

# 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 Spillover risk is complex and, even within a pool of susceptible wildlife hosts, driven by a multiplicity of  
2 factors (Plowright et al. 2017). Although host richness is often used as a coarse proxy for hotspots of  
3 emergence risk (Anthony et al. 2017, Ruiz-Aravena et al. 2022), these approaches deliberately  
4 oversimplify interspecific heterogeneity in immunology, behavior, and other traits. Global maps of  
5 spillover risk often struggle to distill these features into interpretable risk maps, and overlook  
6 highly-unique host pools that allow for the rapid evolution of highly divergent viruses (Agosta et al. 2010).  
7 In the case of generalist pathogens like betacoronaviruses, there is conceptual and empirical support to the  
8 idea that these community-level mechanisms are even more important in driving hotspots of overall risk  
9 (Power and Mitchell 2004).

10 These kinds of dynamics lend themselves to interpretation through the lens of the Geographic Mosaic  
11 Theory of Coevolution (GMTC; Thompson 2005), which connects microevolutionary dynamics in  
12 symbiotic interactions to macroecological dynamics in host communities. The GMTC posits that  
13 coevolutionary processes between pairs (Thompson 1994) or complexes (Janzen 1980) of species are  
14 structured in space by the rippling effects of abiotic conditions onto evolutionary mechanism, resulting in  
15 spatially fragmented evolutionary dynamics, coupled only by dispersal-related processes (Gomulkiewicz  
16 et al. 2000). In turn, these spatially fragmented processes can lead taxonomically homogeneous systems to  
17 have different structure and dynamics over large spatial extents (see e.g. Price 2002). The GMTC predicts a  
18 spatial fragmentation of coevolutionary dynamics under the joint action of three processes (see notably  
19 Gomulkiewicz et al. 2007), which all have the potential to act on outbreak potential, pathogen transmission,  
20 and disease virulence (Parratt et al. 2016, Turner et al. 2021). First hot- and coldspots of coevolution can  
21 appear when the intensity of *interaction* (in terms of reciprocal fitness consequences) varies spatially,  
22 because of e.g. partial range overlap between organisms (Nuismer et al. 2003). Second, the GMTC  
23 supposes the existence of selection mosaics, wherein the intensity of *selection* varies across space; the  
24 strength of reciprocal selection responds to the biotic complexity of the community (locally diverse hosts  
25 and viruses are more biotically complex; Thrall et al. 2007) and to the local favorability of the environment  
26 (Hochberg and Baalen 1998). Finally, trait remixing occurs when coevolutionary dynamics are driven by  
27 the arrival (or departure) of functional traits, through changes in community composition due to  
28 invasions, meta-community dynamics, and dispersal.

29 Each of these elements can be applied to a relatively quantifiable aspect of host-virus ecology: (i) viral  
30 sharing among hosts, representing the strength of potential interaction between viruses and any one host

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

38 Here,

39 We turn the processes on the GMTC into definitions of spillover risk from viruses to hosts (focusing on the  
40 bats-betacoronavirus complex), with a specific emphasis on the potential to create independent  
41 coevolutionary dynamics (and therefore spatial fragmentation in the risk) through heterogeneity.

## 42 Methods

### 43 Known *Betacoronavirus* hosts

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

### 51 Bats occurrences

52 We downloaded the rangemap of every current bat species that was classified as an empirically  
53 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data (IUCN 2021).  
54 The range maps were subsequently rasterized using the rasterize function from GDAL (Rouault et al.  
55 2022) at a resolution of approximately 100kmx100km. For every pixel in the resulting raster where at least  
56 one bat host of *Betacoronavirus* was present, we extract the species pool (list of all competent bat hosts),

57 which was used to calculate the following risk assessment components: bat phylogenetic diversity, bat  
58 compositional uniqueness, and predicted viral sharing risk.

## 59 **Bats phylogenetic diversity**

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

## 70 **Bats compositional uniqueness**

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

82 **Viral sharing between hosts**

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

89 **Composite risk map**

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

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

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

107 **Viral phylogeography and evolutionary diversification**

108 To next represent phylogeography of betacoronaviruses in bats, we aggregated and analyzed  
109 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data  
110 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR  
111 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR  
112 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated  
113 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained  
114 words indicating recombinant or laboratory strains including “patent,” “mutant,” “GFP,” and  
115 “recombinant.” We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East  
116 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus  
117 RdRp sequences were then aligned using MAFFT v 1.4.0 [Katoh and Standley (2013); Algorithm  
118 FFT-NS-2, Scoring matrix 200PAM / k=2, gap open penalty 1.53m offset value 0.123] and a maximum  
119 likelihood tree reconstructed in IQ-TREE v 1.6.12 (Nguyen et al. 2015) with ModelFinder  
120 (Kalyaanamoorthy et al. 2017) ultrafast bootstrap approximation (Hoang et al. 2018) with a general time  
121 reversible model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of  
122 nucleotide substitution (GTR+F+R5).

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

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

132 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the  
133 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,  
134 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their

135 original method, a distance matrix - where each row or column represents a geographic raster's grid cell,  
136 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes  
137 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected  
138 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel  
139 methodology. First, we measure the phylogenetic distance between the different viruses in the  
140 betacoronaviruses tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently,  
141 we take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this  
142 case) to project the viral tree into an n-dimensional space. We then take the first two principal  
143 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value  
144 and projected them using a four-color bivariate map.

## 145 Results and discussion

### 146 Host richness does not predict virus distinctiveness

147 Bats, the second most diverse groups of mammals, are found worldwide and serve as the main animal  
148 reservoir for different strains of betacoronaviruses (Drexler et al. 2014). This has attracted attention to  
149 areas where high diversity of bats, and therefore presumably high diversity of betacoronaviruses, can be an  
150 important issue for human health (Calisher et al. 2006, Moratelli and Calisher 2015). By overlaying the  
151 IUCN range maps for confirmed bat hosts of betacoronaviruses [fig. 1; top], we see that the the main  
152 hotspots (both in terms of size and higher values) of host richness are primarily South-Eastern Asia, parts  
153 of Southern Europe, and to a lesser extent parts of Africa in the -25-0 range of latitudes. The description of  
154 host richness is an important first step towards understanding risk, as previous research (Anthony et al.  
155 2017, Mollentze and Streicker 2020) states that locally diverse bat communities could maintain more  
156 viruses and hence, a higher probability of having a pathogen that could represent a risk for human health.

157 [Figure 1 about here.]

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

161 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness  
162 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as Yangochiroptera and  
163 Yinpterochiroptera elsewhere (Teeling et al. 2005, Springer 2013). Specifically, we would expect that the  
164 so-called “New World” group of bats, being more evolutionary distinct, would also have evolutionary  
165 distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a distinct  
166 bat-betacoronaviruses complex. This can be explained by the fact that Yangochiroptera, although not  
167 limited to the western hemisphere, contain the highly diverse adaptive radiation in the Phyllostomidae  
168 (Villalobos and Arita 2010), which is restricted to the western hemisphere. By contrast, South-Eastern  
169 Asia has a lot of non-evolutionary distinct bats, who nevertheless hosted evolutionary-distinct viruses.

170 It is noteworthy that outside of South America, viral evolutionary distinctiveness does not accurately track  
171 host diversity, with some areas having over-distinct viruses (eastern China but, oddly, not the rest of  
172 southeast Asia). There are a number of likely explanations. First, given the richness of bats in southeast  
173 Asia, many betacoronaviruses likely remain to be discovered in this region. Indeed, global predictions by  
174 Becker et al. (2022) highlight that southeast Asia is a likely hotspot of unconfirmed hosts of  
175 betacoronaviruses, which would likely result in additional viral discoveries. This idea is unsurprising  
176 given the growing realization, especially since the emergence of SARS-CoV-2, that unique lineages of  
177 similar viruses are widespread in bats but still mostly undescribed. The most distinct  
178 bats-betacoronaviruses complex is found in South America, a region with a comparatively lower number  
179 of hosts; this matches with the isolation through variance of the host group, and may highlight a different  
180 co-evolutionary dynamic. Alternatively, this distinctiveness hotspot may be a product of under-sampling:  
181 South-America is one of the places where the fewest *Betacoronavirus* sequences have been discovered  
182 (Anthony et al. 2017, Olival et al. 2017, Allen et al. 2017), resulting in sparser phylogenetic tree, thereby  
183 artificially inflating distinctiveness. Adding more viruses would bring the distinctiveness of known  
184 sequences down. Finally, South-America is the range of Phyllostomidae, a group of bats that underwent  
185 explosive diversification events (Villalobos and Arita 2010), which may drive the emergence of multiple  
186 viral lineages.

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

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

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

200 [Figure 2 about here.]

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

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

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

231 [Figure 3 about here.]

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

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

250 [Figure 4 about here.]

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

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

259 Based on the previous result, we extracted the risk component from the composite map (see Methods), to  
260 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. As this map  
261 represents the potential risk, it must be weighed by the potential for contacts with humans. As a proxy for  
262 this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a reasonable  
263 proxy for the density of humans per unit area, which increases the probability of pathogen spread more  
264 widely (Hazarie et al. 2021). Since human activity is required to amplify the frequency of virus encounters  
265 and thus create areas of viral amplification, mapping the potential risk against measures of land use is  
266 required to generate a more actionable assessment of risk. This map is presented in fig. 5. Most of South  
267 America and Europe are at comparatively lower risk, as although densely populated, settlements tend to  
268 be in areas with lower potential risk. Regions like Malaysia and the North coast of Australia have a high  
269 risk component, but should represent a relatively lower effective risk due to low human density. However,  
270 this mapping reveals that South-East Asia, the Indian subcontinent, and parts of sub-Saharan Africa, are  
271 at high risk due to the overlap between built areas and bat communities representing more opportunities  
272 for cross-species transmission of betacoronaviruses. In looking for the origins of SARS in China, Xu et al.  
273 (2004) present serological evidence that strongest human-animal contact results in higher risk of virus  
274 exposure, regardless of the animal species, but that different types of contact had different impacts. Ideally,  
275 finer-grained information about human activity (rather than human presence through anthropisation)

276 could allow to partition this risk further, albeit at the cost of more hypotheses required to estimate the  
277 amount of risk represented by each activity. Our map of risk overlays with recent results from Cohen et al.  
278 (2022) – areas of purported high risk/diversification potential (Madagascar, South-America) overlay with  
279 sampling gaps for *Betacoronavirus*, stressing the need for spatially targeted monitoring and discovery.

280 [Figure 5 about here.]

## 281 Conclusion

282 RELOCATE: Indeed, it only assumes the action of well described evolutionary mechanisms. The benefit  
283 of this approach is to provide the potential for a more dynamic and nuanced understanding of risk: not  
284 only on ecological timescales, but also by providing clues about which areas can change over  
285 micro-evolutionary timescales.

286 Our study focuses largely on the biogeography of hosts. Specifically, we identify the Amazon and  
287 South-Eastern Asia as hotspots where the phylogenetic distinctiveness of *Betacoronavirus* is the highest  
288 (Anthony et al. 2017); surprisingly, current data suggest that viral sharing between hosts is high in the  
289 Amazon and low in South-Eastern Asia, which has the potential to result in different evolutionary  
290 dynamics between these two regions, hinting at different futures for their viral communities. This work is  
291 important both as a description of the bats-betacoronaviruses complex, but also because more broadly,  
292 bats are known reservoirs for a variety of emerging viruses and other pathogens (Calisher et al. 2006,  
293 Melaun et al. 2014), making balancing the needs for bat conservation and disease prevention most likely  
294 very difficult and a source of human-wildlife conflicts, especially in more densely populated areas (Stone  
295 et al. 2015, Rego et al. 2015).

296 Yet, we know that viruses with high host plasticity, that is, the ability of a given virus to adapt to various  
297 taxonomic orders and ecological groups (Kreuder Johnson et al. 2015), are more likely to amplify viral  
298 spillover, followed by secondary human-to-human transmission, and geographical spread (Hazarie et al.  
299 2021). High viral host plasticity is an especially important trait for RNA viruses like betacoronaviruses  
300 (Haddad et al. 2021). Indeed, our analysis of viral sequences reveals that Latin America is a hotspot of  
301 viral distinctiveness, suggesting that this part of the bats-betacoronaviruses complex may be undergoing  
302 independent evolutionary dynamics (related species sharing viruses that are different from the rest of the

303 global pool). The other hotspot of viral distinctiveness is S.E. Asia, in which richness is high but sharing is  
304 low; this suggests a different type of evolutionary dynamics (unrelated viruses coevolving with  
305 evolutionarily distinct hosts, generating high diversity locally, see e.g. Latinne et al. 2020). Both of these  
306 areas should be priority areas for sampling, especially since Becker et al. (2022) advance that they harbor  
307 undiscovered hosts of beta-coronaviruses. This diversity of hosts, and the mechanisms by which the  
308 exchange of viruses occurs between species, is largely affected by the local environmental conditions and  
309 environmental change.

310 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to  
311 human health (Letko et al. 2020, Van Brussel and Holmes 2022). Chiropterans emerged around 64 million  
312 years ago and are one of the most diverse mammalian orders, with an estimated richness of more than  
313 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020). They exhibit a broad variety of habitat  
314 use, behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of  
315 several ecosystem services, tied to important ecosystem-derived benefits to human (Kasso and  
316 Balakrishnan 2013). For example, bats are an essential component of many seed-dispersal networks  
317 (Mello et al. 2011). Over two-thirds of bats are know to be either obligate or facultative insectivores,  
318 therefore actively contributing for agricultural pest control (Williams-Guillén et al. 2008, Voigt and  
319 Kingston 2016), and vectors of pathogens that put a risk on human health (Gonsalves et al. 2013a, b).  
320 Because bats are globally distributed and have a long evolutionary history, phylogeographic and  
321 biogeographic approaches are required to shed light on the contemporary distribution of coevolutionary  
322 processes between bats and the pathogens they host. Not all areas in which bats, viruses, and human are  
323 co-occurring are facing a risk of spillover towards human populations, and the areas in which this risk exist  
324 may not be facing risks of the same nature and magnitude.

325 There are several factors that drive changes in the diversity of bats (Alves et al. 2018), but human  
326 activities' effects on the ecosystem (like modifications of land use) could significantly decrease it.  
327 Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their  
328 biogeographic variation, and human population density and other anthropogenic factors are decisive  
329 moderators for its implications in public health. With the increase of contact between humans and  
330 potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al. 2020), as previous  
331 studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal  
332 (Gryseels et al. 2017). One of these scenarios where interaction between bats and humans can occur can

333 be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats  
334 not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and  
335 humans (Deshpande et al. 2022). This represents a challenge for conservation strategies and disease  
336 ecology since some areas can have both potential for the acquisition of zoonotic viruses and bat-human  
337 interactions; in particular, the challenge lies in the fact that actual exposure must then be quantified  
338 accounting for several transmission scenarios, including both direct and indirect bat - human interaction.

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

## 347 References

- 348 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and  
349 implications for emerging infectious disease. - *Zoologia* (Curitiba) 27: 151–162.
- 350 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -  
351 *Nature Communications* 11: 2260.
- 352 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - *Nature*  
353 *Communications* in press.
- 354 Alves, D. M. C. C. et al. 2018. Geographic variation in the relationship between large-scale environmental  
355 determinants and bat species richness. - *Basic and Applied Ecology* 27: 1–8.
- 356 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - *Virus Evolution* in press.
- 357 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - *Frontiers in Immunology* 11: 26.
- 358 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -  
359 *The Lancet Microbe* in press.

- 360 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - Clinical Microbiology  
361 Reviews 19: 531–545.
- 362 Carlson, C. J. et al. 2022. Climate change increases cross-species viral transmission risk. - Nature: 1–1.
- 363 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - Ecol. Lett.  
364 12: 693–715.
- 365 Cohen, L. E. et al. 2022. Sampling strategies and pre-pandemic surveillance gaps for bat coronaviruses.:  
366 2022.06.15.496296.
- 367 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species  
368 distribution modelling. - Oikos n/a: e09063.
- 369 Deshpande, K. et al. 2022. Forbidden fruits? Ecosystem services from seed dispersal by fruit bats in the  
370 context of latent zoonotic risk. - Oikos (Copenhagen, Denmark): oik.08359.
- 371 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of  
372 SARS. - Antiviral Research 101: 45–56.
- 373 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1–10.
- 374 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -  
375 The American Naturalist 156: 156–174.
- 376 Gomulkiewicz, R. et al. 2007. Dos and don'ts of testing the geographic mosaic theory of coevolution. -  
377 Heredity 98: 249–258.
- 378 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:  
379 10–24.
- 380 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE  
381 8: e77183.
- 382 Gorbunova, V. et al. 2020. The World Goes Bats: Living Longer and Tolerating Viruses. - Cell Metabolism  
383 32: 31–43.
- 384 Gryseels, S. et al. 2017. When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure  
385 in the Spatial Spread of Arenaviruses (JH Kuhn, Ed.). - PLOS Pathogens 13: e1006073.

- 386 Haddad, D. et al. 2021. SARS-CoV-2: Possible recombination and emergence of potentially more virulent  
387 strains (H Attoui, Ed.). - PLOS ONE 16: e0251368.
- 388 Hazarie, S. et al. 2021. Interplay between population density and mobility in determining the spread of  
389 epidemics in cities. - Communications Physics 4: 191.
- 390 Hoang, D. T. et al. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. - Molecular Biology  
391 and Evolution 35: 518–522.
- 392 Hochberg, M. E. and Baalen, M. 1998. Antagonistic coevolution over productivity gradients. - The  
393 American Naturalist 152: 620–634.
- 394 Irving, A. T. et al. 2021. Lessons from the host defences of bats, a unique viral reservoir. - Nature 589:  
395 363–370.
- 396 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.  
397 - PLOS ONE 2: e296.
- 398 IUCN 2021. The IUCN Red List of Threatened Species.
- 399 Janzen, D. H. 1980. When is it Coevolution? - Evolution 34: 611–612.
- 400 Jebb, D. et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. - Nature 583:  
401 578–584.
- 402 Johnson, C. K. et al. 2020. Global shifts in mammalian population trends reveal key predictors of virus  
403 spillover risk. - Proceedings of the Royal Society B: Biological Sciences 287: 20192736.
- 404 Kalyaanamoorthy, S. et al. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. -  
405 Nature Methods 14: 587–589.
- 406 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -  
407 ISRN Biodiversity 2013: e187415.
- 408 Katoh, K. and Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7:  
409 Improvements in Performance and Usability. - Molecular Biology and Evolution 30: 772–780.
- 410 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010  
411 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.

- 412 Kettenburg, G. et al. 2022. Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a  
413 Unique Evolutionary History for This Coronavirus Clade. - *Frontiers in Public Health* in press.
- 414 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - *Proceedings of*  
415 *the National Academy of Sciences* 104: 5925–5930.
- 416 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species  
417 distributions. - *Journal of Biogeography* 37: 2029–2053.
- 418 Kreuder Johnson, C. et al. 2015. Spillover and pandemic properties of zoonotic viruses with high host  
419 plasticity. - *Scientific Reports* 5: 14830.
- 420 Latinne, A. et al. 2020. Origin and cross-species transmission of bat coronaviruses in China. - *bioRxiv*:  
421 The Preprint Server for Biology: 2020.05.31.116061.
- 422 Lednicky, J. A. et al. 2021. Isolation of a Novel Recombinant Canine Coronavirus from a Visitor to Haiti:  
423 Further Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. - *Clinical*  
424 *Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*: ciab924.
- 425 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity  
426 coefficients and partitioning (H Morlon, Ed.). - *Ecology Letters* 16: 951–963.
- 427 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado  
428 Island forest dynamics plot, Panama. - *Forest Ecosystems* 6: 7.
- 429 Letko, M. et al. 2020. Bat-borne virus diversity, spillover and emergence. - *Nature Reviews Microbiology*  
430 18: 461–471.
- 431 Melaun, C. et al. 2014. Bats as Potential Reservoir Hosts for Vector-Borne Diseases. - In: Kliment, S. and  
432 Mehlhorn, H. (eds), *Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths*.  
433 *Parasitology Research Monographs*. Springer, pp. 25–61.
- 434 Mello, M. A. R. et al. 2011. The Missing Part of Seed Dispersal Networks: Structure and Robustness of  
435 Bat-Fruit Interactions. - *PLOS ONE* 6: e17395.
- 436 Mollentze, N. and Streicker, D. G. 2020. Viral zoonotic risk is homogenous among taxonomic orders of  
437 mammalian and avian reservoir hosts. - *Proceedings of the National Academy of Sciences* 117:  
438 9423–9430.

- 439 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with  
440 emerging deadly viruses? - *Memórias do Instituto Oswaldo Cruz* 110: 1–22.
- 441 Moreno Santillán, D. D. et al. 2021. Large-scale genome sampling reveals unique immunity and metabolic  
442 adaptations in bats. - *Molecular Ecology*: mec.16027.
- 443 Muylaert, R. L. et al. 2022. Present and future distribution of bat hosts of sarbecoviruses: Implications for  
444 conservation and public health. - *Proceedings of the Royal Society B: Biological Sciences* 289:  
445 20220397.
- 446 Nguyen, L.-T. et al. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating  
447 Maximum-Likelihood Phylogenies. - *Molecular Biology and Evolution* 32: 268–274.
- 448 Nuismer, S. L. et al. 2003. Coevolution between hosts and parasites with partially overlapping geographic  
449 ranges. - *Journal of Evolutionary Biology* 16: 1337–1345.
- 450 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - *Nature* 546:  
451 646–650.
- 452 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary  
453 analyses in R. - *Bioinformatics* 35: 526–528.
- 454 Parratt, S. R. et al. 2016. Infectious Disease Dynamics in Heterogeneous Landscapes. - *Annual Review of  
455 Ecology, Evolution, and Systematics* 47: 283–306.
- 456 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across  
457 spatial scales. - *BMC ecology* 18: 18.
- 458 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - *Nature Reviews Microbiology* 15: 502–510.
- 459 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - *The American  
460 Naturalist* 164: S79–S89.
- 461 Price, P. W. 2002. Macroevolutionary Theory on Macroecological Patterns. - Cambridge University Press.
- 462 Ramshaw, R. E. et al. 2019. A database of geopositioned Middle East Respiratory Syndrome Coronavirus  
463 occurrences. - *Scientific Data* 6: 318.
- 464 Rego, K. M. da C. et al. 2015. Assessing human-bat interactions around a protected area in northeastern  
465 Brazil. - *Journal of Ethnobiology and Ethnomedicine* 11: 80.

- 466 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 467 Ruiz-Aravena, M. et al. 2022. Ecology, evolution and spillover of coronaviruses from bats. - Nature  
468 Reviews Microbiology 20: 299–314.
- 469 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - Limnology and Oceanography Letters 3:  
470 49–56.
- 471 Shi, J. J. et al. 2014. A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with  
472 Evidence for Widespread Panmixia. - Acta Chiropterologica 16: 279–292.
- 473 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic  
474 database.
- 475 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - Current Biology 23: R999–R1001.
- 476 Stone, E. et al. 2015. Managing Conflict between Bats and Humans: The Response of Soprano Pipistrelles  
477 (*Pipistrellus pygmaeus*) to Exclusion from Roosts in Houses. - PLoS ONE 10: e0131825.
- 478 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.  
479 - Science (New York, N.Y.) 307: 580–584.
- 480 Thompson, J. N. 1994. The Coevolutionary Process. - University of Chicago Press.
- 481 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.
- 482 Thrall, P. H. et al. 2007. Coevolution of symbiotic mutualists and parasites in a community context. -  
483 Trends in Ecology & Evolution 22: 120–126.
- 484 Turner, W. C. et al. 2021. The roles of environmental variation and parasite survival in  
485 virulence–transmission relationships. - Royal Society open science 8: 210088.
- 486 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in  
487 ecology, evolution, and conservation. - PLOS Biology 17: e3000494.
- 488 Van Brussel, K. and Holmes, E. C. 2022. Zoonotic disease and virome diversity in bats. - Current Opinion  
489 in Virology 52: 192–202.
- 490 Villalobos, F. and Arita, H. T. 2010. The diversity field of New World leaf-nosed bats (Phyllostomidae). -  
491 Global Ecology and Biogeography 19: 200–211.

- 492 Vlasova, A. N. et al. 2022. Animal alphacoronaviruses found in human patients with acute respiratory  
493 illness in different countries. - Emerging Microbes & Infections 11: 699–702.
- 494 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,  
495 Eds.). - Springer International Publishing.
- 496 Wang, N. et al. 2018. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.  
497 - Virologica Sinica 33: 104–107.
- 498 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - Science 320:  
499 70–70.
- 500 Xu, R.-H. et al. 2004. Epidemiologic Clues to SARS Origin in China. - Emerging Infectious Diseases 10:  
501 1030–1037.

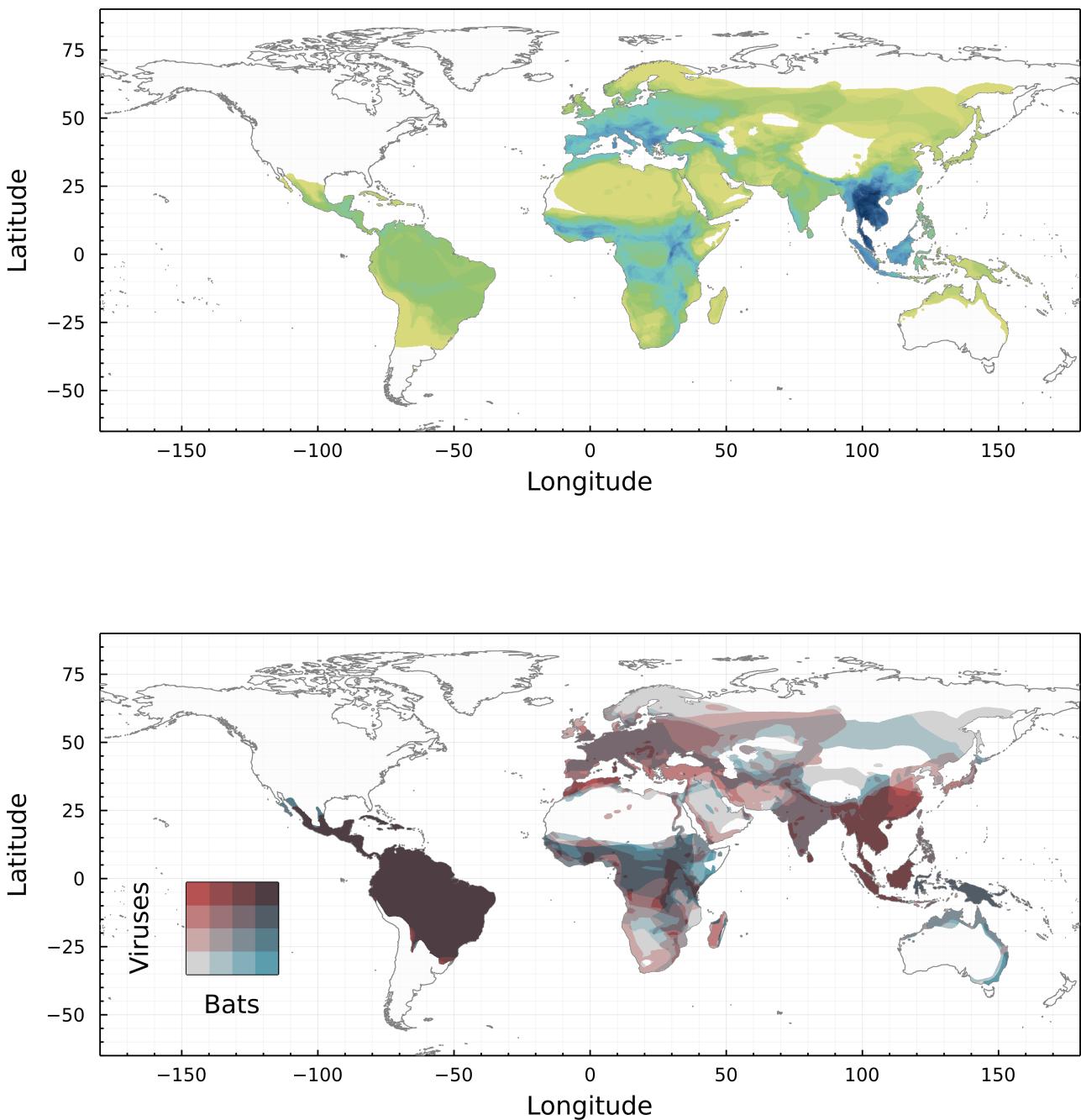


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).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

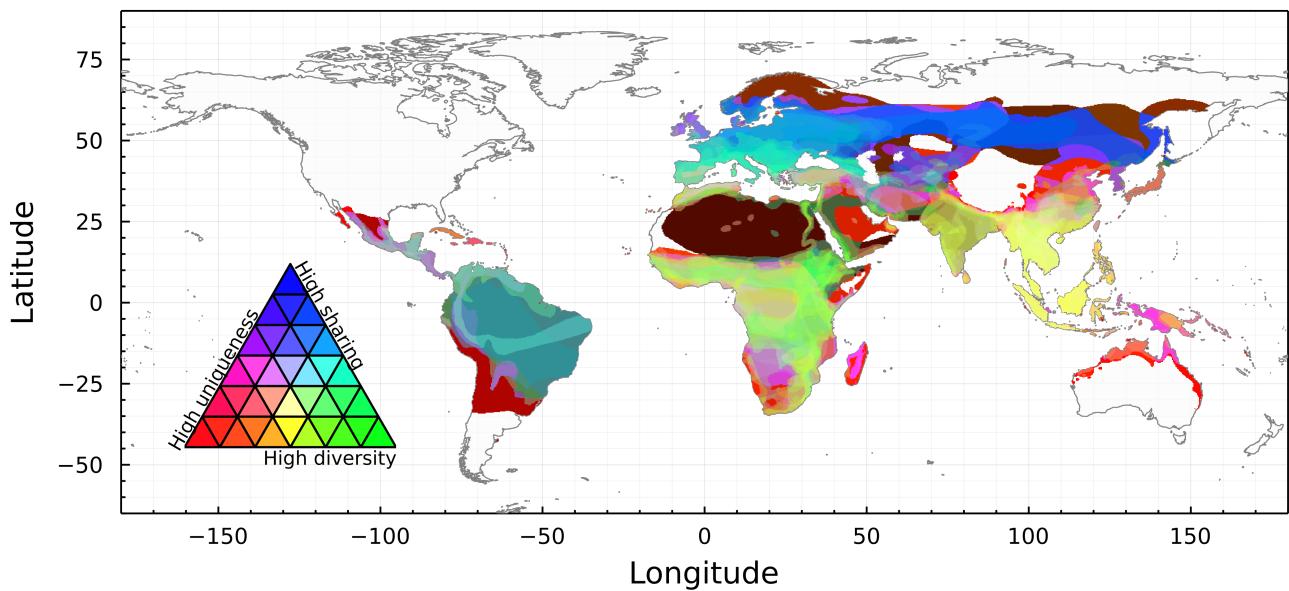


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible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

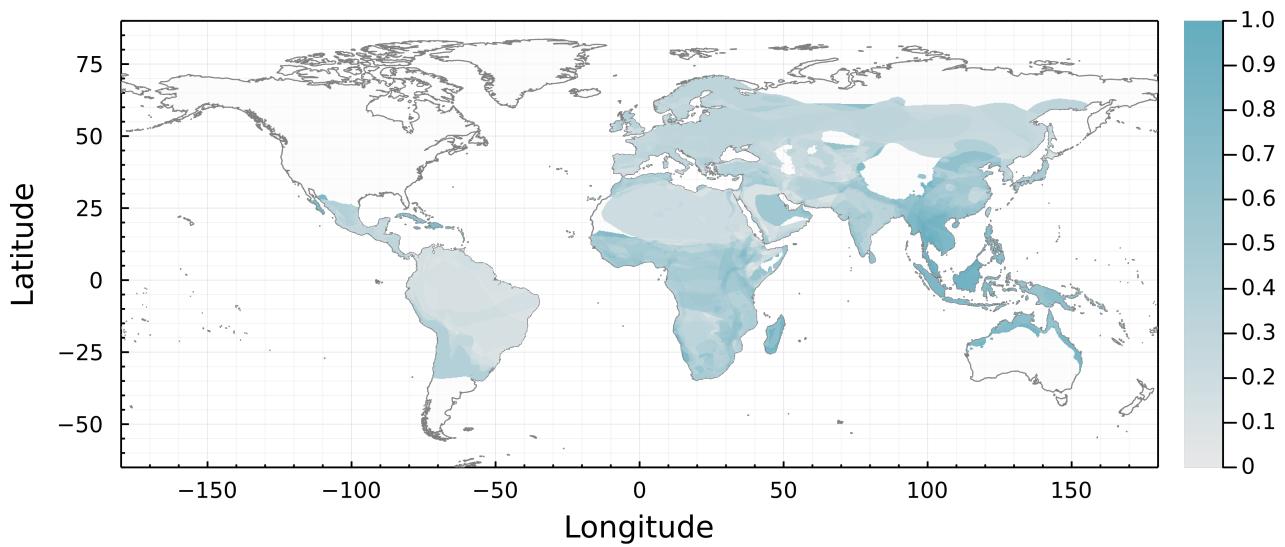


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

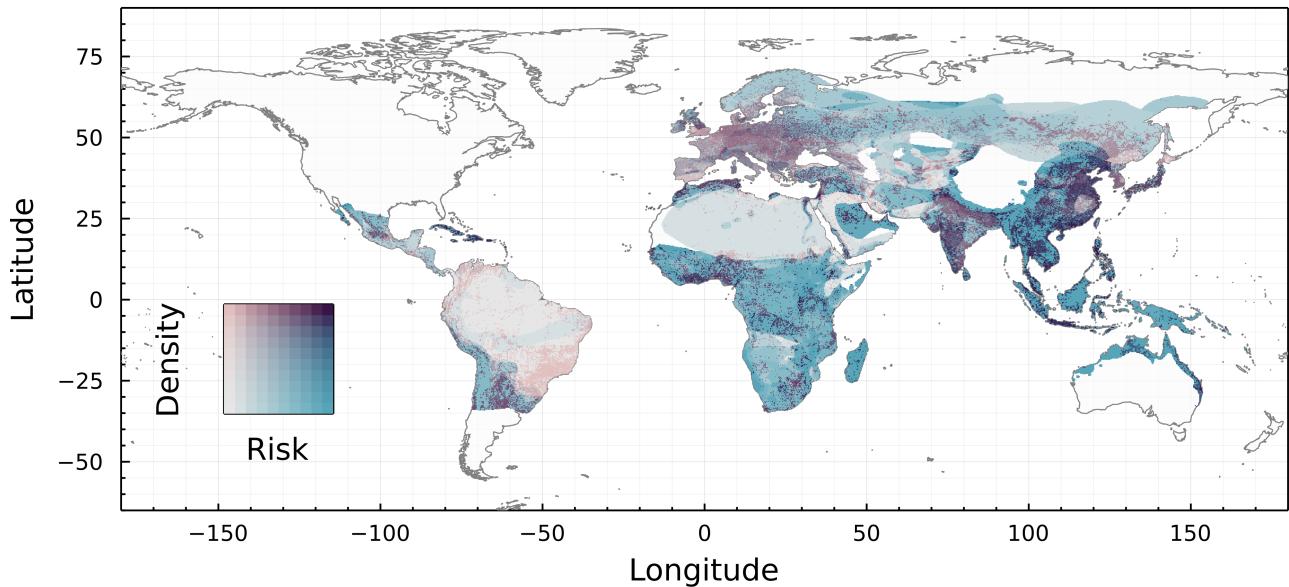


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk,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