

The coevolutionary mosaic of bat betacoronavirus emergence risk

Norma R Forero-Muñoz^{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

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

1 Disease emergence is complex, and is driven not only by animal-human contact, but also by the
2 underlying evolutionary dynamics in viral reservoirs.¹ Although host richness is often used as a superficial
3 proxy for spillover risk,[Anthony2017GloPat?](#),[Ruiz-Aravena2022EcoEvo?](#),[Sanchez2022Strategy?](#) these approaches
4 oversimplify the relevant interspecific heterogeneity in immunology, behavior, and other traits, and
5 therefore overlook unique host pools that allow for the rapid evolution of highly divergent viruses.² In the
6 case of generalist pathogens like betacoronaviruses, there is conceptual and empirical support to the idea
7 that these community-level mechanisms are even more important,³ particularly given that cross-species
8 transmission may, as a rule, structure viral evolution more than co-divergence with hosts.⁴ This creates a
9 disconnect between coevolutionary theory and most existing ecological frameworks for mapping spillover
10 risk.

11 The geographic mosaic theory of coevolution (GMTC) attempts to explicitly connect microevolutionary
12 dynamics to the macroecology and biogeography of symbiotic interactions.⁵ The GMTC posits that
13 coevolutionary processes among pairs⁶ or complexes⁷ of species are structured in space by the rippling
14 effects of abiotic conditions onto evolutionary mechanisms, giving rise to fragmented systems with
15 different ecologies over large spatial extents.⁸ The GMTC predicts a spatial fragmentation of
16 coevolutionary dynamics under the joint action of three processes:⁹ coevolutionary hot- and coldspots,
17 which appear when the intensity of *interaction* (in terms of reciprocal fitness consequences) varies
18 spatially; selection mosaics, wherein the intensity of *selection* varies across space, driven by both the biotic
19 complexity of the community (locally diverse hosts and viruses are more biotically complex) and the local
20 favorability of the environment;¹⁰ and trait remixing, which occurs when coevolutionary dynamics change
21 when community-level *functional traits* change through meta-community dynamics.

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—¹¹a 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.^{Anthony2017GloPat?, Ruiz-Aravena2022EcoEvo?} Projecting bat and betacoronavirus
44 phylogeny over space (fig. 1), we find support for the idea that bat community assembly is directly
45 responsible for a global mosaic of viral evolution. The distinct groupings (represented by different colors,
46 symbolizing positions in a subspace formed by the first two phylogenetic principal components) are
47 essentially equivalent between the two groups, and can be coarsely delineated as (1) south and southeast
48 Asia; (2) east Asia (including northern China), west Asia, and the Mediterranean coast; (3) Eurasia above
49 a northing of 40; and (4) Africa and Latin America. In some cases, this diverges from expectations about
50 coronavirus biogeography: for example, previous work has rarely flagged India as a region of interest, but
51 for both bats and betacoronaviruses, the subcontinent falls into the same regions as the southeast Asian
52 peninsula (and indeed, the region is home to known bat hosts of multiple betacoronavirus subgenera,
53 including nobecoviruses, sarbecoviruses, and merbecoviruses).¹²

54 [Figure 1 about here.]

55 Overall, these results suggest that the boundaries of bat and betacoronavirus biogeographic regions are
56 broadly consistent at a global scale; perfect matching between the biogeographic regions would have

57 indicated that the signal of virus distribution is fully predicted by bat hosts ranges. Areas for which the
58 biogeographic regions for bats and betacoronaviruses differ are primarily (i) southeast Asia and southern
59 China, and (ii) the Arabian peninsula, which are both regions where zoonotic transmission has been
60 documented (potentially driving a unique level of viral sampling effort that generates these patterns).
61 These spatially limited mismatches notwithstanding, the large level of congruence may be surprising,
62 given that cross-species transmission may play a stronger role in coronavirus diversification than
63 cospeciation—¹¹a property that would theoretically allow for substantial broad divergence in their
64 biogeography. However, host jumps at the family level or higher are relatively rare and significant events
65 in coronavirus evolutionary history;^{11,13} as a result, the mosaic of betacoronavirus phylogeography is
66 assembled from a set of overlapping smaller coevolutionary systems, superimposed in space and filtered
67 by the importance of different subgroups in local host communities. For example, the most speciose and
68 cosmopolitan family of bats, the vesper bats (Vespertilionidae), are considered the primary hosts of the
69 subgenus *Merbecovirus* (MERS-like viruses);^{12,13} but in the Americas, where merbecoviruses are the only
70 lineage present, they have only been found in other bat taxa (e.g., Molossidae, Phyllostomidae).^{14–17} At the
71 coarsest scale, these heterogeneities are lost, and betacoronavirus biogeography tracks the deep rifts in bat
72 evolutionary history—but within broad regions, the component coevolutionary systems may have very
73 different dynamics.

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

75 Bats, the second most diverse group of mammals, are found worldwide; gradients in their species
76 richness generally track broader patterns of mammal diversity,^{Tanalgo2022Mapping?} with a striking
77 Neotropical hotspot (especially in the Amazon basin) and a secondary hotspot centered in Indochina.
78 These hotspots of bat diversity are generally presumed to be hotspots of viral adaptive radiation, and
79 therefore areas of concern for human health.^{Anthony2017GloPat?,Olival2017HosVir?} However, the hotspots of
80 known bat betacoronavirus hosts show a distinct pattern, with primary hotspots (both in terms of area and
81 higher values) of host richness situated in southeast Asia, parts of southern Europe, and to a lesser extent
82 parts of Africa in the -25-0 range of latitudes (fig. 2; top). Although hundreds of species likely host
83 undiscovered betacoronaviruses, machine learning predictions have suggested that these undiscovered
84 reservoirs should follow the same diversity gradient.^{Becker2022OptPre?} In principle, these hotspots of
85 locally-diverse, virus-rich bat communities should drive more adaptive diversification in their viruses.

[Figure 2 about here.]

87 However, we find that the global pattern of betacoronavirus phylogenetic distinctiveness is quite distinct
88 from both bat host richness and phylogenetic distinctiveness (fig. 2; bottom). In contrast to the sparsity of
89 Neotropical betacoronavirus hosts, South and Central America have the most evolutionary distinct hosts
90 and viruses, followed by secondary hotspots in southeast Asia and the Rift Valley region have mostly
91 distinct viruses. Some degree of sampling bias may contribute to these patterns: for example, the
92 Neotropics are one of the places where the fewest bat betacoronavirus sequences have been
93 generated, Worobey2022HuaMar?, Temmam2022BatCor?, Boni2020EvoOri? resulting in a sparser phylogenetic tree, and
94 artificially inflating distinctiveness; conversely, disproportionate research effort in eastern
95 China Cohen2022SamStr? may have led to a more complete inventory of the local diversity of coronaviruses,
96 again inflating these metrics relative to underlying patterns. Even accounting for these potential biases,
97 though, there is obvious heterogeneity in betacoronavirus evolutionary distinctiveness that is distinct from
98 overall bat diversity.

99 Overall, these patterns recapitulate the evolutionary history of both the order Chiroptera and the genus
100 *Betacoronavirus*. Horseshoe bats (Rhinolophidae) include the reservoirs of the SARS-like viruses
101 (subgenus *Sarbecovirus*), the group of pandemic threats that have been of the greatest interest to
102 researchers Latinne2020OriCro? (and so have been sampled most intensively). Cohen2022SamStr? The hotspots of
103 host richness and viral diversity in southeast Asia—both of which are disproportionately high,
104 considering the global landscape of bat species richness—are almost entirely driven by viral adaptive
105 radiation through host switching within this clade Becker2022OptPre?, Ruiz-Aravena2022EcoEvo?. In contrast, the
106 Neotropical hotspot of viral distinctiveness is driven by isolation by host vicariance. Out of the four main
107 groups of betacoronaviruses, only merbecoviruses have been found in animals in the Americas—an
108 introduction that is generally presumed to be ancient. Ruiz-Aravena2022EcoEvo?, Olival2020PosRev? While
109 comparatively understudied, New World merbecoviruses have been found in the ghost-faced bats
110 (Mormoopidae), Neotropical leaf-nosed bats (Phyllostomidae), and free-tailed bats
111 (Molossidae). Anthony2013CorBat?, Goes2013NovBat?, Goes2016GenDiv?, Brandao2008CorDet? The former two groups and
112 a clade of the latter are endemic to the Neotropics, while the explosive adaptive radiations of the
113 phyllostomids are responsible for the hotspot of bat diversity in the Amazon. Ammerman2012FirMol? Together,
114 these clades of New World bats play host to a distinct regime of betacoronavirus coevolution.

115 Our approach is potentially limited by sampling bias: key hotspots identified by our model have, indeed,
116 been sampled intensely following major zoonotic emergence events. In these areas, more betacoronavirus
117 hosts will have been discovered, leading to higher overall diversity and potentially higher sharing.
118 Similarly, hotspots of evolutionary uniqueness - as in the Arabian peninsula - could reflect much broader
119 lineages that have only been sampled in focal areas for public health. While the discovery of new branches
120 of bat-betacoronavirus coevolution is certainly likely, and might change some of the observed patterns, our
121 framework is likely to be fairly robust: the 126 hosts in our study capture nearly 10% of global bat diversity,
122 and the underlying evolutionary patterns they represent are much less sensitive to new information than
123 any inferences about viral evolution.

124 **Coevolutionary regimes structure evolutionary potential for zoonotic emergence**

125 The existence of well-defined cophylogenetic regions suggests that the bat-betacoronavirus system is
126 spatially fragmented enough to create divergent coevolutionary trajectories; in turn, this coevolutionary
127 mosaic may alter the risk of zoonotic emergence. These ideas are, respectively, supported by the existence
128 of hotspots of viral uniqueness and the diverse origins of human betacoronaviruses. Together, this
129 framework points to a predictable relationship between host community structure and coevolutionary
130 pressure: phylogeographic structure in bat hosts (and their diverse immune strategies)^{Banerjee2020NovIns?}
131 creates a landscape of selective pressure; the trajectory of viruses' coevolutionary response is, in turn,
132 constrained by their opportunities for either specialization or diversification through host jumps and
133 recombination.

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

144 coevolutionary mosaics can form within cophylogenetic regions.

145 [Figure 3 about here.]

146 The greatest evolutionary potential for zoonotic emergence exists where pathogen pools have a high
147 genetic diversity and high propensity for cross-species transmission. In our framework, emergence risk is
148 therefore maximized under higher phylogenetic diversity (viruses are exposed to different host clades),
149 higher host uniqueness (viruses are experiencing novel, heterogeneous host traits combinations), and low
150 to medium viral sharing (host-virus pairs can coevolve independently, but divergent viruses may still have
151 opportunities for recombination). In fig. 3, this corresponds to yellow areas (dynamics dominated by low
152 viral sharing, with equal contributions of selection mosaics and trait remixing; southeast Asia, and the
153 Indian sub-continent), green-yellow areas (dynamics with low viral sharing but dominated by the
154 selection mosaic effect of host diversity; sub-Saharan Africa), and red-yellow areas (dynamics with low
155 viral sharing but dominated by trait remixing in host communities; the Middle East). Translating this axis
156 of variation back into a univariate risk map (fig. 4) highlights that this evolutionary landscape has a
157 striking correspondence to regions where zoonotic betacoronaviruses have previously emerged. Our
158 findings align with predictions regarding the spatial location of cross-species transmission. These
159 locations not only pose a potential risk of viral jumps that could endanger human health but also provide
160 valuable information for monitoring wildlife health. This could guide us to determine where and what
161 measures to implement for effectively monitoring wildlife and human betacoronavirus outbreaks before
162 they escalate to critical levels. Nevertheless, there are actually very few documented cases of emergence
163 events, and similarities could be some degree of coincidental.

164 Compared to approaches that map emergence risk based only on the number of known bat hosts of
165 betacoronaviruses, our framework suggests regions where high viral sharing dominates coevolutionary
166 dynamics—such as Latin America, or Eurasia above a northing of 30—would pose less of a relative risk of
167 zoonotic emergence. Nevertheless, areas of high host uniqueness coupled with high viral sharing
168 (red-to-pink in fig. 3) could create hotspots facilitated by viral codivergence. Our framework identifies
169 Madagascar, where most bat species are endemic following evolutionary divergence from sister species in
170 both African and Asian continents, ^{Shi2014DeeDiv?} as one such hotspot; interestingly, a recent
171 study ^{Kettenburg2022FulGen?} reported a novel and highly divergent lineage of nobecoviruses from
172 Madagascar-endemic pteropid bat species (*Pteropus rufus* and *Rousettus madagascariensis*), again

173 supporting the predictive power of the coevolutionary framework.

174 [Figure 4 about here.]

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

176 The relationship between the underlying pathogen pool and emergence risk is mediated by both
177 human-wildlife interfaces (the probability of spillover) and opportunities for onward horizontal
178 transmission (the probability that spillovers become epidemics)^{Plowright2017PatZoo?}. It must be noted that
179 the assessment of risk based on the GMTC mechanisms does not account for human presence; for this
180 reason, it represents “potential” level of risk, which must be re-evaluated in the light of human presence.

181 As a proxy for both, we finally overlaid the risk component from the composite map (see above) with the
182 proportion of built land, as a proxy for a mix of habitat disturbance, potential for bat synanthropy or
183 contact with bridge hosts like livestock,^{Rulli2021LanCha?,Cui2019OriEvo?} and human population density and
184 connectivity^{Plowright2017PatZoo?,Muylaert2022PreFut?,Hassell2017UrbDis?} (fig. 5). Accounting for these factors,
185 most of South America and Europe are at comparatively lower risk, as—although densely
186 populated—settlements tend to be in areas with lower potential risk. Conversely, regions like Malaysia and
187 the northern coast of Australia have a high evolutionary risk component, but should represent a relatively
188 lower effective risk due to low human density. However, southeast Asia, the Indian subcontinent, and
189 scattered hotspots in sub-Saharan Africa are at high risk due to the overlap between human populations
190 and natural opportunities for cross-species transmission of betacoronaviruses.

191 [Figure 5 about here.]

192 Reassuringly, these predictions correspond to the geographic origins of the three bat-origin coronaviruses
193 that have recently emerged in human populations. While available information puts the spillover of
194 SARS-CoV-2 in a live animal market in Wuhan, China, the ultimate origin of the virus is almost certainly
195 in a divergent lineage of sarbecoviruses from Indochina that was poorly characterized prior to the
196 pandemic.^{Worobey2022HuaMar?,Temmam2022BatCor?,Boni2020EvoOri?} Similarly, the SARS-CoV outbreak began in
197 Guangdong province in 2002, reaching humans through small carnivore bridge hosts, but was eventually
198 traced back to a set of likely progenitor viruses found in cave-dwelling horseshoe bats in Yunnan
199 province;^{Hu2017DisRic?} nearby, antibody evidence has indicated human exposure to SARS-like

200 viruses. Wang2018SerEvi? MERS-CoV was first detected in Jordan, but is widespread in camels in East Africa
201 and the Middle East, and may have reached its bridge host decades earlier than originally
202 supposed; Muller2014MerCor? as a result, the geography of the original bat-to-camel transmission is still
203 widely regarded as uncertain. All of these are broadly consistent with the risk factors we identify. Notably,
204 India and west Africa are additional hotspots that have yet to experience the emergence of a bat
205 coronavirus into human populations, but may still be at risk—particularly given known gaps in bat
206 surveillance, Cohen2022SamStr? and a dense population in both regions with global connectivity. In any of
207 these regions, surveillance on viral reservoirs can be paired with targeted monitoring of high-risk human
208 populations (i.e., those with regular wildlife contact) Xu2004EpiClu? for maximum impact.

209 Conclusion

210 Bats emerged around 64 million years ago, and are one of the most diverse mammalian orders, with more
211 than 1,400 estimated species. Peixoto2018SynEco?, Simmons2020BatSpe? They exhibit a broad variety of habitat use,
212 behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of several
213 ecosystem services, tied to important ecosystem-derived benefits to humans. Kasso2013EcoEco? Over
214 two-thirds of bats are known to be either obligate or facultative insectivores, therefore actively
215 contributing for agricultural pest control, Voigt2016BatAnt?, Williams-Guillen2008BatLim? and vectors of pathogens
216 that put a risk on human health; Gonsalves2013MosCon?, Gonsalves2013MosInf? some other species are essential
217 links in many seed-dispersal networks. Mello2011MisPar? However, many of these species face a high risk of
218 extinction, particularly given persecution and killings that sometimes follows from messaging about their
219 role in disease emergence. Areas where bats, viruses, and humans co-occur are not always hotspots of risk
220 for human health; as such, developing more precise ways to map zoonotic hazards can help bats and
221 humans coexist safely, and support the conservation of these important and unique animals.

222 Here, we propose a simple framework with broad explanatory power that helps contextualize discoveries
223 like highly divergent nobecoviruses in Madagascar and the once-neglected adaptive radiation of
224 sarbecoviruses in the Indochinese peninsula. In doing so, it advances ecological theory beyond the current
225 state of the art for global maps of emergence risk. For example, previous studies that have used host
226 richness as a proxy have predicted a high diversity of unsampled bat viruses, Olival2017HosVir? bat
227 coronaviruses, Anthony2017GloPat? and even specifically betacoronaviruses Becker2022OptPre? in both the

228 Amazon and southeast Asia. While we find that both regions are characterized by unique and diverse
229 communities of both hosts and viruses, our framework is able to identify key differences between the two
230 systems. We find that the merbecovirus complex in Latin America has been a unique branch of evolution
231 separate from the rest of the global pool, but with limited potential for viral diversification—a finding that
232 is supported by previous work indicating a higher rate of codivergence in Latin
233 America. [Anthony2017GloPat?](#), [Caraballo2022CroTra?](#) In contrast, in southeast Asia, host richness and viral
234 distinctiveness are high but sharing is low; this suggests a different type of evolutionary dynamics that
235 could generate high local diversity of viruses through host switching and viral recombination (see
236 e.g., [Latinne2020OriCro?](#) as well as the discovery of recombinant viruses with genetic material from both the
237 SARS-CoV and SARS-CoV-2 branches of the Sarbecovirus lineage). [Wu2021ComSur?](#) Both of these regions are
238 priority areas for sampling, especially given predictions that they contain many bat hosts of undiscovered
239 betacoronaviruses. [Becker2022OptPre?](#), [Cohen2022SamStr?](#) However, both the evolutionary and ecological aspects
240 of emergence risk are higher in southeast Asia—a fact that will only become more relevant, as bats track
241 shifting climates and exchange viruses with other species, creating a hotspot of elevated cross-species
242 transmission unique to the region. [Carlson2022CliCha?](#), [Muylaert2022PreFut?](#)

243 Our trivariate additive mapping of components of risk (fig. 3) aims to elicit the complexity of spatial
244 cross-species transmission risk beyond the mere presence or absence of the pathogen host in a specific
245 location. By considering coevolutionary factors such as viral sharing and host uniqueness, we suggest
246 insights that can aid in identifying potential locations for surveillance of betacoronavirus circulation and
247 assessing the risk of cross-species transmission to other mammals. In communities characterized by
248 diverse but unique host populations, with limited viral sharing between them, we could encounter viruses
249 that specialize in targeting the immune system of specific hosts. This implies a low likelihood of infecting
250 novel hosts but, once locally introduced into a new host (either a new species, or an immunologically
251 naïve population), the specialized virus could spread relatively easily due to encountering little immune
252 resistance (CITE PLOWRIGHT). With the right combination of viral traits, such as low disease-induced
253 mortality or high transmission rate, this could lead to successfully spread within the new host community.
254 However, while high adaptation to a specific host can be advantageous, it may also lead to maladaptation
255 when the pathogen encounters a new unsuitable host, potentially resulting in its extinction.

256 Plowright RK, Foley P, Field HE, Dobson AP, Foley JE, Eby P, Daszak P. Urban habituation,
257 ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying

258 foxes (*Pteropus* spp.). Proceedings of the Royal Society B: Biological Sciences. 2011 Dec
259 22;278(1725):3703-12.

260 Bats—and the spillover of their viruses—are also sensitive to anthropogenic factors others than climate
261 change, including deforestation and other kinds of habitat loss, increased stress, and greater contact with
262 potential bridge hosts like domesticated
263 species.^{Alves2018GeoVar?, Treitler2016EffLoc?, Rulli2021LanCha?, Mendenhall2014PreBio?} This represents a challenge for
264 both conservation strategies and pandemic prevention,^{Amman2011InvRoI?} but identifying areas at risk, and
265 protecting the health of bats and ecosystems within those zones, can be a win-win intervention for
266 both.^{Hopkins2021HowIde?, Plowright2021LanUse?, OHHLEP2022OneHea?} As we scale these predictions down in space
267 to finer spatial resolutions to guide public health actions,¹⁸ the incorporation of human activity predictors
268 will become more importyant.¹⁹

269 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
270 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and
271 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research
272 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des
273 Données (IVADO). This research was enabled in part by support provided by Calcul Québec
274 (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). NF is funded by the NSERC
275 BIOS² CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by
276 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation. DJB was
277 supported by the National Institute of General Medical Sciences of the National Institutes of Health
278 (P20GM134973).

279 **Methods**

280 **Known *Betacoronavirus* hosts**

281 We downloaded the data on bats hosts of *Betacoronavirus* from
282 <https://www.viralemergence.org/betacov> on Apr. 2022, Becker2022OptPre? and filtered it to “known”
283 hosts (established before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling
284 and competence assays since the initial data collection). The original database was assembled by a
285 combination of data mining and literature surveys, including automated alerts on the “bats” and
286 “coronavirus” keywords to identify novel empirical evidence of bats-betacoronaviruses associations; this
287 yielded a total of 126 known hosts, 47 of which were novel hosts. This host-virus list of interactions was
288 obtained through a comprehensive aggregation of GenBank data as well as systematic literature
289 searches, Becker2022OptPre?, Cohen2022SamStr? such that we have high confidence in its fitness for the purpose
290 of inference at a large spatial scale.

291 **Bat occurrences**

292 We downloaded the rangemap of every current bat species that was classified as an empirically
293 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data. IUCN2021IucRed?
294 The IUCN data have been assembled to support wildlife conservation efforts, and therefore we do not
295 expect that they are biased by wildlife disease sampling efforts or priority. The range maps were
296 subsequently rasterized using the rasterize function from GDAL RouaultEven2022GdaOgr? at a resolution of
297 approximately 100kmx100km at the equator. For every pixel in the resulting raster where at least one bat
298 host of *Betacoronavirus* was present, we extract the species pool (list of all known bat hosts), which was
299 used to calculate the following risk assessment components: bat phylogenetic diversity, bat compositional
300 uniqueness, and predicted viral sharing risk.

301 **Bat phylogenetic diversity**

302 For every pixel, we measured Faith’s Phylogenetic Diversity Faith1992ConEva? based on a recent synthetic tree
303 with robust time calibration, covering about 6000 mammalian species. Upham2019InfMam? Faith’s PD
304 measures the sum of unique branches from an arbitrary root to a set of tips, and comparatively larger

305 values indicate a more phylogenetic diverse species pool. We measured phylogenetic diversity starting
306 from the root of the entire tree (and not from Chiroptera); this bears no consequences on the resulting
307 values, since all branches leading up to Chiroptera are only counted once per species pool, and (as we
308 explain when describing the assembly of the composite risk map), all individual risk components are
309 ranged in [0,1]. This measure incorporates a richness component, which we chose not to correct for; the
310 interpretation of the phylogenetic diversity is therefore a weighted species richness, that accounts for
311 phylogenetic over/under-dispersal in some places.

312 **Bat compositional uniqueness**

313 For every species pool, we measured its Local Contribution to Beta-Diversity;^{Legendre2013BetDiv?} LCBD
314 works from a species-data matrix (traditionally noted as **Y**), where species are rows and sites are columns,
315 and a value of 1 indicates occurrence. We extracted the Y matrix assuming that every pixel represents a
316 unique location, and following best practices^{Legendre2019SpaTem?} transformed it using Hellinger's distance
317 to account for unequal bat richness at different pixels. The correction of raw community data is
318 particularly important for two reasons: first, it prevents the artifact of richer sites having higher
319 importance; second, it removes the effect of overall species richness, which is already incorporated in the
320 phylogenetic diversity component. High values of LCBD indicate that the pixel has a community that is on
321 average more dissimilar in species composition than what is expected knowing the entire matrix, i.e. a
322 more unique community. Recent results by^{Dansereau2022EvaEco?} shows that LCBD measures are robust with
323 regards to spatial scale, and are therefore applicable at the global scale.

324 **Viral sharing between hosts**

325 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a
326 previously published generalized additive mixed model of virus sharing by a tensor function of
327 phylogenetic distance and geographic range overlap across mammals.^{Albery2020PreGlo?} This network stores
328 pairwise values of viral community similarity, measured for all hosts (to maintain consistency with the
329 phylogenetic diversity measure) across all viruses; therefore, we consider that it accounts for some overall
330 similarity in the way hosts deal with viruses, and not only betacoronaviruses. There is empirical evidence
331 that capacity for cross-species transmission even between divergent species is generally high,²⁰ especially

332 for beta-coronaviruses.¹³ To project viral sharing values into a single value for every pixel, we averaged the
333 pairwise scores. High values of the average sharing propensity means that this specific extant bat
334 assemblage is likely to be proficient at exchanging viruses.

335 Composite risk map

336 To visualize the aggregated risk at the global scale, we combine the three individual risk components
337 (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color
338 model. [Seekell2018GeoLak?](#) In this approach, every risk component gets assigned a component in the RGB
339 color model (phylogenetic diversity is green, compositional uniqueness is red, and viral sharing is blue).
340 In order to achieve a valid RGB measure, all components are re-scaled to the [0,1] interval, so that a pixel
341 with no sharing, no phylogenetic diversity, and no compositional uniqueness is black, and a pixel with
342 maximal values for each is white. This additive model conveys both the intensity of the overall risk, but
343 also the nature of the risk as colors diverge towards combinations of values for three risk components. Out
344 of the possible combinations, the most risky in terms of rapid diversification and spillover potential is high
345 phylogenetic diversity and low viral sharing, [Gomulkiewicz2000HotSpo?](#) in that this allows multiple
346 independent host-virus coevolutionary dynamics to take place in the same location. In the colorimetric
347 space, this correspond to yellow – because the HSV space is more amenable to calculations for feature
348 extraction, [Keke2010StuSki?](#) we measured the risk level by calculating the angular distance of the hue of each
349 pixel to a reference value of 60 (yellow), and weighted this risk level by the value component. Specifically,
350 given a pixel with colorimetric coordinates (h, s, v) , its ranged weighted risk value is

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

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

352 Viral phyogeography and evolutionary diversification

353 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed
354 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data
355 from the GenBank Nucleotide database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR

356 betacoronavirus[All Fields]) NOT (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR
357 sars-cov-2[All Fields]). We added a single representative sequence for SARS-CoV-2 and manually curated
358 to remove sequences without the RNA-dependent RNA polymerase (RdRp) sequence or that contained
359 words indicating recombinant or laboratory strains including “patent”, “mutant”, “GFP”, and
360 “recombinant”. We filtered over-represented taxa including betacoronavirus 1, hCoV-OC43, Middle East
361 respiratory syndrome coronavirus, Murine hepatitis virus, and hCoV-HKU1. Curated betacoronavirus
362 RdRp sequences were then aligned using MAFFT^{Katoh2013MafMul?} v1.4.0 (Algorithm FFT-NS-2, Scoring
363 matrix 200PAM / k=2, gap open penalty 1.53m offset value 0.123) and a maximum likelihood tree
364 reconstructed in IQ-TREE^{Nguyen2015IqtFas?} v1.6.12 with ModelFinder^{Kalyaanamoorthy2017ModFas?} ultrafast
365 bootstrap approximation^{Hoang2018UfbImp?} with a general time reversible model with empirical base
366 frequencies and the 5-discrete-rate-category FreeRate model of nucleotide substitution (GTR+F+R5).

367 We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat
368 diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the
369 phylogeny, so there was a 1:1 correspondence between data sources) against the “mean evolutionary
370 distinctiveness” of the associated viruses. To calculate this, we derived the fair proportions evolutionary
371 distinctiveness^{Isaac2007MamEdg?} for each of the viruses in the tree, then averaged these at the bat species
372 level, projected these values onto their geographic distributions, and averaged across every bat found in a
373 given pixel. As such, this can be thought of as a map of the mean evolutionary distinctiveness of the
374 known viral community believed to be associated with a particular subset of bats present.

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

376 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the
377 biogeography of their hosts. To test this idea, we loosely adapted a method
378 from, ^{Kreft2007GloPat?,Kreft2010FraDel?} who proposed a phylogenetic method for the delineation of animal
379 biogeographic regions. In their original method, a distance matrix - where each row or column represents
380 a geographic raster’s grid cell, and the dissimilarity values are the “beta diversity similarity” of their
381 community assemble - undergoes non-metric multidimensional scaling (NMDS); the first two axes of the
382 NMDS are projected geographically using a four-color bivariate map. Here, we build on this idea with an
383 entirely novel methodology. First, we measure the phylogenetic distance between the different viruses in
384 the betacoronaviruses tree by using the cophenetic function in ape;^{Paradis2019ApeEnv?} subsequently, we take

385 a principal components analysis of that distance matrix (readily interchangeable for NMDS in this case) to
386 project the viral tree into an n-dimensional space. We then take the first two principal components and, as
387 with the evolutionary distinctiveness analysis, aggregated these to a mean host value and projected them
388 using a four-color bivariate map.

389 **Data availability statement**

390 The code to reproduce these analyses, as well as the data (with the exception of the IUCN rangemaps,
391 which must be downloaded from their website) are available in the [viralemergence/betamap](#) repository
392 on GitHub.

393 **References**

- 394 1.
- 395 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology* **15**, 502–510 (2017).
- 396 2.
- 397 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).
- 398 3.
- 399 Power, A. G. & Mitchell, C. E. Pathogen Spillover in Disease Epidemics. *The American Naturalist* **164**, S79–S89 (2004).
- 400 4.
- 401 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).
- 402 5.
- 403 Thompson, J. N. *The Geographic Mosaic of Coevolution*. (University Of Chicago Press, 2005).
- 404 6.
- 405 Thompson, J. N. *The Coevolutionary Process*. (University of Chicago Press, 1994).
- 406 7.
- 407 Janzen, D. H. When is it Coevolution? *Evolution* **34**, 611–612 (1980).
- 408 8.
- 409 Price, P. W. *Macroevolutionary Theory on Macroecological Patterns*. (Cambridge University Press, 2002).
doi:[10.1017/CBO9780511615030](https://doi.org/10.1017/CBO9780511615030).
- 410 9.
- 411 Gomulkiewicz, R. *et al.* Dos and don’ts of testing the geographic mosaic theory of coevolution. *Heredity* **98**, 249–258 (2007).
- 412 10.
- 413 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).
- 414 11.
- 415 Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evolution* **3**, (2017).
- 416 12.

- 417 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology* **20**, 299–314 (2022).
- 418 13.
- 419 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).
- 420 14.
- 421 Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *The Journal of General Virology* **94**, 1028–1038 (2013).
- 422 15.
- 423 Góes, L. G. B. *et al.* Novel Bat Coronaviruses, Brazil and Mexico. *Emerging Infectious Diseases* **19**, 1711–1713 (2013).
- 424 16.
- 425 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).
- 426 17.
- 427 Brandão, P. E. *et al.* A coronavirus detected in the vampire bat Desmodus rotundus. *Brazilian Journal of Infectious Diseases* **12**, 466–468 (2008).
- 428 18.
- 429 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).
- 430 19.
- 431 Ka-Wai Hui, E. Reasons for the increase in emerging and re-emerging viral infectious diseases. *Microbes and Infection* **8**, 905–916 (2006).
- 432 20.
- 433 Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences* **117**, 9423–9430 (2020).

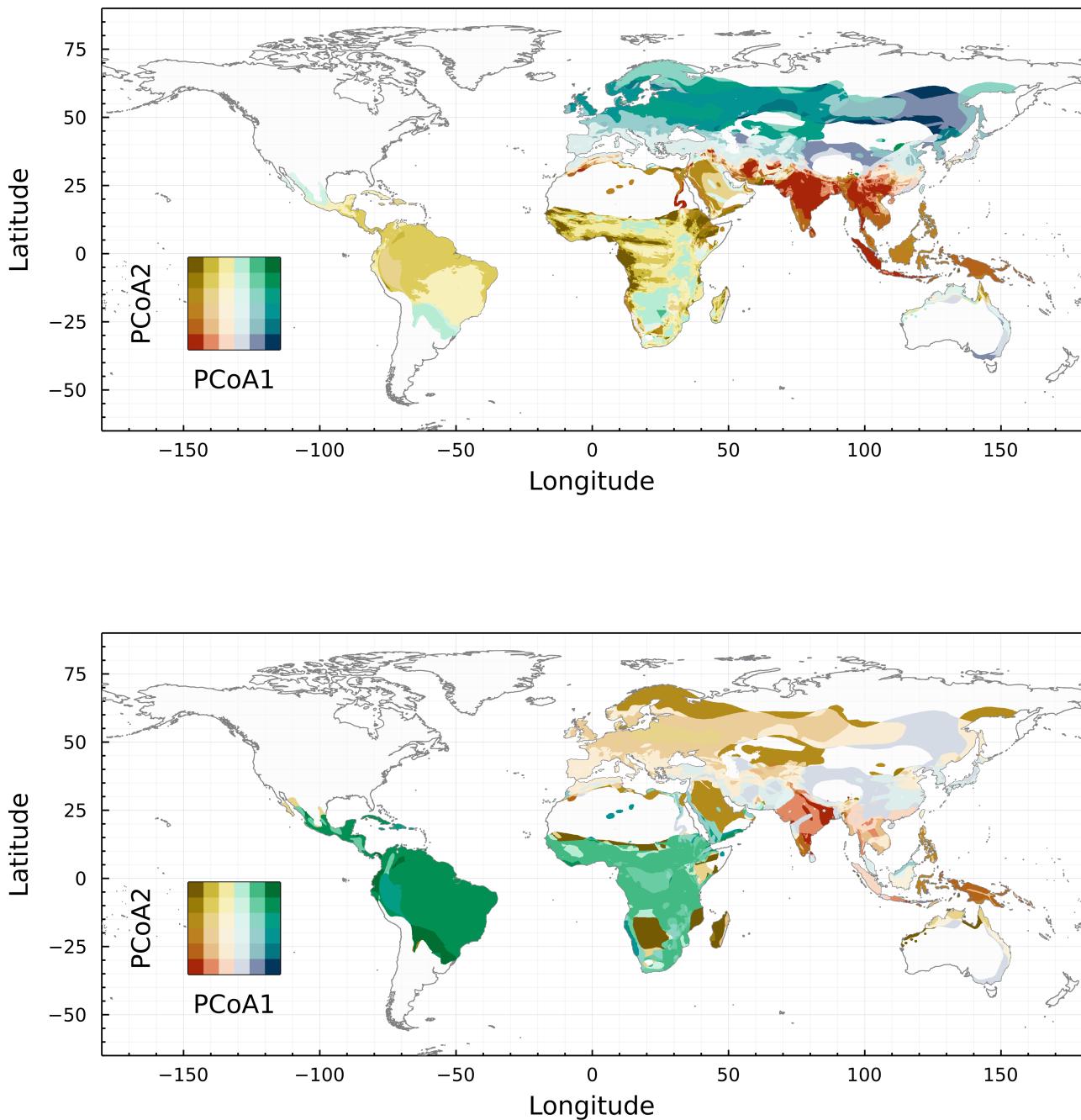


Figure 1: Bat and betacoronavirus biogeographic regions. Phylogeography of bats (top) and viruses (bottom) is categorized based on an analysis of bat distributions, paired with bat or virus phylogeny. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent, and so cannot be compared directly – that being said, the fact that different regions cluster in the same way across maps be directly compared.

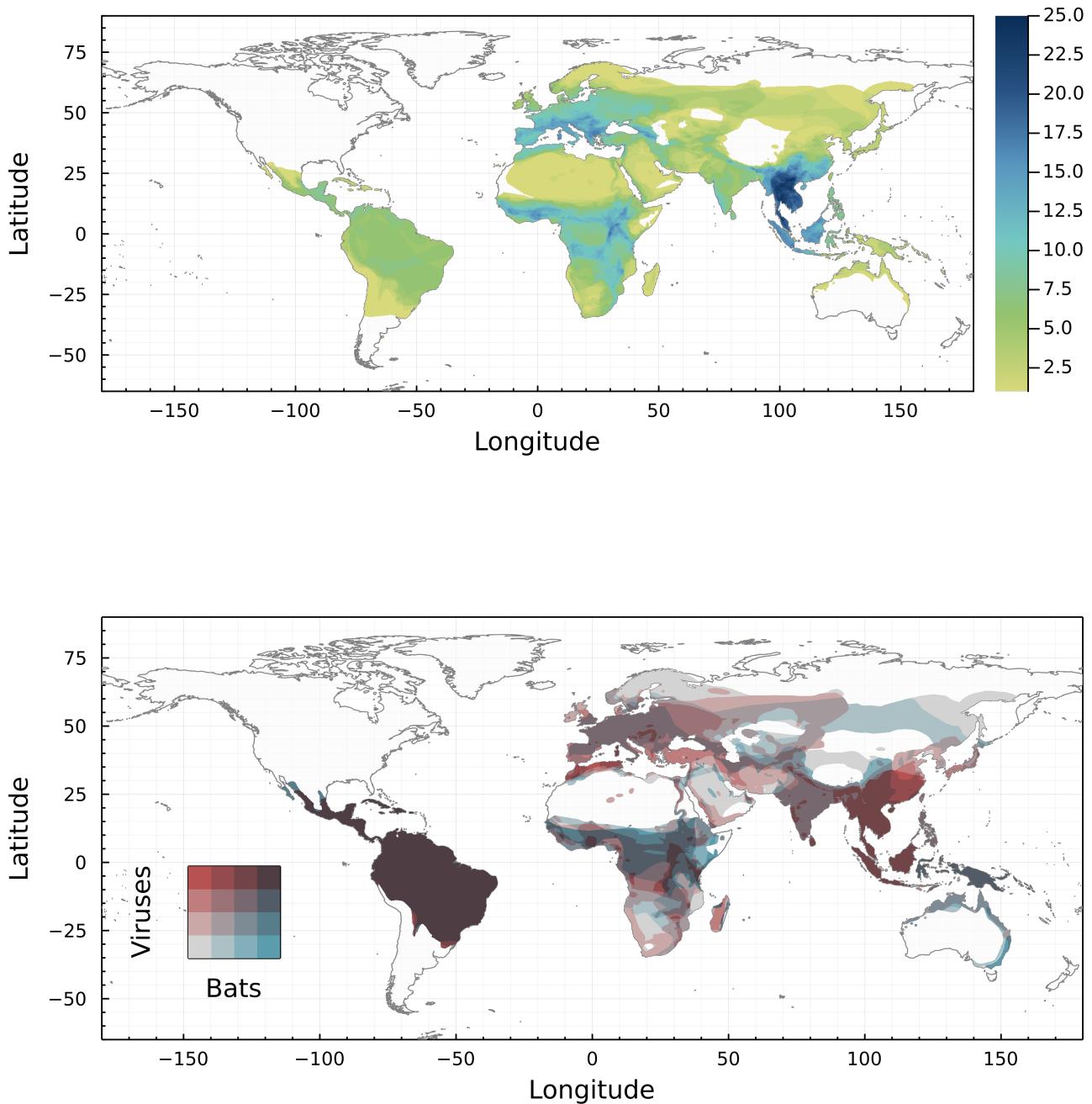


Figure 2: **Bat and betacoronavirus diversity.** Top panel: diversity of known bat hosts of betacoronaviruses in our dataset. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the *evolutionary* distinctiveness of the hosts (grey to blue) and the viruses (grey to red). Darker areas have higher combined evolutionary distinctiveness for the entire bat-virus system.

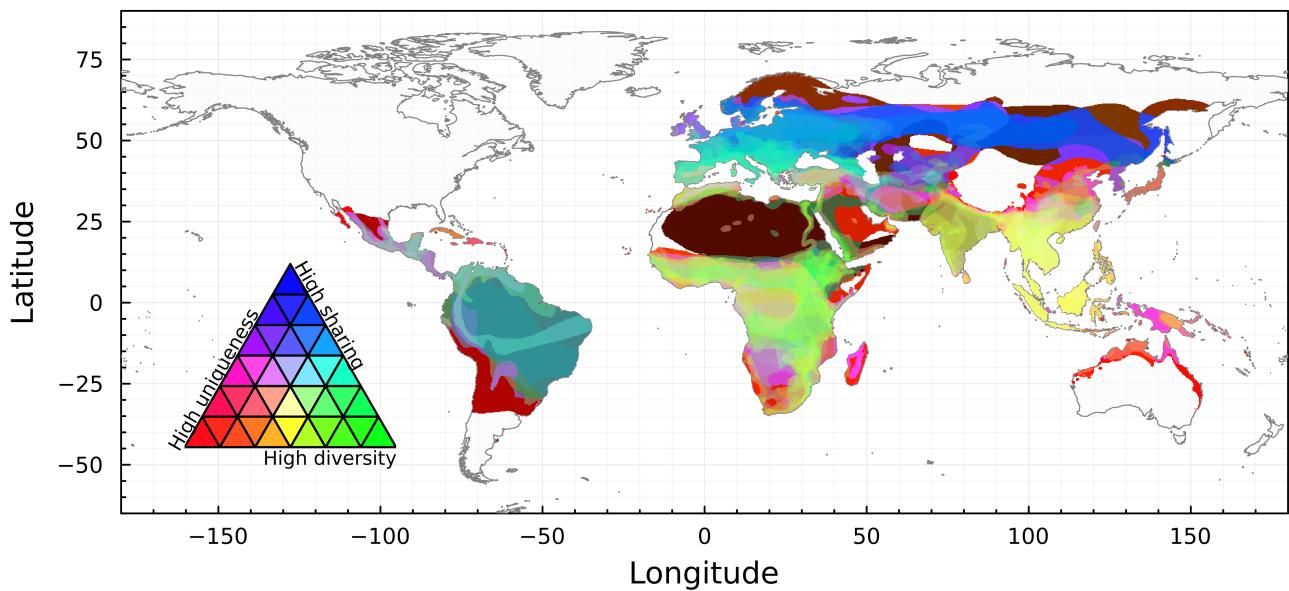


Figure 3: Trivariate additive mapping of the components of risk. Viral sharing runs from yellow (low) to blue (high); host phylogenetic diversity runs from pink (low) to high (green); and host compositional uniqueness runs from cyan (low) to red (high). The GMTC suggests that the highest evolutionary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

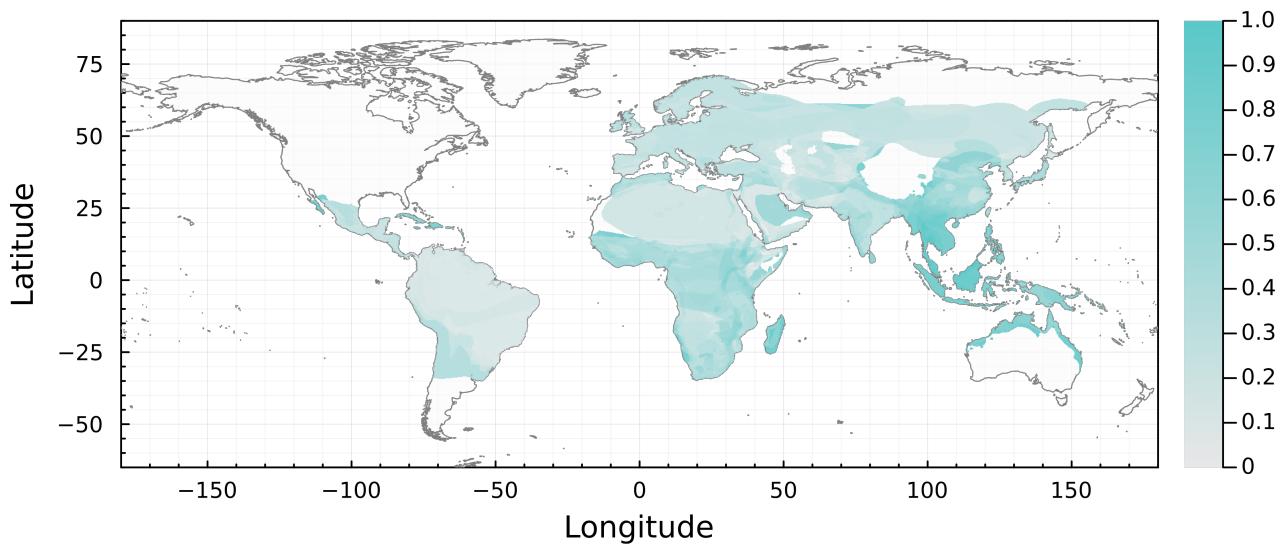


Figure 4: Evolutionary potential for zoonotic emergence of bat-origin betacoronaviruses. Risk is a composite measure of the color value and angular distance to the yellow hue in fig. 3 (see Methods). Darker pixels represent areas where the co-evolutionary mechanisms are likely to introduce a strong risk of emergence.

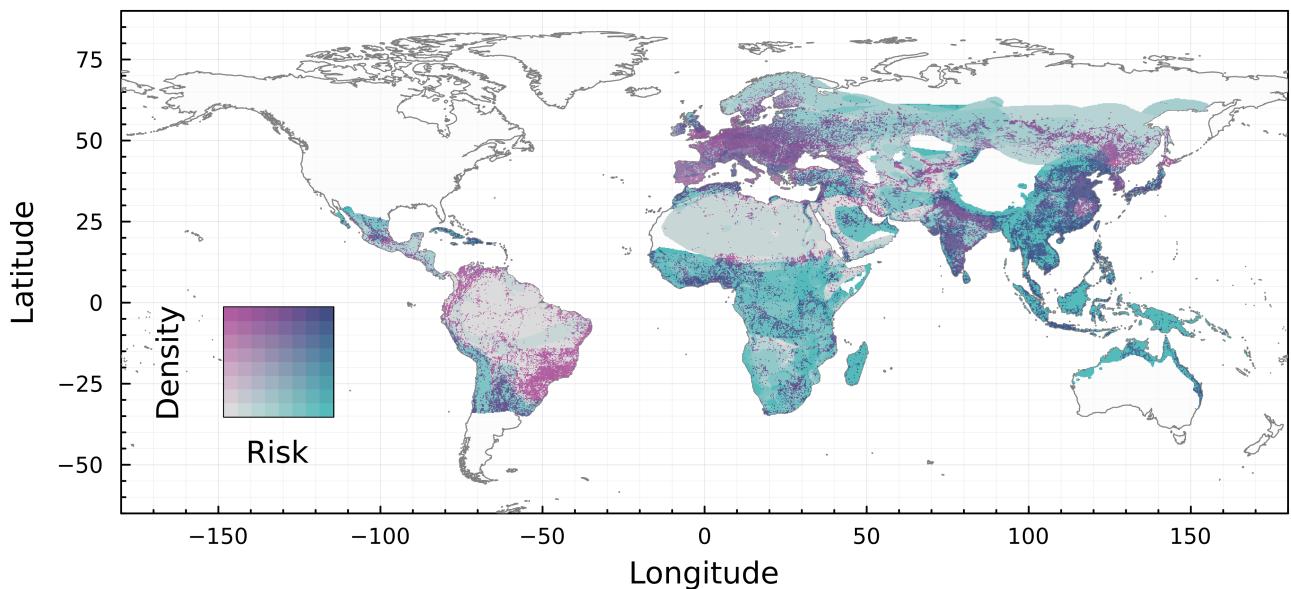


Figure 5: **Overlap between evolutionary potential and ecological opportunity for zoonotic emergence.** Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk, in that the GMTC-derived risk of fig. 4 is high *and* the pixel is densely occupied by human populations.