146 increasing the random chance of the emergence of a virus with the raw genomic components required for  
147 the potential to infect humans.

148 [Figure 4 about here.]

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

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

157 Based on the previous result, we extracted the risk component from the composite map (see Methods), to  
158 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. As this map  
159 represents the potential risk, it must be weighed by the potential for contacts with humans. As a proxy for  
160 this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a reasonable  
161 proxy for the density of humans per unit area, which increases the probability of pathogen spread more  
162 widely (Hazarie et al. 2021). Since human activity is required to amplify the frequency of virus encounters  
163 and thus create areas of viral amplification, mapping the potential risk against measures of land use is  
164 required to generate a more actionable assessment of risk. This map is presented in fig. 5. Most of South  
165 America and Europe are at comparatively lower risk, as although densely populated, settlements tend to  
166 be in areas with lower potential risk. Regions like Malaysia and the North coast of Australia have a high  
167 risk component, but should represent a relatively lower effective risk due to low human density. However,  
168 this mapping reveals that South-East Asia, the Indian subcontinent, and parts of sub-Saharan Africa, are  
169 at high risk due to the overlap between built areas and bat communities representing more opportunities  
170 for cross-species transmission of betacoronaviruses. In looking for the origins of SARS in China, Xu et al.  
171 (2004) present serological evidence that strongest human-animal contact results in higher risk of virus  
172 exposure, regardless of the animal species, but that different types of contact had different impacts. Ideally,  
173 finer-grained information about human activity (rather than human presence through anthropisation)

174 could allow to partition this risk further, albeit at the cost of more hypotheses required to estimate the  
175 amount of risk represented by each activity. Our map of risk overlays with recent results from Cohen et al.  
176 (2022) – areas of purported high risk/diversification potential (Madagascar, South-America) overlay with  
177 sampling gaps for *Betacoronavirus*, stressing the need for spatially targeted monitoring and discovery.

178 [Figure 5 about here.]

## 179 Conclusion

180 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to  
181 human health (Letko et al. 2020, Van Brussel and Holmes 2022). Chiropterans emerged around 64 million  
182 years ago and are one of the most diverse mammalian orders, with an estimated richness of more than  
183 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020). They exhibit a broad variety of habitat  
184 use, behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of  
185 several ecosystem services, tied to important ecosystem-derived benefits to human (Kasso and  
186 Balakrishnan 2013). For example, bats are an essential component of many seed-dispersal networks  
187 (Mello et al. 2011). Over two-thirds of bats are known to be either obligate or facultative insectivores,  
188 therefore actively contributing for agricultural pest control (Williams-Guillén et al. 2008, Voigt and  
189 Kingston 2016), and vectors of pathogens that put a risk on human health (Gonsalves et al. 2013a, b).  
190 Because bats are globally distributed and have a long evolutionary history, phylogeographic and  
191 biogeographic approaches are required to shed light on the contemporary distribution of coevolutionary  
192 processes between bats and the pathogens they host. Not all areas in which bats, viruses, and human are  
193 co-occurring are facing a risk of spillover towards human populations, and the areas in which this risk exist  
194 may not be facing risks of the same nature and magnitude.

195 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries  
196 like highly divergent nobcoviruses in Madagascar and the previously-neglected adaptive radiation of  
197 sarbecoviruses outside of southern China and throughout southeast Asia. In doing so, it advances  
198 ecological theory beyond the current state of the art for global maps of emergence risk. For example,  
199 previous studies that have used host richness as proxy have predicted a high diversity of unsampled bat  
200 viruses [Olival], bat coronaviruses [Anthony], and even specifically betacoronaviruses [Becker] in both the

201 Amazon and southeast Asia. While we find that both regions are characterized by highly divergent host  
202 and viral communities, our framework identifies key differences between the regions. We find that Latin  
203 America is a hotspot of both host and viral distinctiveness, suggesting that this branch of the  
204 bat-betacoronavirus complex may be undergoing independent evolutionary dynamics from the rest of the  
205 global pool, but with limited potential for viral diversification—a finding that is supported by previous  
206 work indicating a higher rate of codivergence in Latin America [Anthony]. In contrast, in southeast Asia,  
207 host richness and viral distinctiveness are high but sharing is low; this suggests a different type of  
208 evolutionary dynamics that could generate high local diversity of viruses through host switching and viral  
209 recombination (see e.g. (Latinne et al. 2020), as well as the discovery of recombinant viruses that share  
210 genetic material from both the SARS-CoV and SARS-CoV-2 branches of the Sarbecovirus lineage [Wu]).  
211 Both of these regions are priority areas for sampling, especially given predictions by Becker et al. (2022)  
212 that they contain many bat hosts of undiscovered betacoronaviruses. However, both the evolutionary and  
213 ecological aspects of emergence risk are likely higher in southeast Asia—a fact that will only become more  
214 relevant, as bats track shifting climates and exchange viruses with other species, creating a hotspot of  
215 cross-species transmission unique to the region [Carlson].

216 [Tim tinker with para about conservation] There are several factors that drive changes in the diversity of  
217 bats (Alves et al. 2018), but human activities' effects on the ecosystem (like modifications of land use)  
218 could significantly decrease it. Therefore, it can be suggested that changes in the diversity of  
219 betacoronaviruses in bats are linked to their biogeographic variation, and human population density and  
220 other anthropogenic factors are decisive moderators for its implications in public health. With the  
221 increase of contact between humans and potential hosts, we also increase the risk of emergence of novel  
222 diseases (Johnson et al. 2020), as previous studies on RNA viruses suggest the importance of host  
223 phylogeography at the time of virus dispersal (Gryseels et al. 2017). One of these scenarios where  
224 interaction between bats and humans can occur can be seed dispersal in tropical agroecosystems. It opens  
225 the discussion of whether the fruits thrown by bats not only disperse seeds but could also be a source of  
226 indirect interaction between viruses of bat origin and humans (Deshpande et al. 2022). This represents a  
227 challenge for conservation strategies and disease ecology since some areas can have both potential for the  
228 acquisition of zoonotic viruses and bat-human interactions; in particular, the challenge lies in the fact that  
229 actual exposure must then be quantified accounting for several transmission scenarios, including both  
230 direct and indirect bat - human interaction.

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

239 **Methods**

240 **Known *Betacoronavirus* hosts**

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

248 **Bat occurrences**

249 We downloaded the rangemap of every current bat species that was classified as an empirically  
250 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data (IUCN 2021).  
251 The range maps were subsequently rasterized using the rasterize function from GDAL (Rouault et al.  
252 2022) at a resolution of approximately 100kmx100km. For every pixel in the resulting raster where at least  
253 one bat host of *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which  
254 was used to calculate the following risk assessment components: bat phylogenetic diversity, bat  
255 compositional uniqueness, and predicted viral sharing risk.

256 **Bat phylogenetic diversity**

257 For every pixel, we measured Faith’s Phylogenetic Diversity (Faith 1992) based on a recent synthetic tree  
258 with robust time calibration, covering about 6000 mammalian species (Upham et al. 2019). Faith’s PD  
259 measures the sum of unique branches from an arbitrary root to a set of tips, and comparatively larger  
260 values indicate a more phylogenetic diverse species pool. We measured phylogenetic diversity starting  
261 from the root of the entire tree (and not from Chiroptera); this bears no consequences on the resulting  
262 values, since all branches leading up to Chiroptera are only counted one per species pool, and (as we  
263 explain when describing the assembly of the composite risk map), all individual risk components are  
264 ranged in [0,1]. This measure incorporates a richness component, which we chose not to correct for; the

265 interpretation of the phylogenetic diversity is therefore a weighted species richness, that accounts for  
266 phylogenetic over/under-dispersal in some places.

## 267 **Bat compositional uniqueness**

268 For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres  
269 2013); LCBD works from a species-data matrix (traditionally noted as  $\mathbf{Y}$ ), where species are rows and sites  
270 are columns, and a value of 1 indicates occurrence. We extracted the  $\mathbf{Y}$  matrix assuming that every pixel  
271 represents a unique location, and following best practices (Legendre and Condit 2019) transformed it  
272 using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw  
273 community data is particularly important for two reasons: first, it prevents the artifact of richer sites  
274 having higher importance; second, it removes the effect of overall species richness, which is already  
275 incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a  
276 community that is on average more dissimilar in species composition than what is expected knowing the  
277 entire matrix, i.e. a more unique community. Recent results by Dansereau et al. (2022) shows that LCBD  
278 measures are robust with regards to spatial scale, and are therefore applicable at the global scale.

## 279 **Viral sharing between hosts**

280 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a  
281 previously published generalized additive mixed model of virus sharing by a tensor function of  
282 phylogenetic distance and geographic range overlap (Albery et al. 2020). This network stores pairwise  
283 values of viral community similarity. To project viral sharing values into a single value for every pixel, we  
284 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant  
285 bat assemblage is likely to be proficient at exchanging viruses.

## 286 **Composite risk map**

287 To visualize the aggregated risk at the global scale, we combine the three individual risk components  
288 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model  
289 (Seekell et al. 2018). In this approach, every risk component gets assigned a component in the RGB color  
290 model (phylogenetic diversity is green, compositional uniqueness is red, and viral sharing is blue). In

order to achieve a valid RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the risk as colors diverge towards combinations of values for three risk components. Out of the possible combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic diversity and low viral sharing (Gomulkiewicz et al. 2000, Cavender-Bares et al. 2009), in that this allows multiple independent host-virus coevolutionary dynamics to take place in the same location. In the colorimetric space, this correspond to yellow – because the HSV space is more amenable to calculations for feature extraction (see e.g. Keke et al. 2010), we measured the risk level by calculating the angular distance of the hue of each pixel to a reference value of 60, and weighted this risk level by the value component. Specifically, given a pixel with colorimetric coordinates  $(h, s, v)$ , its ranged weighted risk value is

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

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

## 304 **Viral phyogeography and evolutionary diversification**

To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained words indicating recombinant or laboratory strains including “patent,” “mutant,” “GFP,” and “recombinant.” We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus RdRp sequences were then aligned using MAFFT v 1.4.0 [Katoh and Standley (2013); Algorithm FFT-NS-2, Scoring matrix 200PAM / k=2, gap open penalty 1.53m offset value 0.123] and a maximum likelihood tree reconstructed in IQ-TREE v 1.6.12 (Nguyen et al. 2015) with ModelFinder

317 (Kalyaanamoorthy et al. 2017) ultrafast bootstrap approximation (Hoang et al. 2018) with a general time  
318 reversible model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of  
319 nucleotide substitution (GTR+F+R5).

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

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

329 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
330 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,  
331 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their  
332 original method, a distance matrix - where each row or column represents a geographic raster’s grid cell,  
333 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes  
334 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected  
335 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel  
336 methodology. First, we measure the phylogenetic distance between the different viruses in the  
337 betacoronaviruses tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently,  
338 we take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this  
339 case) to project the viral tree into an n-dimensional space. We then take the first two principal  
340 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value  
341 and projected them using a four-color bivariate map.

342 **References**

- 343 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and  
344 implications for emerging infectious disease. - *Zoologia (Curitiba)* 27: 151–162.
- 345 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -  
346 *Nature Communications* 11: 2260.
- 347 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - *Nature*  
348 *Communications* in press.
- 349 Alves, D. M. C. C. et al. 2018. Geographic variation in the relationship between large-scale environmental  
350 determinants and bat species richness. - *Basic and Applied Ecology* 27: 1–8.
- 351 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - *Virus Evolution* in press.
- 352 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - *Frontiers in Immunology* 11: 26.
- 353 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -  
354 *The Lancet Microbe* in press.
- 355 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - *Clinical Microbiology*  
356 *Reviews* 19: 531–545.
- 357 Carlson, C. J. et al. 2022. Climate change increases cross-species viral transmission risk. - *Nature*: 1–1.
- 358 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - *Ecol. Lett.*  
359 12: 693–715.
- 360 Cohen, L. E. et al. 2022. Sampling strategies and pre-pandemic surveillance gaps for bat coronaviruses.:  
361 2022.06.15.496296.
- 362 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species  
363 distribution modelling. - *Oikos* n/a: e09063.
- 364 Deshpande, K. et al. 2022. Forbidden fruits? Ecosystem services from seed dispersal by fruit bats in the  
365 context of latent zoonotic risk. - *Oikos (Copenhagen, Denmark)*: oik.08359.
- 366 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of  
367 SARS. - *Antiviral Research* 101: 45–56.

- 368 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1–10.
- 369 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -  
370 The American Naturalist 156: 156–174.
- 371 Gomulkiewicz, R. et al. 2007. Dos and don'ts of testing the geographic mosaic theory of coevolution. -  
372 Heredity 98: 249–258.
- 373 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:  
374 10–24.
- 375 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE  
376 8: e77183.
- 377 Gorbunova, V. et al. 2020. The World Goes Bats: Living Longer and Tolerating Viruses. - Cell Metabolism  
378 32: 31–43.
- 379 Gryseels, S. et al. 2017. When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure  
380 in the Spatial Spread of Arenaviruses (JH Kuhn, Ed.). - PLOS Pathogens 13: e1006073.
- 381 Hazarie, S. et al. 2021. Interplay between population density and mobility in determining the spread of  
382 epidemics in cities. - Communications Physics 4: 191.
- 383 Hoang, D. T. et al. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. - Molecular Biology  
384 and Evolution 35: 518–522.
- 385 Irving, A. T. et al. 2021. Lessons from the host defences of bats, a unique viral reservoir. - Nature 589:  
386 363–370.
- 387 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.  
388 - PLOS ONE 2: e296.
- 389 IUCN 2021. The IUCN Red List of Threatened Species.
- 390 Janzen, D. H. 1980. When is it Coevolution? - Evolution 34: 611–612.
- 391 Jebb, D. et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. - Nature 583:  
392 578–584.
- 393 Johnson, C. K. et al. 2020. Global shifts in mammalian population trends reveal key predictors of virus  
394 spillover risk. - Proceedings of the Royal Society B: Biological Sciences 287: 20192736.

- 395 Kalyaanamoorthy, S. et al. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. -  
396 Nature Methods 14: 587–589.
- 397 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -  
398 ISRN Biodiversity 2013: e187415.
- 399 Katoh, K. and Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7:  
400 Improvements in Performance and Usability. - Molecular Biology and Evolution 30: 772–780.
- 401 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010  
402 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.
- 403 Kettenburg, G. et al. 2022. Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a  
404 Unique Evolutionary History for This Coronavirus Clade. - Frontiers in Public Health in press.
- 405 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - Proceedings of  
406 the National Academy of Sciences 104: 5925–5930.
- 407 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species  
408 distributions. - Journal of Biogeography 37: 2029–2053.
- 409 Latinne, A. et al. 2020. Origin and cross-species transmission of bat coronaviruses in China. - bioRxiv:  
410 The Preprint Server for Biology: 2020.05.31.116061.
- 411 Lednicky, J. A. et al. 2021. Isolation of a Novel Recombinant Canine Coronavirus from a Visitor to Haiti:  
412 Further Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. - Clinical  
413 Infectious Diseases: An Official Publication of the Infectious Diseases Society of America: ciab924.
- 414 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity  
415 coefficients and partitioning (H Morlon, Ed.). - Ecology Letters 16: 951–963.
- 416 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado  
417 Island forest dynamics plot, Panama. - Forest Ecosystems 6: 7.
- 418 Letko, M. et al. 2020. Bat-borne virus diversity, spillover and emergence. - Nature Reviews Microbiology  
419 18: 461–471.
- 420 Mello, M. A. R. et al. 2011. The Missing Part of Seed Dispersal Networks: Structure and Robustness of  
421 Bat-Fruit Interactions. - PLOS ONE 6: e17395.

- 422 Mollentze, N. and Streicker, D. G. 2020. Viral zoonotic risk is homogenous among taxonomic orders of  
423 mammalian and avian reservoir hosts. - *Proceedings of the National Academy of Sciences* 117:  
424 9423–9430.
- 425 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with  
426 emerging deadly viruses? - *Memórias do Instituto Oswaldo Cruz* 110: 1–22.
- 427 Moreno Santillán, D. D. et al. 2021. Large-scale genome sampling reveals unique immunity and metabolic  
428 adaptations in bats. - *Molecular Ecology*: mec.16027.
- 429 Muylaert, R. L. et al. 2022. Present and future distribution of bat hosts of sarbecoviruses: Implications for  
430 conservation and public health. - *Proceedings of the Royal Society B: Biological Sciences* 289:  
431 20220397.
- 432 Nguyen, L.-T. et al. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating  
433 Maximum-Likelihood Phylogenies. - *Molecular Biology and Evolution* 32: 268–274.
- 434 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - *Nature* 546:  
435 646–650.
- 436 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary  
437 analyses in R. - *Bioinformatics* 35: 526–528.
- 438 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across  
439 spatial scales. - *BMC ecology* 18: 18.
- 440 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - *Nature Reviews Microbiology* 15: 502–510.
- 441 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - *The American  
442 Naturalist* 164: S79–S89.
- 443 Price, P. W. 2002. Macroevolutionary Theory on Macroecological Patterns. - Cambridge University Press.
- 444 Ramshaw, R. E. et al. 2019. A database of geopositioned Middle East Respiratory Syndrome Coronavirus  
445 occurrences. - *Scientific Data* 6: 318.
- 446 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 447 Ruiz-Aravena, M. et al. 2022. Ecology, evolution and spillover of coronaviruses from bats. - *Nature  
448 Reviews Microbiology* 20: 299–314.

- 449 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - Limnology and Oceanography Letters 3:  
450 49–56.
- 451 Shi, J. J. et al. 2014. A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with  
452 Evidence for Widespread Panmixia. - Acta Chiropterologica 16: 279–292.
- 453 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic  
454 database.
- 455 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - Current Biology 23: R999–R1001.
- 456 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.  
457 - Science (New York, N.Y.) 307: 580–584.
- 458 Thompson, J. N. 1994. The Coevolutionary Process. - University of Chicago Press.
- 459 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.
- 460 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in  
461 ecology, evolution, and conservation. - PLOS Biology 17: e3000494.
- 462 Van Brussel, K. and Holmes, E. C. 2022. Zoonotic disease and virome diversity in bats. - Current Opinion  
463 in Virology 52: 192–202.
- 464 Villalobos, F. and Arita, H. T. 2010. The diversity field of New World leaf-nosed bats (Phyllostomidae). -  
465 Global Ecology and Biogeography 19: 200–211.
- 466 Vlasova, A. N. et al. 2022. Animal alphacoronaviruses found in human patients with acute respiratory  
467 illness in different countries. - Emerging Microbes & Infections 11: 699–702.
- 468 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,  
469 Eds.). - Springer International Publishing.
- 470 Wang, N. et al. 2018. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.  
471 - Virologica Sinica 33: 104–107.
- 472 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - Science 320:  
473 70–70.
- 474 Xu, R.-H. et al. 2004. Epidemiologic Clues to SARS Origin in China. - Emerging Infectious Diseases 10:  
475 1030–1037.

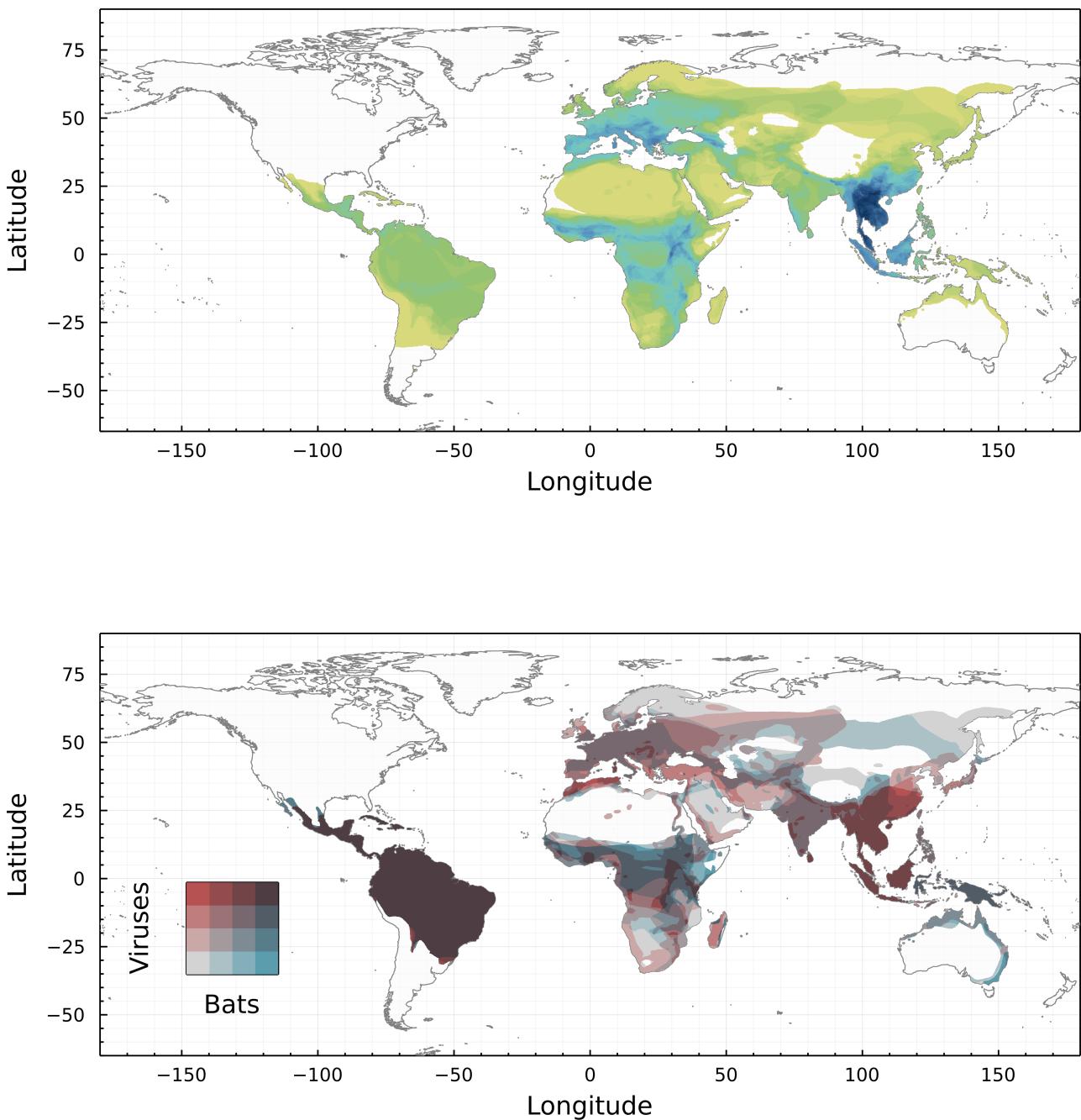


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).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

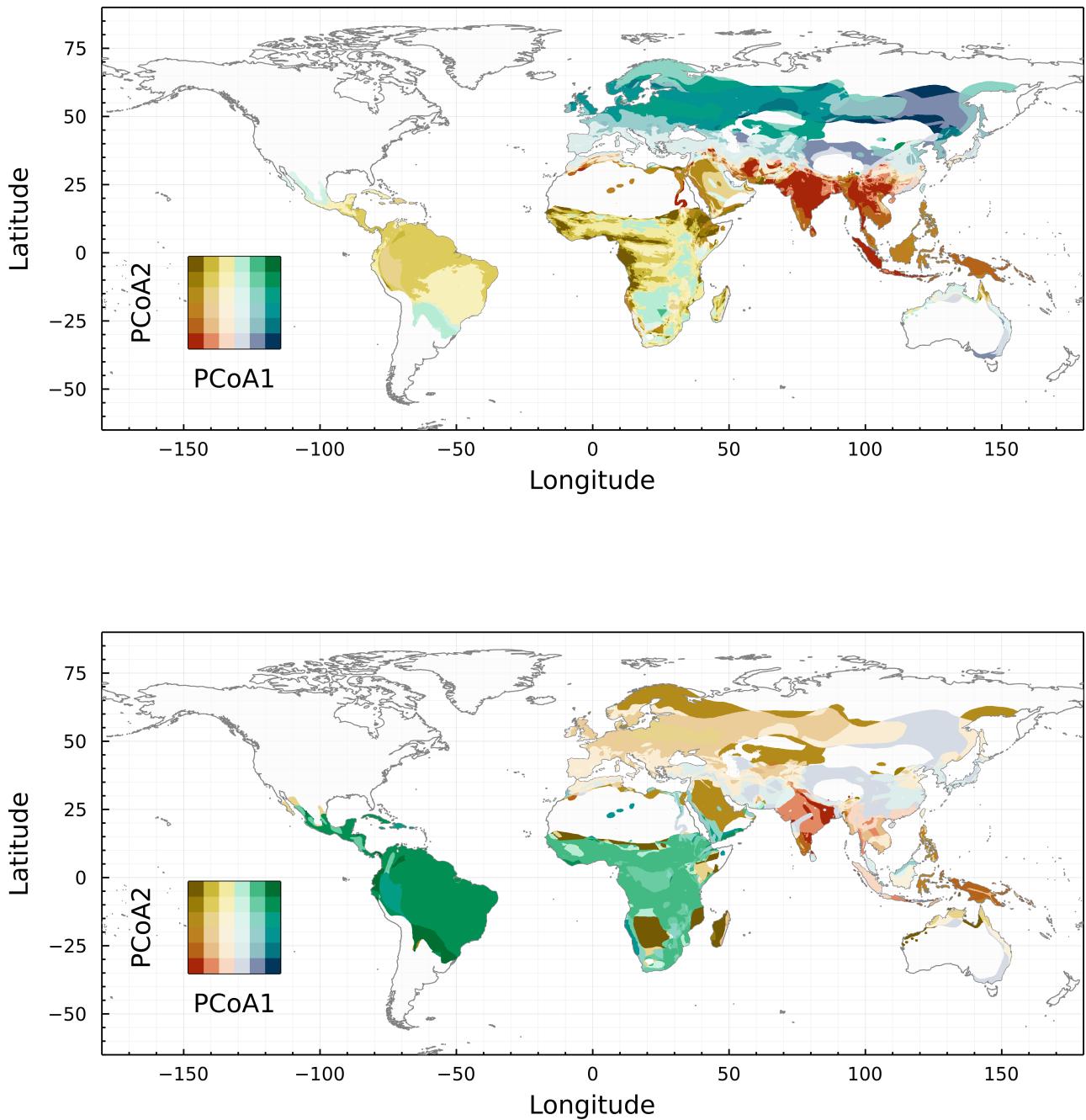


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

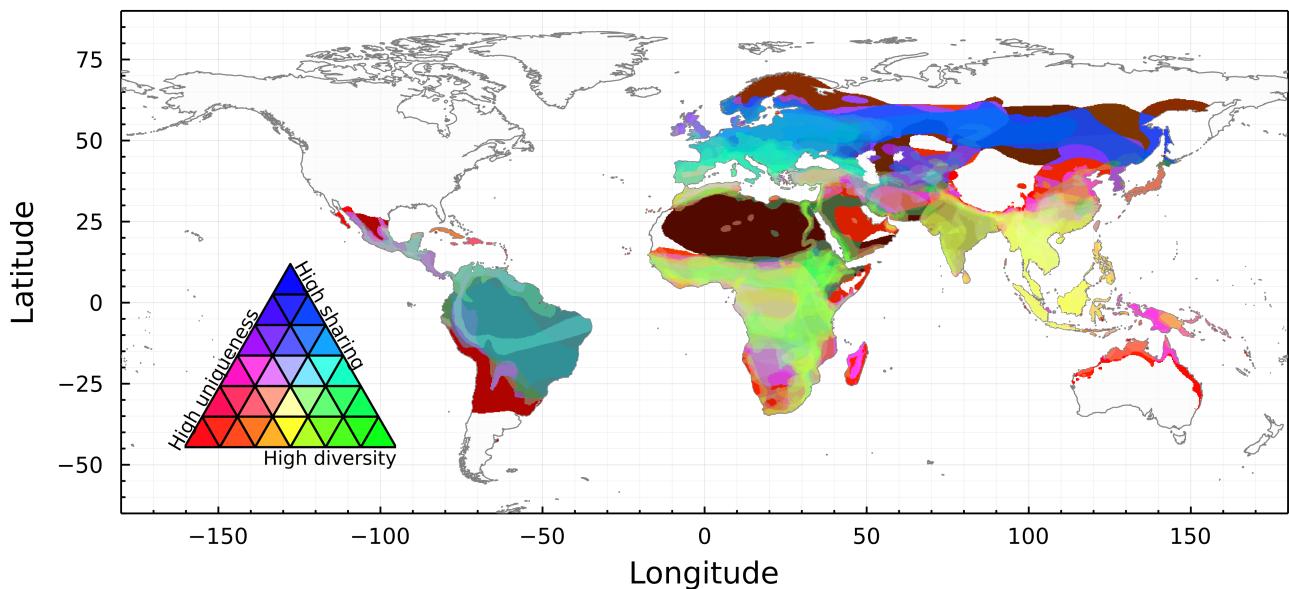


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible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

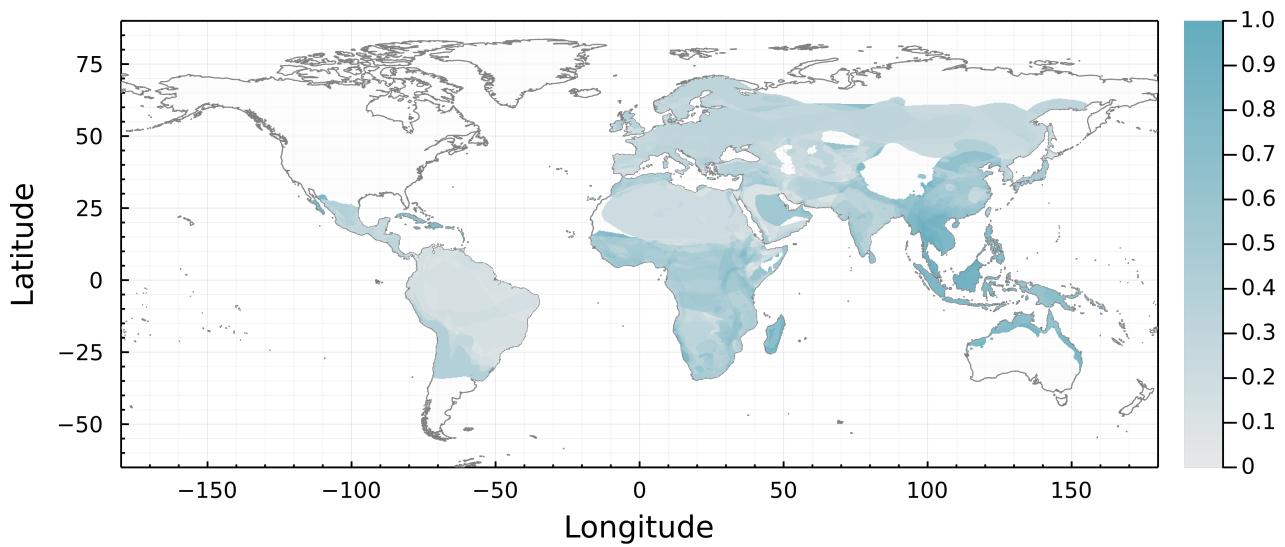


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

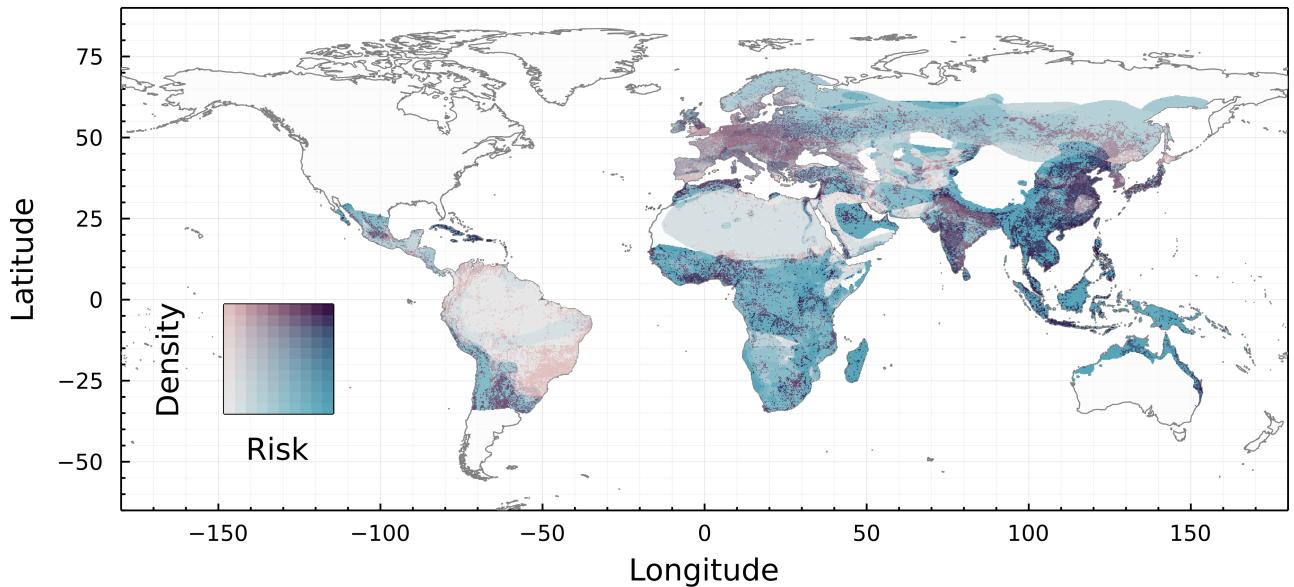


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk,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