

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

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

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

39 Results and Discussion

40 Bat and betacoronavirus biogeography are broadly consistent

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

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

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

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

79 [Figure 2 about here.]

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

85 Neotropics are one of the places where the fewest bat betacoronavirus sequences have been generated
86 (cite2), resulting in a sparser phylogenetic tree, and artificially inflating distinctiveness; conversely,
87 disproportionate research effort in eastern China²⁰ may have led to a more complete inventory of the local
88 diversity of coronaviruses, again inflating these metrics relative to underlying patterns. Even accounting
89 for these potential biases, though, there is obvious heterogeneity in betacoronavirus evolutionary
90 distinctiveness that is distinct from overall bat diversity.

91 Overall, these patterns recapitulate the evolutionary history of both the order Chiroptera and the genus
92 *Betacoronavirus*. Horseshoe bats (Rhinolophidae) include the reservoirs of the SARS-like viruses
93 (subgenus *Sarbecovirus*), the group of pandemic threats that have been of the greatest interest to
94 researchers¹³ (and so have been sampled most intensively).²⁰ The hotspots of host richness and viral
95 diversity in southeast Asia—both of which are disproportionately high, considering the global landscape
96 of bat species richness—are almost entirely driven by viral adaptive radiation through host switching
97 within this clade^{3,19}. In contrast, the Neotropical hotspot of viral distinctiveness is driven by isolation by
98 host vicariance. Out of the four main groups of betacoronaviruses, only merbecoviruses have been found
99 in animals in the Americas—an introduction that is generally presumed to be ancient.^{3,21} While
100 comparatively understudied, New World merbecoviruses have been found in the ghost-faced bats
101 (Mormoopidae), Neotropical leaf-nosed bats (Phyllostomidae), and free-tailed bats (Molossidae).^{14–17} The
102 former two groups and a clade of the latter are endemic to the Neotropics, while the explosive adaptive
103 radiations of the phyllostomids are responsible for the hotspot of bat diversity in the Amazon.²² Together,
104 these clades of New World bats play host to a distinct regime of betacoronavirus coevolution.

105 Host landscapes generate a coevolutionary mosaic

106 The existence of well-defined cophylogenetic regions suggests that the bat-betacoronavirus system is
107 spatially fragmented enough to create differential coevolutionary processes; in turn, the coevolutionary
108 mosaic contributes to heterogeneity in emergence risk. These ideas are, respectively, supported by the
109 existence of hotspots of viral uniqueness and the diverse origins of human betacoronaviruses. Together,
110 these ideas point to a predictable relationship between host community structure and coevolutionary
111 pressure: phylogeographic structure in bat hosts—and their diverse immune strategies—²³ creates a
112 landscape of selective pressure; the trajectory of viruses' coevolutionary response is, in turn, constrained
113 by the opportunities they have for either coevolutionary specialization or diversification through host

114 jumps and recombination.

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

126 [Figure 3 about here.]

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

139 Compared to approaches that map emergence risk based only on the number of known bat hosts of
140 betacoronaviruses, our framework suggests regions where high viral sharing dominates coevolutionary
141 dynamics—such as Latin America, or Eurasia above a northing of 30—would pose less of a relative risk of
142 zoonotic emergence. Nevertheless, areas of high host uniqueness coupled with high viral sharing

143 (red-to-pink in fig. 3) could create hotspots facilitated by viral codivergence. Our framework identifies
144 Madagascar, where most bat species are endemic following evolutionary divergence from sister species in
145 both African and Asian continents,²⁴ as one such hotspot; interestingly, a recent study²⁵ reported a novel
146 and highly divergent lineage of nobecoviruses from Madagascar-endemic pteropid bat species (*Pteropus*
147 *rufus* and *Rousettus madagascariensis*), again supporting the predictive power of the coevolutionary
148 framework.

149 [Figure 4 about here.]

150 Human landscapes filter the geography of emergence risk

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

163 [Figure 5 about here.]

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

170 progenitor viruses found in cave-dwelling horseshoe bats in Yunnan province;³³ nearby, antibody
171 evidence has indicated human exposure to SARS-like viruses.³⁴ MERS-CoV was originally detected in
172 Saudi Arabia, accompanied by a nearly identical virus sequenced from an Egyptian tomb bat (*Taphozous*
173 *perforatus*),³⁵ but is widespread in camels in East Africa and the Middle East, and may have reached its
174 bridge host decades earlier than originally supposed;³⁶ as a result, the geography of the original
175 bat-to-camel transmission is still widely regarded as uncertain. All of these are broadly consistent with the
176 risk factors we identify. Notably, India and west Africa are additional hotspots that have yet to experience
177 the emergence of a bat coronavirus into human populations, but may still be at risk—particularly given
178 known gaps in bat surveillance,²⁰ and a dense population in both regions with global connectivity. In any
179 of these regions, surveillance on viral reservoirs can be paired with targeted monitoring of high-risk
180 human populations (i.e., those with regular wildlife contact)³⁷ for maximum impact.

181 Conclusion

182 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to
183 human health.^{38,39} Chiropterans emerged around 64 million years ago and are one of the most diverse
184 mammalian orders, with an estimated richness of more than 1400 species.^{40,41} They exhibit a broad variety
185 of habitat use, behaviour, and feeding strategies, putting them at key positions in the delivery and
186 provisioning of several ecosystem services, tied to important ecosystem-derived benefits to human.⁴² For
187 example, bats are an essential component of many seed-dispersal networks.⁴³ Over two-thirds of bats are
188 known to be either obligate or facultative insectivores, therefore actively contributing for agricultural pest
189 control,^{44,45} and vectors of pathogens that put a risk on human health.^{46,47} Because bats are globally
190 distributed and have a long evolutionary history, phylogeographic and biogeographic approaches are
191 required to shed light on the contemporary distribution of coevolutionary processes between bats and the
192 pathogens they host. Not all areas in which bats, viruses, and human are co-occurring are facing a risk of
193 spillover towards human populations, and the areas in which this risk exist may not be facing risks of the
194 same nature and magnitude.

195 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries
196 like highly divergent nobecoviruses in Madagascar and the previously-neglected adaptive radiation of
197 sarbecoviruses outside of southern China and throughout southeast Asia. In doing so, it advances

198 ecological theory beyond the current state of the art for global maps of emergence risk. For example,
199 previous studies that have used host richness as proxy have predicted a high diversity of unsampled bat
200 viruses,¹⁸ bat coronaviruses,² and even specifically betacoronaviruses¹⁹ in both the Amazon and southeast
201 Asia. While we find that both regions are characterized by highly divergent host and viral communities,
202 our framework identifies key differences between the regions. We find that Latin America is a hotspot of
203 both host and viral distinctiveness, suggesting that this branch of the bat-betacoronavirus complex may be
204 undergoing independent evolutionary dynamics from the rest of the global pool, but with limited potential
205 for viral diversification—a finding that is supported by previous work indicating a higher rate of
206 codivergence in Latin America.² In contrast, in southeast Asia, host richness and viral distinctiveness are
207 high but sharing is low; this suggests a different type of evolutionary dynamics that could generate high
208 local diversity of viruses through host switching and viral recombination (see e.g.,¹³ as well as the
209 discovery of recombinant viruses that share genetic material from both the SARS-CoV and SARS-CoV-2
210 branches of the Sarbecovirus lineage).⁴⁸ Both of these regions are priority areas for sampling, especially
211 given predictions that they contain many bat hosts of undiscovered betacoronaviruses.^{19,20} However, both
212 the evolutionary and ecological aspects of emergence risk are likely higher in southeast Asia—a fact that
213 will only become more relevant, as bats track shifting climates and exchange viruses with other species,
214 creating a hotspot of cross-species transmission unique to the region.⁴⁹

215 The diversity and diversification potential of bats responds to anthropogenic factors others than shifting
216 climates.⁵⁰ Land use changes could significantly decrease bat suitability, notably through effects on diet
217 and availability of habitats.⁵¹ As our results establish that the diversification of bats betacoronaviruses
218 happens on top of processes affecting hosts, biogeographic variation in human population density and
219 anthropogenic disturbances may feed into co-evolutionary dynamics. Increase in humans-hosts contacts
220 also increase the risk of emergence of novel diseases,⁵² so does the changes in landscape connectivity at
221 local/regional scales.⁵³ This represents a challenge for both conservation strategies and disease ecology:
222 some areas can a high emergence risk and more potential for the acquisition of zoonotic viruses through
223 bat-human encounters.⁵⁴ In particular, the challenge ahead lies in the need to quantify actual exposure
224 (and risk) accounting for several transmission scenarios, including both direct and indirect bat - human
225 interactions, and feeding back into the provision of ecosystem services by bats.

226 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
227 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and

228 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research
229 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des
230 Données (IVADO). This research was enabled in part by support provided by Calcul Québec
231 (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). NF is funded by the NSERC
232 BIOS² CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by
233 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation. DJB was
234 supported by the National Institute of General Medical Sciences of the National Institutes of Health
235 (P20GM134973).

236 **Methods**

237 **Known *Betacoronavirus* hosts**

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

245 **Bat occurrences**

246 We downloaded the rangemap of every current bat species that was classified as an empirically
247 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data.⁵⁵ The range
248 maps were subsequently rasterized using the rasterize function from GDAL⁵⁶ at a resolution of
249 approximately 100kmx100km. For every pixel in the resulting raster where at least one bat host of
250 *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was used to
251 calculate the following risk assessment components: bat phylogenetic diversity, bat compositional
252 uniqueness, and predicted viral sharing risk.

253 **Bat phylogenetic diversity**

254 For every pixel, we measured Faith’s Phylogenetic Diversity⁵⁷ based on a recent synthetic tree with robust
255 time calibration, covering about 6000 mammalian species.⁵⁸ Faith’s PD measures the sum of unique
256 branches from an arbitrary root to a set of tips, and comparatively larger values indicate a more
257 phylogenetic diverse species pool. We measured phylogenetic diversity starting from the root of the entire
258 tree (and not from Chiroptera); this bears no consequences on the resulting values, since all branches
259 leading up to Chiroptera are only counted one per species pool, and (as we explain when describing the
260 assembly of the composite risk map), all individual risk components are ranged in [0,1]. This measure
261 incorporates a richness component, which we chose not to correct for; the interpretation of the

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

264 **Bat compositional uniqueness**

265 For every species pool, we measured its Local Contribution to Beta-Diversity;⁵⁹ LCBD works from a
266 species-data matrix (traditionally noted as \mathbf{Y}), where species are rows and sites are columns, and a value of
267 1 indicates occurrence. We extracted the \mathbf{Y} matrix assuming that every pixel represents a unique location,
268 and following best practices⁶⁰ transformed it using Hellinger's distance to account for unequal bat
269 richness at different pixels. The correction of raw community data is particularly important for two
270 reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect
271 of overall species richness, which is already incorporated in the phylogenetic diversity component. High
272 values of LCBD indicate that the pixel has a community that is on average more dissimilar in species
273 composition than what is expected knowing the entire matrix, i.e. a more unique community. Recent
274 results by⁶¹ shows that LCBD measures are robust with regards to spatial scale, and are therefore
275 applicable at the global scale.

276 **Viral sharing between hosts**

277 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a
278 previously published generalized additive mixed model of virus sharing by a tensor function of
279 phylogenetic distance and geographic range overlap across mammals.⁶² This network stores pairwise
280 values of viral community similarity. To project viral sharing values into a single value for every pixel, we
281 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant
282 bat assemblage is likely to be proficient at exchanging viruses.

283 **Composite risk map**

284 To visualize the aggregated risk at the global scale, we combine the three individual risk components
285 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model.⁶³ In
286 this approach, every risk component gets assigned a component in the RGB color model (phylogenetic
287 diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid

288 RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel with no sharing, no
289 phylogenetic diversity, and no compositional uniqueness is black, and a pixel with maximal values for
290 each is white. This additive model conveys both the intensity of the overall risk, but also the nature of the
291 risk as colors diverge towards combinations of values for three risk components. Out of the possible
292 combinations, the most risky in terms of rapid diversification and spillover potential is high phylogenetic
293 diversity and low viral sharing,⁶⁴ in that this allows multiple independent host-virus coevolutionary
294 dynamics to take place in the same location. In the colorimetric space, this correspond to yellow – because
295 the HSV space is more amenable to calculations for feature extraction,⁶⁵ we measured the risk level by
296 calculating the angular distance of the hue of each pixel to a reference value of 60 (yellow), and weighted
297 this risk level by the value component. Specifically, given a pixel with colorimetric coordinates (h, s, v), its
298 ranged weighted risk value is

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

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

300 **Viral phyogeography and evolutionary diversification**

301 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed
302 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data
303 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR
304 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR
305 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated
306 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained
307 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and
308 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East
309 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus
310 RdRp sequences were then aligned using MAFFT⁶⁶ v1.4.0 (Algorithm FFT-NS-2, Scoring matrix 200PAM /
311 k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree reconstructed in
312 IQ-TREE⁶⁷ v1.6.12 with ModelFinder⁶⁸ ultrafast bootstrap approximation⁶⁹ with a general time reversible
313 model with empirical base frequencies and the 5-discrete-rate-category FreeRaye model of nucleotide

314 substitution (GTR+F+R5).

315 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat
316 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the
317 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary
318 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary
319 distinctiveness⁷⁰ for each of the viruses in the tree, then averaged these at the bat species level, projected
320 these values onto their geographic distributions, and averaged across every bat found in a given pixel. As
321 such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral
322 community believed to be associated with a particular subset of bats present.

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

324 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the
325 biogeography of their hosts. To test this idea, we loosely adapted a method from,^{71,72} who proposed a
326 phylogenetic method for the delineation of animal biogeographic regions. In their original method, a
327 distance matrix - where each row or column represents a geographic raster’s grid cell, and the dissimilarity
328 values are the “beta diversity similarity” of their community assemble - undergoes non-metric
329 multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a
330 four-color bivariate map. Here, we build on this idea with an entirely novel methodology. First, we
331 measure the phylogenetic distance between the different viruses in the betacoronaviruses tree by using the
332 cophenetic function in ape;⁷³ subsequently, we take a principal components analysis of that distance
333 matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional
334 space. We then take the first two principal components and, as with the evolutionary distinctiveness
335 analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

336 **References**

- 337 1.
- 338 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 339 2.
- 340 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 341 3.
- 342 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 343 4.
- 344 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).
- 345 5.
- 346 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 347 6.
- 348 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).
- 349 7.
- 350 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 351 8.
- 352 Thompson, J. N. *The Coevolutionary Process*. (University Of Chicago Press, 1994).
- 353 9.
- 354 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 355 10.
- 356 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 357 11.
- 358 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).

- 359 12.
- 360 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).
- 361 13.
- 362 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).
- 363 14.
- 364 Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *The Journal of General Virology* **94**, 1028–1038
(2013).
- 365 15.
- 366 Góes, L. G. B. *et al.* Novel Bat Coronaviruses, Brazil and Mexico. *Emerging Infectious Diseases* **19**, 1711–1713
(2013).
- 367 16.
- 368 Góes, L. G. B. *et al.* Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil.
*Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious
Diseases* **44**, 510–513 (2016).
- 369 17.
- 370 Brandão, P. E. *et al.* A coronavirus detected in the vampire bat Desmodus rotundus. *Brazilian Journal of
Infectious Diseases* **12**, 466–468 (2008).
- 371 18.
- 372 Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
- 373 19.
- 374 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).
- 375 20.
- 376 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).
- 377 21.
- 378 Olival, K. J. *et al.* Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A
case study of bats. *PLoS pathogens* **16**, e1008758 (2020).
- 379 22.
- 380 Ammerman, L. K., Lee, D. N. & Tipps, T. M. First molecular phylogenetic insights into the evolution of free-
tailed bats in the subfamily Molossinae (Molossidae, Chiroptera). *Journal of Mammalogy* **93**, 12–28 (2012).

- 381 23.
- 382 Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Frontiers in Immunology* **11**, 26 (2020).
- 383 24.
- 384 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).
- 385 25.
- 386 Kettenburg, G. *et al.* Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a Unique Evolutionary History for This Coronavirus Clade. *Frontiers in Public Health* **10**, (2022).
- 387 26.
- 388 Rulli, M. C., D’Odorico, P., Galli, N. & Hayman, D. T. Land-use change and the livestock revolution increase the risk of zoonotic coronavirus transmission from rhinolophid bats. *Nature Food* **2**, 409–416 (2021).
- 389 27.
- 390 Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* **17**, 181–192 (2019).
- 391 28.
- 392 Muylaert, R. L. *et al.* Present and future distribution of bat hosts of sarbecoviruses: Implications for conservation and public health. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220397 (2022).
- 393 29.
- 394 Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. *Trends in ecology & evolution* **32**, 55–67 (2017).
- 395 30.
- 396 Worobey, M. *et al.* The Huanan market was the epicenter of SARS-CoV-2 emergence. *Zenodo* (2022)
doi:[10.5281/zenodo.6299600](https://doi.org/10.5281/zenodo.6299600).
- 397 31.
- 398 Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
- 399 32.
- 400 Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature microbiology* **5**, 1408–1417 (2020).
- 401 33.
- 402 Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS pathogens* **13**, e1006698 (2017).

- 403 34.
- 404 Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica* **33**, 104–107 (2018).
- 405 35.
- 406 Memish, Z. A. *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerging infectious diseases* **19**, 1819 (2013).
- 407 36.
- 408 Müller, M. A. *et al.* MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983–1997. *Emerging infectious diseases* **20**, 2093 (2014).
- 409 37.
- 410 Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10**, 1030–1037 (2004).
- 411 38.
- 412 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).
- 413 39.
- 414 Van Brussel, K. & Holmes, E. C. Zoonotic disease and virome diversity in bats. *Current Opinion in Virology* **52**, 192–202 (2022).
- 415 40.
- 416 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).
- 417 41.
- 418 Simmons, N. B. & Cirranello, A. L. Bat Species of the World: A taxonomic and geographic database. <https://batnames.org/> (2020).
- 419 42.
- 420 Kasso, M. & Balakrishnan, M. Ecological and Economic Importance of Bats (Order Chiroptera). *ISRN Biodiversity* **2013**, e187415 (2013).
- 421 43.
- 422 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).
- 423 44.
- 424 *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).

- 425 45.
- 426 Williams-Guillén, K., Perfecto, I. & Vandermeer, J. Bats Limit Insects in a Neotropical Agroforestry System. *Science* **320**, 70–70 (2008).
- 427 46.
- 428 Gonsalves, L., Bicknell, B., Law, B., Webb, C. & Monamy, V. Mosquito Consumption by Insectivorous Bats: Does Size Matter? *PLOS ONE* **8**, e77183 (2013).
- 429 47.
- 430 Gonsalves, L., Lamb, S., Webb, C., Law, B. & Monamy, V. Do mosquitoes influence bat activity in coastal habitats? *Wildlife Research* **40**, 10–24 (2013).
- 431 48.
- 432 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).
- 433 49.
- 434 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).
- 435 50.
- 436 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).
- 437 51.
- 438 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).
- 439 52.
- 440 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).
- 441 53.
- 442 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).
- 443 54.
- 444 Amman, B. R. *et al.* Investigating the role of bats in emerging zoonoses: Balancing ecology, conservation and public health interest. (FAO, 2011).
- 445 55.
- 446 IUCN. The IUCN Red List of Threatened Species. <https://www.iucnredlist.org> (2021).

- 447 56.
- 448 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).
- 449 57.
- 450 Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10 (1992).
- 451 58.
- 452 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).
- 453 59.
- 454 Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients
and partitioning. *Ecology Letters* **16**, 951–963 (2013).
- 455 60.
- 456 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).
- 457 61.
- 458 Dansereau, G., Legendre, P. & Poisot, T. Evaluating ecological uniqueness over broad spatial extents using
species distribution modelling. *Oikos* **n/a**, e09063 (2022).
- 459 62.
- 460 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).
- 461 63.
- 462 Seekell, D. A., Lapierre, J.-F. & Cheruvilil, K. S. A geography of lake carbon cycling. *Limnology and
Oceanography Letters* **3**, 49–56 (2018).
- 463 64.
- 464 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).
- 465 65.
- 466 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).
- 467 66.
- 468 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).

- 469 67.
- 470 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).
- 471 68.
- 472 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).
- 473 69.
- 474 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).
- 475 70.
- 476 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).
- 477 71.
- 478 Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proceedings of the National
Academy of Sciences* **104**, 5925–5930 (2007).
- 479 72.
- 480 Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions.
Journal of Biogeography **37**, 2029–2053 (2010).
- 481 73.
- 482 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.

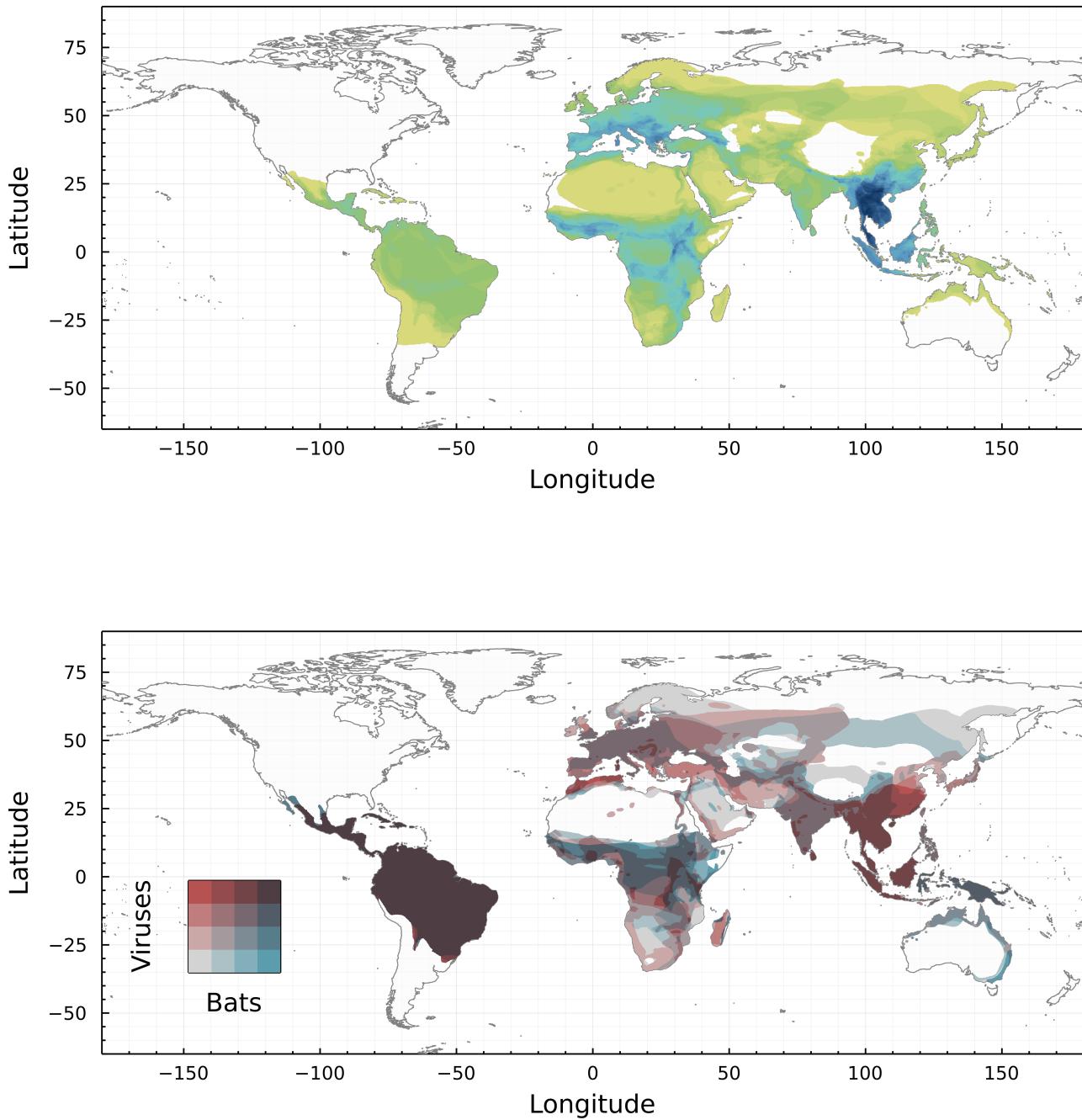


Figure 2: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).

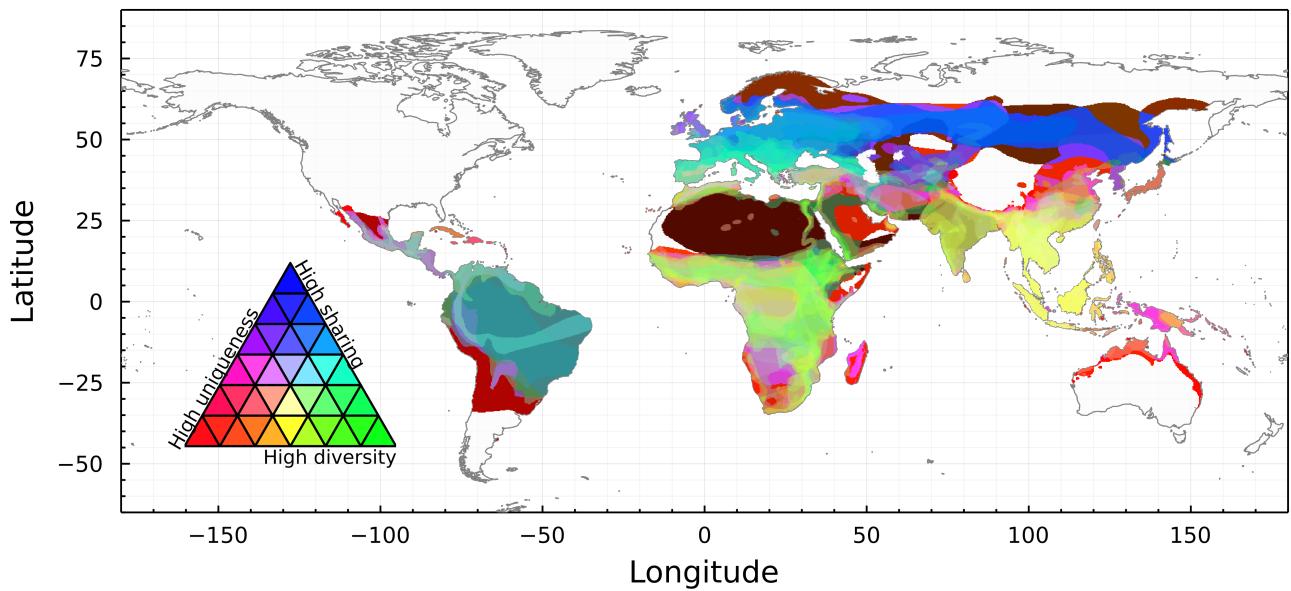


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. The main driver of emergence risk (possibility of spatially overlapped but coevolutionarily independent host-viral dynamics) corresponds to low viral sharing, *i.e.* pixels around yellow.

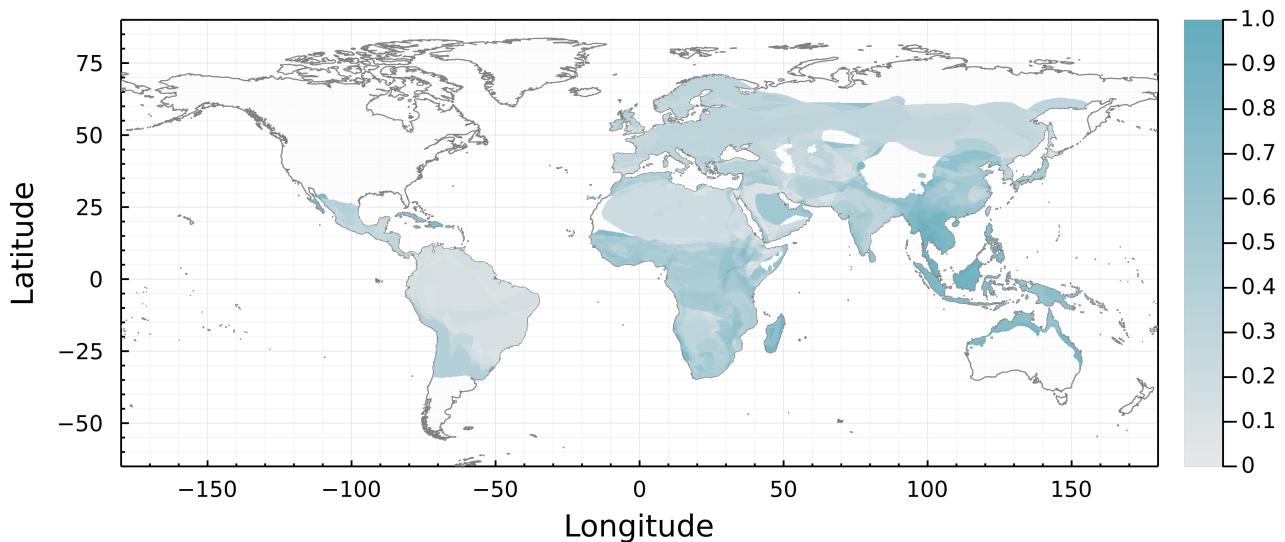


Figure 4: Extraction of a measure of *Betacoronavirus* emergence risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on these analyses, South-Eastern Asia, Madagascar, the Middle-East, and Africa below the Sahara desert have the highest relative risk of emergence.

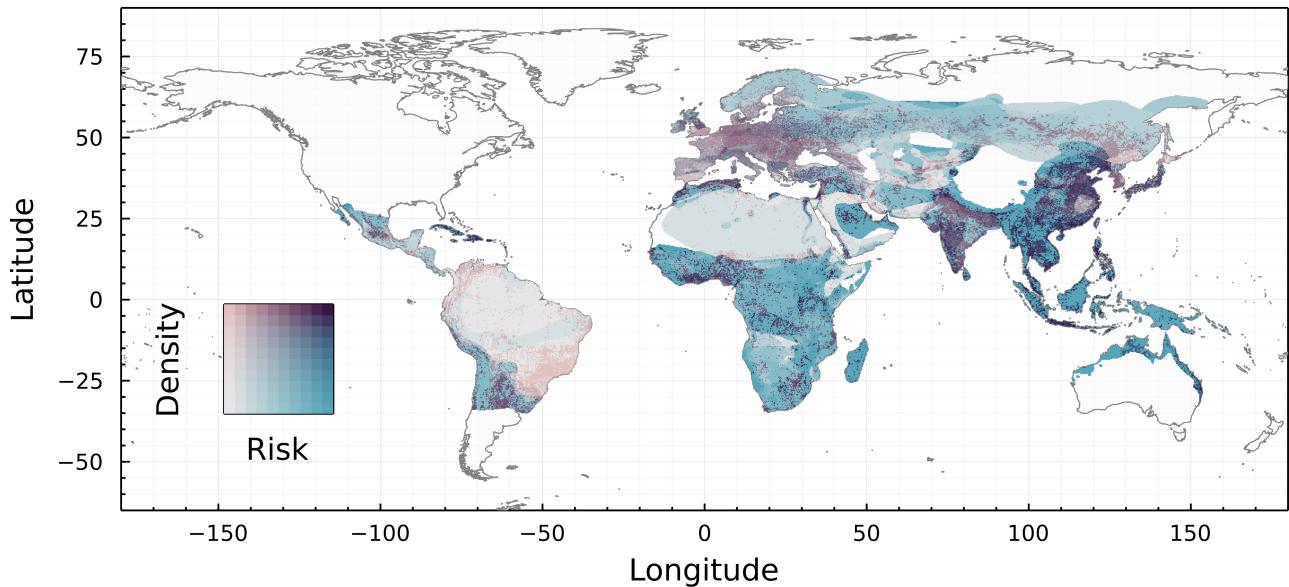


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk,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