

The coevolutionary mosaic of bat betacoronavirus emergence risk

Norma Forero Rocio Munoz^{1,2,‡} Renata L. Muylaert³ Stephanie N. Seifert⁴ Gregory F. Albery⁵

Daniel J. Becker⁶ Colin J. Carlson^{7,8,9,‡} [Timothée Poisot](#)^{1,2,‡}

¹ Université de Montréal ² Québec Centre for Biodiversity Sciences ³ Molecular Epidemiology and Public Health Laboratory, School of Veterinary Science, Massey University, New Zealand ⁴ Paul G. Allen School for Global Health, Washington State University, Pullman, WA, United States ⁵ Department of Biology, Georgetown University, Washington, DC, USA ⁶ Department of Biology, University of Oklahoma, Norman, OK, USA ⁷ Department of Biology, Georgetown University, Washington, DC,

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

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Driven by the need to understand the ecological factors involved in the emergence of *Betacoronavirus* (the genus causing the SARS, MERS, and COVID-19 in human) through bat hosts, we develop an approach to the assessment of spillover risk based on the Geographic Mosaic Theory of Coevolution. In doing so, we provide a global mapping of the spillover risk posed by betacoronaviruses, reflecting the fact that this risk is best understood through the multi-faceted prism of ecological and evolutionary mechanisms. Our framework reveals that host richness alone, although a component of viral hazard, is not a sufficiently integrative predictor of risk. We offer alternative insights based on viral sharing, host compositional uniqueness, and host phylogenetic diversity and phylogeographic regions. By comparing our aggregated measure of risk to a proxy for human density, namely the proportion of each pixel that is covered by urban or built land, we provide a synthetic risk map, allowing the identification of hotspots where the bat-betacoronavirus interaction network may facilitate the emergence of novel viruses and their spillover into human populations.

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

10 The Geographic Mosaic Theory of Coevolution (GMTC) attempts to explicitly connect microevolutionary
11 dynamics to the macroecology and biogeography of symbiotic interactions.⁷ The GMTC posits that
12 coevolutionary processes among pairs⁸ or complexes⁹ of species are structured in space by the rippling
13 effects of abiotic conditions onto evolutionary mechanism, giving rise to fragmented systems with
14 different structure and ecologically dynamics over large spatial extents.¹⁰ The GMTC predicts a spatial
15 fragmentation of coevolutionary dynamics under the joint action of three processes:¹¹ coevolutionary hot-
16 and coldspots, which appear when the intensity of *interaction* (in terms of reciprocal fitness consequences)
17 varies spatially; selection mosaics, wherein the intensity of *selection* varies across space, driven by both the
18 biotic complexity of the community (locally diverse hosts and viruses are more biotically complex) and the
19 local favorability of the environment;¹² and trait remixing, which occurs when coevolutionary dynamics
20 are driven by the arrival (or departure) of *functional traits*, through changes in community composition
21 due to invasions, meta-community dynamics, and dispersal.

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

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

40 Results and Discussion

41 Bat and betacoronavirus biogeography are broadly consistent

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

53 [Figure 1 about here.]

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

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

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

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

78 [Figure 2 about here.]

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

85 phylogenetic tree, and artificially inflating distinctiveness; conversely, disproportionate research effort in
86 eastern China¹⁷ may have led to a more complete inventory of the local diversity of coronaviruses, again
87 inflating these metrics relative to underlying patterns. Even accounting for these potential biases, though,
88 there is obvious heterogeneity in betacoronavirus evolutionary distinctiveness that is distinct from overall
89 bat diversity.

90 Overall, these patterns recapitulate the evolutionary history of both the order Chiroptera and the genus
91 *Betacoronavirus*. Horseshoe bats (Rhinolophidae) are both the reservoirs of the SARS-like viruses
92 (subgenus *Sarbecovirus*), the group of pandemic threats that have been of the greatest interest to
93 researchers¹³ (and so have been sampled most intensively).¹⁷ The hotspots of host richness and viral
94 diversity in southeast Asia—both of which are disproportionately high, considering the global landscape
95 of bat species richness—are almost entirely driven by viral adaptive radiation through host switching
96 within this clade^{3,15}. In contrast, the Neotropical hotspot of viral distinctiveness is driven by isolation by
97 host vicariance. Out of the four main groups of betacoronaviruses, only the subgenus *Merbecovirus*
98 (MERS-like viruses) has been found in animals in the Americas—an introduction that is generally
99 presumed to be ancient.³ While comparatively understudied, New World merbecoviruses have been found
100 in the ghost-faced bats (Mormoopidae), New World leaf-nosed bats (Phyllostomidae), and free-tailed bats
101 (Molossidae) (add cite: Olival 2020 PLoS Pathogens). The former two groups are endemic to the
102 Neotropics, while the explosive adaptive radiations of the latter two (and particularly the phyllostomids)
103 are responsible for the hotspot of bat diversity in the Amazon. Together, these clades of New World bats
104 play host to a distinct regime of betacoronavirus coevolution.

105 Coevolution-informed emergence risk is different in space

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

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

117 [Figure 3 about here.]

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

136 [Figure 4 about here.]

137 From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide
138 hotspots of *Betacoronavirus* risk through mixing of unique viruses (via codivergence) and in turn
139 recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat
140 species are endemic following evolutionary divergence from sister species in both African and Asian
141 continents.¹⁸ Recent surveillance¹⁹ has identified a novel *Betacoronavirus* (in the subgenus *Nobecovirus*) in

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

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

145 The relationship between the underlying pathogen pool and emergence risk is mediated by both
146 human-wildlife interfaces (the probability of spillover) and opportunities for onward transmission (the
147 probability that spillovers become epidemics). As a proxy for both, we finally overlaid the risk component
148 from the composite map (see above) with the proportion of built land, as a proxy for a mix of habitat
149 disturbance, potential for bat synanthropy or contact with bridge hosts like livestock (cite Rulli et al 2021
150 Nature Food; cite Cui et al 2019 Nature Reviews Microbiology), and human population density and
151 connectivity^{1,20} (add: Muylaert 2022 Proc B; Hassell 2017 TREE) (fig. 5). Accounting for these factors,
152 most of South America and Europe are at comparatively lower risk, as—although densely
153 populated—settlements tend to be in areas with lower potential risk. Conversely, regions like Malaysia and
154 the northern coast of Australia have a high evolutionary risk component, but should represent a relatively
155 lower effective risk due to low human density. However, southeast Asia, the Indian subcontinent, and
156 scattered hotspots in sub-Saharan Africa are at high risk due to the overlap between human populations
157 and natural opportunities for cross-species transmission of betacoronaviruses.

158 [Figure 5 about here.]

159 Reassuringly, these predictions are validated by the sites where the three highly pathogenic coronavirus
160 are believed to have recently emerged. While available information puts the spillover of SARS-CoV-2 in a
161 live animal market in Wuhan, China, the ultimate origin of the virus is almost certainly in a divergent
162 lineage of sarbecoviruses from the Indochinese peninsula that was poorly characterized prior to the
163 pandemic [cite Worobey 2022 Zenodo; Temmam 2022 Nature; Boni 2020 Nature Microbiology]. Similarly,
164 the SARS-CoV outbreak began in Guangdong province in 2002, reaching humans through small carnivore
165 bridge hosts, but was eventually traced back to a set of likely progenitor viruses found in cave-dwelling
166 horseshoe bats in Yunnan province (cite Hu et al. 2017 PLoS Pathogens); nearby, antibody evidence has
167 indicated human exposure to SARS-like viruses.²¹ MERS-CoV was originally detected in Saudi Arabia,
168 accompanied by a nearly identical virus sequenced from an Egyptian tomb bat (*Taphozous perforatus*)
169 (cite Memish et al. 2013 EID), but is widespread in camels in East Africa and the Middle East, and may

170 have reached its bridge host decades earlier than originally supposed (cite Muller et al 2014 EID); as a
171 result, the geography of the original bat-to-camel transmission is still widely regarded as uncertain. All of
172 these are broadly consistent with the risk factors we identify. Notably, India and west Africa are additional
173 hotspots that have yet to experience the emergence of a bat coronavirus into human populations, but may
174 still be at risk—particularly given known gaps in bat surveillance,¹⁷ and a dense population in both
175 regions with global connectivity. In any of these regions, surveillance on viral reservoirs can be paired
176 with targeted monitoring of high-risk human populations (i.e., those with regular wildlife contact)²² for
177 maximum impact.

178 Conclusion

179 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to
180 human health.^{23,24} Chiropterans emerged around 64 million years ago and are one of the most diverse
181 mammalian orders, with an estimated richness of more than 1400 species.^{25,26} They exhibit a broad variety
182 of habitat use, behaviour, and feeding strategies, putting them at key positions in the delivery and
183 provisioning of several ecosystem services, tied to important ecosystem-derived benefits to human.²⁷ For
184 example, bats are an essential component of many seed-dispersal networks.²⁸ Over two-thirds of bats are
185 know to be either obligate or facultative insectivores, therefore actively contributing for agricultural pest
186 control,^{29,30} and vectors of pathogens that put a risk on human health.^{31,32} Because bats are globally
187 distributed and have a long evolutionary history, phylogeographic and biogeographic approaches are
188 required to shed light on the contemporary distribution of coevolutionary processes between bats and the
189 pathogens they host. Not all areas in which bats, viruses, and human are co-occurring are facing a risk of
190 spillover towards human populations, and the areas in which this risk exist may not be facing risks of the
191 same nature and magnitude.

192 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries
193 like highly divergent nobecoviruses in Madagascar and the previously-neglected adaptive radiation of
194 sarbecoviruses outside of southern China and throughout southeast Asia. In doing so, it advances
195 ecological theory beyond the current state of the art for global maps of emergence risk. For example,
196 previous studies that have used host richness as proxy have predicted a high diversity of unsampled bat
197 viruses,¹⁴ bat coronaviruses,² and even specifically betacoronaviruses¹⁵ in both the Amazon and southeast

198 Asia. While we find that both regions are characterized by highly divergent host and viral communities,
199 our framework identifies key differences between the regions. We find that Latin America is a hotspot of
200 both host and viral distinctiveness, suggesting that this branch of the bat-betacoronavirus complex may be
201 undergoing independent evolutionary dynamics from the rest of the global pool, but with limited potential
202 for viral diversification—a finding that is supported by previous work indicating a higher rate of
203 codivergence in Latin America.² In contrast, in southeast Asia, host richness and viral distinctiveness are
204 high but sharing is low; this suggests a different type of evolutionary dynamics that could generate high
205 local diversity of viruses through host switching and viral recombination (see e.g.,¹³ as well as the
206 discovery of recombinant viruses that share genetic material from both the SARS-CoV and SARS-CoV-2
207 branches of the Sarbecovirus lineage).³³ Both of these regions are priority areas for sampling, especially
208 given predictions that they contain many bat hosts of undiscovered betacoronaviruses.^{15,17} However, both
209 the evolutionary and ecological aspects of emergence risk are likely higher in southeast Asia—a fact that
210 will only become more relevant, as bats track shifting climates and exchange viruses with other species,
211 creating a hotspot of cross-species transmission unique to the region.³⁴

212 The diversity and diversification potential of bats responds to anthropogenic factors others than shifting
213 climates.³⁵ Land use changes could significantly decrease bat suitability, notably through effects on diet
214 and availability of habitats.³⁶ As our results establish that the diversification of bats betacoronaviruses
215 happens on top of processes affecting hosts, biogeographic variation in human population density and
216 anthropogenic disturbances may feed into co-evolutionary dynamics. Increase in humans-hosts contacts
217 also increase the risk of emergence of novel diseases,³⁷ so does the changes in landscape connectivity at
218 local/regional scales.³⁸ This represents a challenge for both conservation strategies and disease ecology:
219 some areas can a high emergence risk and more potential for the acquisition of zoonotic viruses through
220 bat-human encounters.³⁹ In particular, the challenge ahead lies in the need to quantify actual exposure
221 (and risk) accounting for several transmission scenarios, including both direct and indirect bat - human
222 interactions, and feeding back into the provision of ecosystem services by bats.

223 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
224 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and
225 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research
226 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des
227 Données (IVADO). This research was enabled in part by support provided by Calcul Québec

228 (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). NF is funded by the NSERC
229 BIOS² CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by
230 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation. DJB was
231 supported by the National Institute of General Medical Sciences of the National Institutes of Health
232 (P20GM134973).

233 **Methods**

234 **Known *Betacoronavirus* hosts**

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

242 **Bat occurrences**

243 We downloaded the rangemap of every current bat species that was classified as an empirically
244 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data.⁴⁰ The range
245 maps were subsequently rasterized using the rasterize function from GDAL⁴¹ at a resolution of
246 approximately 100kmx100km. For every pixel in the resulting raster where at least one bat host of
247 *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was used to
248 calculate the following risk assessment components: bat phylogenetic diversity, bat compositional
249 uniqueness, and predicted viral sharing risk.

250 **Bat phylogenetic diversity**

251 For every pixel, we measured Faith’s Phylogenetic Diversity⁴² based on a recent synthetic tree with robust
252 time calibration, covering about 6000 mammalian species.⁴³ Faith’s PD measures the sum of unique
253 branches from an arbitrary root to a set of tips, and comparatively larger values indicate a more
254 phylogenetic diverse species pool. We measured phylogenetic diversity starting from the root of the entire
255 tree (and not from Chiroptera); this bears no consequences on the resulting values, since all branches
256 leading up to Chiroptera are only counted one per species pool, and (as we explain when describing the
257 assembly of the composite risk map), all individual risk components are ranged in [0,1]. This measure
258 incorporates a richness component, which we chose not to correct for; the interpretation of the

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

261 **Bat compositional uniqueness**

262 For every species pool, we measured its Local Contribution to Beta-Diversity;⁴⁴ LCBD works from a
263 species-data matrix (traditionally noted as \mathbf{Y}), where species are rows and sites are columns, and a value of
264 1 indicates occurrence. We extracted the \mathbf{Y} matrix assuming that every pixel represents a unique location,
265 and following best practices⁴⁵ transformed it using Hellinger's distance to account for unequal bat
266 richness at different pixels. The correction of raw community data is particularly important for two
267 reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect
268 of overall species richness, which is already incorporated in the phylogenetic diversity component. High
269 values of LCBD indicate that the pixel has a community that is on average more dissimilar in species
270 composition than what is expected knowing the entire matrix, i.e. a more unique community. Recent
271 results by⁴⁶ shows that LCBD measures are robust with regards to spatial scale, and are therefore
272 applicable at the global scale.

273 **Viral sharing between hosts**

274 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a
275 previously published generalized additive mixed model of virus sharing by a tensor function of
276 phylogenetic distance and geographic range overlap across mammals.⁴⁷ This network stores pairwise
277 values of viral community similarity. To project viral sharing values into a single value for every pixel, we
278 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant
279 bat assemblage is likely to be proficient at exchanging viruses.

280 **Composite risk map**

281 To visualize the aggregated risk at the global scale, we combine the three individual risk components
282 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model.⁴⁸ In
283 this approach, every risk component gets assigned a component in the RGB color model (phylogenetic
284 diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid

285 RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no
286 phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for
287 each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the
288 risk as colors diverge towards combinations of values for three risk components. Out of the possible
289 combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic
290 diversity and low viral sharing,⁴⁹ in that this allows multiple independent host-virus coevolutionary
291 dynamics to take place in the same location. In the colorimetric space, this corresponds to yellow – because
292 the HSV space is more amenable to calculations for feature extraction,⁵⁰ we measured the risk level by
293 calculating the angular distance of the hue of each pixel to a reference value of 60 (yellow), and weighted
294 this risk level by the value component. Specifically, given a pixel with colorimetric coordinates (h, s, v) , its
295 ranged weighted risk value is

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

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

297 **Viral phyogeography and evolutionary diversification**

298 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed
299 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data
300 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR
301 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR
302 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated
303 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained
304 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and
305 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East
306 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus
307 RdRp sequences were then aligned using MAFFT⁵¹ v1.4.0 (Algorithm FFT-NS-2, Scoring matrix 200PAM /
308 k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree reconstructed in
309 IQ-TREE⁵² v1.6.12 with ModelFinder⁵³ ultrafast bootstrap approximation⁵⁴ with a general time reversible
310 model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of nucleotide

311 substitution (GTR+F+R5).

312 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat
313 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the
314 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary
315 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary
316 distinctiveness⁵⁵ for each of the viruses in the tree, then averaged these at the bat species level, projected
317 these values onto their geographic distributions, and averaged across every bat found in a given pixel. As
318 such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral
319 community believed to be associated with a particular subset of bats present.

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

321 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the
322 biogeography of their hosts. To test this idea, we loosely adapted a method from,^{56,57} who proposed a
323 phylogenetic method for the delineation of animal biogeographic regions. In their original method, a
324 distance matrix - where each row or column represents a geographic raster’s grid cell, and the dissimilarity
325 values are the “beta diversity similarity” of their community assemble - undergoes non-metric
326 multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a
327 four-color bivariate map. Here, we build on this idea with an entirely novel methodology. First, we
328 measure the phylogenetic distance between the different viruses in the betacoronaviruses tree by using the
329 cophenetic function in ape;⁵⁸ subsequently, we take a principal components analysis of that distance
330 matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional
331 space. We then take the first two principal components and, as with the evolutionary distinctiveness
332 analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

333 References

- 334 1.
- 335 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 336 2.
- 337 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 338 3.
- 339 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 340 4.
- 341 Agosta, S. J., Janz, N. & Brooks, D. R. How specialists can be generalists: Resolving the ‘parasite paradox’ and implications for emerging infectious disease. *Zoologia (Curitiba)* **27**, 151–162 (2010).
- 342 5.
- 343 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 344 6.
- 345 Geoghegan, J. L., Duchêne, S. & Holmes, E. C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLOS Pathogens* **13**, e1006215 (2017).
- 346 7.
- 347 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 348 8.
- 349 Thompson, J. N. *The Coevolutionary Process*. (University Of Chicago Press, 1994).
- 350 9.
- 351 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 352 10.
- 353 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 354 11.
- 355 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).

- 356 12.
- 357 Thrall, P. H., Hochberg, M. E., Burdon, J. J. & Bever, J. D. Coevolution of symbiotic mutualists and parasites
in a community context. *Trends in Ecology & Evolution* **22**, 120–126 (2007).
- 358 13.
- 359 Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *bioRxiv: The Preprint
Server for Biology* 2020.05.31.116061 (2020) doi:[10.1101/2020.05.31.116061](https://doi.org/10.1101/2020.05.31.116061).
- 360 14.
- 361 Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
- 362 15.
- 363 Becker, D. J. *et al.* Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The
Lancet Microbe* (2022) doi:[10.1016/S2666-5247\(21\)00245-7](https://doi.org/10.1016/S2666-5247(21)00245-7).
- 364 16.
- 365 Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications* **8**,
(2017).
- 366 17.
- 367 Cohen, L. E., Fagre, A. C., Chen, B., Carlson, C. J. & Becker, D. J. Sampling strategies and pre-pandemic
surveillance gaps for bat coronaviruses. 2022.06.15.496296 (2022) doi:[10.1101/2022.06.15.496296](https://doi.org/10.1101/2022.06.15.496296).
- 368 18.
- 369 Shi, J. J. *et al.* A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with Evidence for
Widespread Panmixia. *Acta Chiropterologica* **16**, 279–292 (2014).
- 370 19.
- 371 Kettenburg, G. *et al.* Full Genome Nobcovirus Sequences From Malagasy Fruit Bats Define a Unique
Evolutionary History for This Coronavirus Clade. *Frontiers in Public Health* **10**, (2022).
- 372 20.
- 373 Hazarie, S., Soriano-Paños, D., Arenas, A., Gómez-Gardeñes, J. & Ghoshal, G. Interplay between population
density and mobility in determining the spread of epidemics in cities. *Communications Physics* **4**, 191 (2021).
- 374 21.
- 375 Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.
Virologica Sinica **33**, 104–107 (2018).
- 376 22.
- 377 Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10**, 1030–1037
(2004).

- 378 23.
- 379 Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover
and emergence. *Nature Reviews Microbiology* **18**, 461–471 (2020).
- 380 24.
- 381 Van Brussel, K. & Holmes, E. C. Zoonotic disease and virome diversity in bats. *Current Opinion in Virology*
52, 192–202 (2022).
- 382 25.
- 383 Peixoto, F. P., Braga, P. H. P. & Mendes, P. A synthesis of ecological and evolutionary determinants of bat
diversity across spatial scales. *BMC ecology* **18**, 18 (2018).
- 384 26.
- 385 Simmons, N. B. & Cirranello, A. L. Bat Species of the World: A taxonomic and geographic database. <https://batnames.org/> (2020).
- 386 27.
- 387 Kasso, M. & Balakrishnan, M. Ecological and Economic Importance of Bats (Order Chiroptera). *ISRN
Biodiversity* **2013**, e187415 (2013).
- 388 28.
- 389 Mello, M. A. R. *et al.* The Missing Part of Seed Dispersal Networks: Structure and Robustness of Bat-Fruit
Interactions. *PLOS ONE* **6**, e17395 (2011).
- 390 29.
- 391 *Bats in the Anthropocene: Conservation of Bats in a Changing World.* (Springer International Publishing,
2016). doi:[10.1007/978-3-319-25220-9](https://doi.org/10.1007/978-3-319-25220-9).
- 392 30.
- 393 Williams-Guillén, K., Perfecto, I. & Vandermeer, J. Bats Limit Insects in a Neotropical Agroforestry System.
Science **320**, 70–70 (2008).
- 394 31.
- 395 Gonsalves, L., Bicknell, B., Law, B., Webb, C. & Monamy, V. Mosquito Consumption by Insectivorous Bats:
Does Size Matter? *PLOS ONE* **8**, e77183 (2013).
- 396 32.
- 397 Gonsalves, L., Lamb, S., Webb, C., Law, B. & Monamy, V. Do mosquitoes influence bat activity in coastal
habitats? *Wildlife Research* **40**, 10–24 (2013).
- 398 33.
- 399 Wu, Z. *et al.* A comprehensive survey of bat sarbecoviruses across China for the origin tracing of SARS-CoV
and SARS-CoV-2. (2021) doi:[10.21203/rs.3.rs-885194/v1](https://doi.org/10.21203/rs.3.rs-885194/v1).

- 400 34.
- 401 Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk. *Nature* 1–1 (2022)
doi:[10.1038/s41586-022-04788-w](https://doi.org/10.1038/s41586-022-04788-w).
- 402 35.
- 403 Alves, D. M. C. C., Diniz-Filho, J. A. F., da Silva e Souza, K., Gouveia, S. F. & Villalobos, F. Geographic
variation in the relationship between large-scale environmental determinants and bat species richness.
Basic and Applied Ecology **27**, 1–8 (2018).
- 404 36.
- 405 Treitler, J. T., Heim, O., Tschapka, M. & Jung, K. The effect of local land use and loss of forests on bats and
nocturnal insects. *Ecology and Evolution* **6**, 4289–4297 (2016).
- 406 37.
- 407 Johnson, C. K. *et al.* Global shifts in mammalian population trends reveal key predictors of virus spillover
risk. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20192736 (2020).
- 408 38.
- 409 Gryseels, S. *et al.* When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure in the
Spatial Spread of Arenaviruses. *PLOS Pathogens* **13**, e1006073 (2017).
- 410 39.
- 411 Amman, B. R. *et al.* *Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and
public health interest.* (FAO, 2011).
- 412 40.
- 413 IUCN. The IUCN Red List of Threatened Species. <https://www.iucnredlist.org> (2021).
- 414 41.
- 415 Rouault, E. *et al.* *GDAL/OGR Geospatial Data Abstraction software Library.* (Zenodo, 2022).
doi:[10.5281/ZENODO.5884351](https://doi.org/10.5281/ZENODO.5884351).
- 416 42.
- 417 Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10 (1992).
- 418 43.
- 419 Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: Species-level sets of phylogenies for
questions in ecology, evolution, and conservation. *PLOS Biology* **17**, e3000494 (2019).
- 420 44.
- 421 Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients
and partitioning. *Ecology Letters* **16**, 951–963 (2013).

- 422 45.
- 423 Legendre, P. & Condit, R. Spatial and temporal analysis of beta diversity in the Barro Colorado Island forest
dynamics plot, Panama. *Forest Ecosystems* **6**, 7 (2019).
- 424 46.
- 425 Dansereau, G., Legendre, P. & Poisot, T. Evaluating ecological uniqueness over broad spatial extents using
species distribution modelling. *Oikos* **n/a**, e09063 (2022).
- 426 47.
- 427 Albery, G. F., Eskew, E. A., Ross, N. & Olival, K. J. Predicting the global mammalian viral sharing network
using phylogeography. *Nature Communications* **11**, 2260 (2020).
- 428 48.
- 429 Seekell, D. A., Lapierre, J.-F. & Cheruvelil, K. S. A geography of lake carbon cycling. *Limnology and
Oceanography Letters* **3**, 49–56 (2018).
- 430 49.
- 431 Gomulkiewicz, R., Thompson, J. N., Holt, R. D., Nuismer, S. L. & Hochberg, M. E. Hot Spots, Cold Spots,
and the Geographic Mosaic Theory of Coevolution. *The American Naturalist* **156**, 156–174 (2000).
- 432 50.
- 433 Keke, S., Peng, Z. & Guohui, L. Study on skin color image segmentation used by Fuzzy-c-means arithmetic.
in *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* vol. 2 612–615 (2010).
- 434 51.
- 435 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in
Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
- 436 52.
- 437 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic
Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274
(2015).
- 438 53.
- 439 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model
selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589 (2017).
- 440 54.
- 441 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast
Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
- 442 55.
- 443 Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. Mammals on the EDGE: Conservation
Priorities Based on Threat and Phylogeny. *PLOS ONE* **2**, e296 (2007).

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

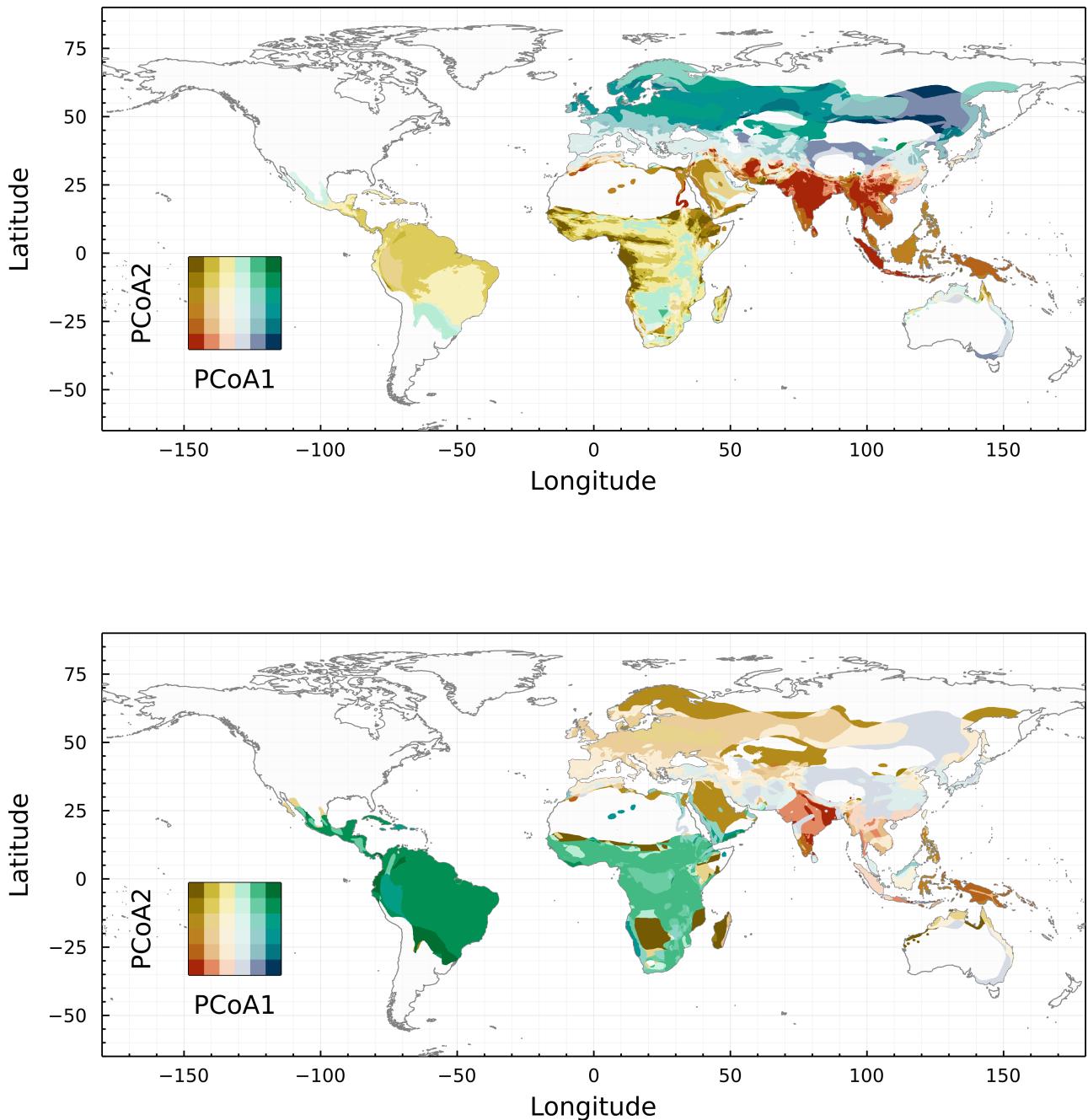


Figure 1: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

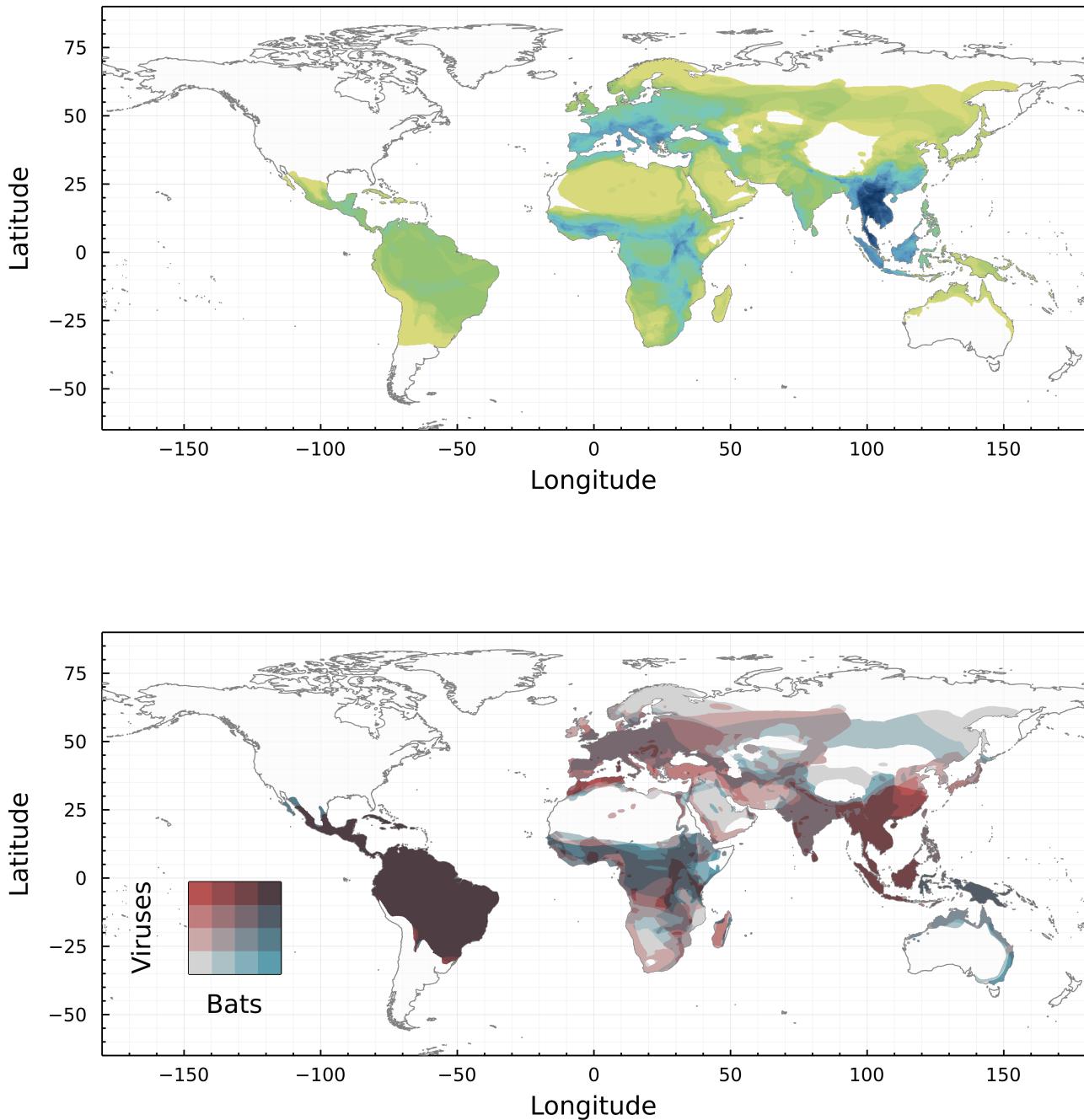


Figure 2: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).

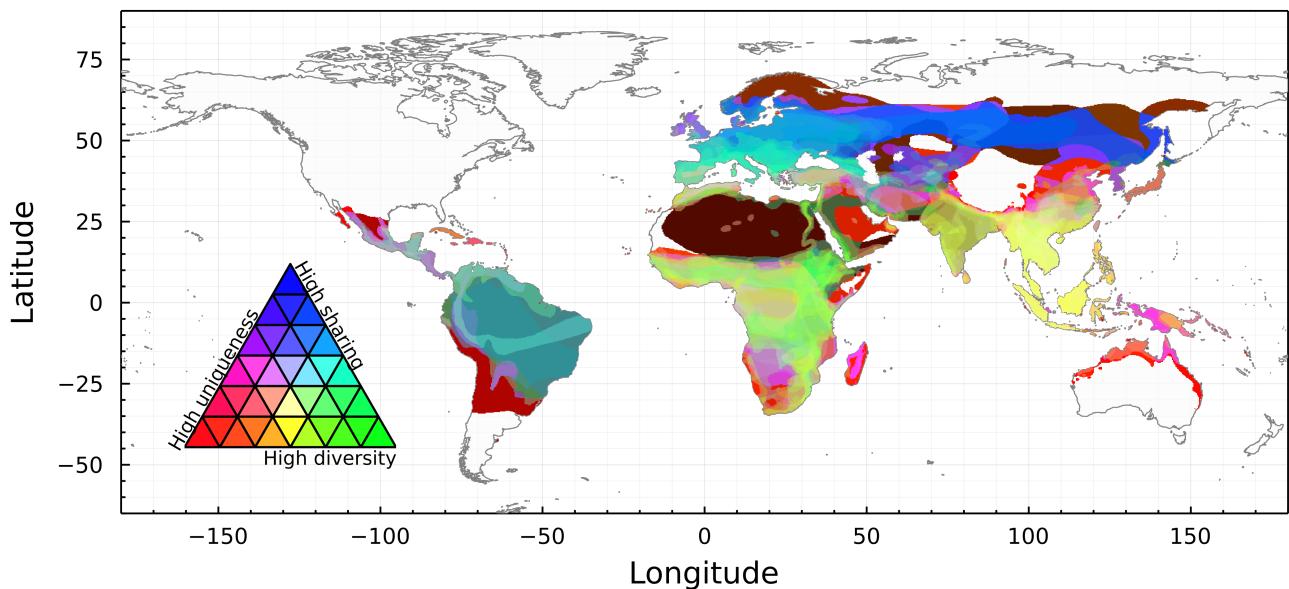


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible risk) would be a pure white (specifically $\text{RGB}(1.0, 1.0, 1.0)$), and a pixel with the lowest possible values would be pure black (specifically $\text{RGB}(0.0, 0.0, 0.0)$). Therefore, lighter values (the sum of the three channels gets closer to 3) indicate higher risk, and the color indicates the proportional distribution of the factors making up the total risk.

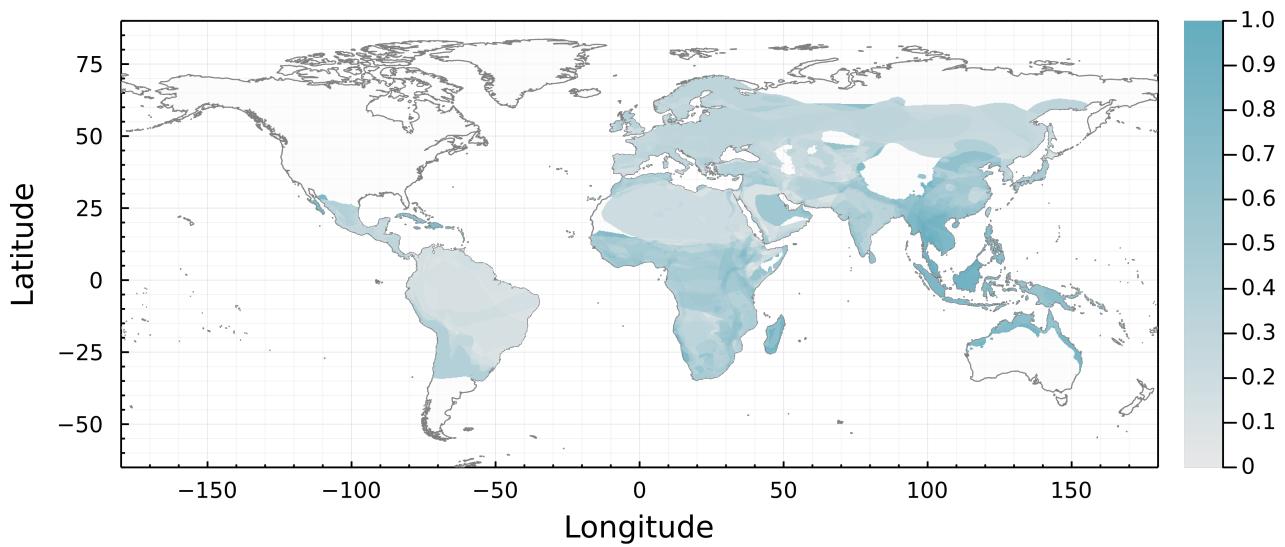


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

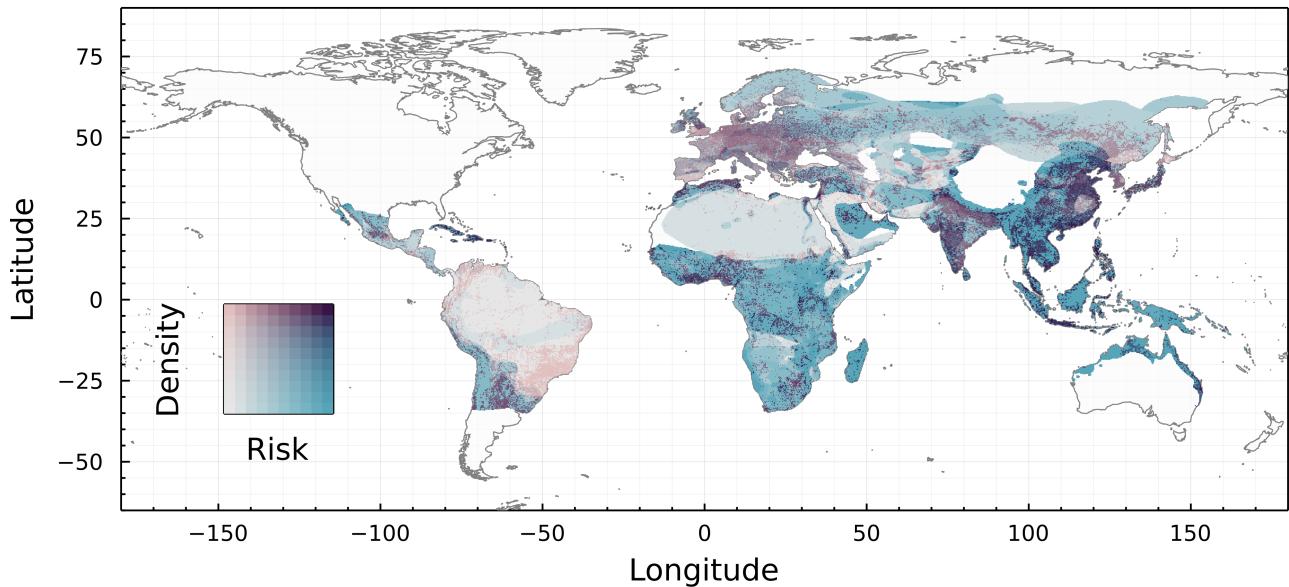


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk, in that the GMTC-derived risk of fig. 4 is high *and* the pixel is densely occupied by human populations. This approach increases the relative risk of several regions in Africa, and highlights the risk in India, southeast China, and the Arabian peninsula where areas of high to moderate risk overlap with areas of denser population.