

The coevolutionary mosaic of bat-betacoronaviruses spillover risk

Norma Forero Rocio Munoz^{1,2} Renata L. Muylaert³ Stephanie N. Seifert⁴ Gregory F. Albery⁵ Colin J. Carlson⁵ **Timothée Poisot**^{1,2,‡}

¹ Université de Montréal ² Québec Centre for Biodiversity Sciences ³ Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand

⁴ Paul G. Allen School for Global Health, Washington State University, Pullman, WA, United States ⁵ ???

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Driven by the need to understand the ecological factors involved in the emergence of betacoronavirus (the genus causing the SARS and MERS disease 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 has many ecological and evolutionary origins. Our framework reveals that reservoir richness alone, although a component of viral hazard, is not a sufficiently integrative predictor of risk. We offer alternative insights based on viral sharing, host compositional uniqueness, and host phylogenetic diversity. By comparing our aggregated measure of risk to a proxy for human density, namely the proportion of each pixel that is covered by urban or built land, we provide a synthetic risk map, allowing the identification of hotspots where the bat-betacoronavirus system may originate novel viruses in humans.

1 Spillover risk is multidimensional and complicated. From the standpoint of an animal community, *i.e.* a
2 pool of suitable hosts, it is driven by a multiplicity of factors (Plowright et al. 2017). Proxies for the global
3 richness of hosts is one such component commonly analysed (see *e.g.* Anthony et al. 2017 for
4 coronaviruses), but there is an argument to be made that species who are not competent (or known) hosts
5 of a specific virus genus may not factor into this (Plowright et al. 2015); this is especially true as
6 competence data increases predictive power when the taxonomic scope of hosts of a viral family increases
7 (Becker et al. 2020, Mull et al. 2022). Similarly, host species who are assumed to share viruses at different
8 rates should be weighted accordingly (Albery et al. 2020). In mammals, key functional traits (for which
9 phylogeny is a reasonable proxy) are determinants of the spillover potential (Olival et al. 2017); these
10 include, notably, body mass and affinity for urban environments (Albery et al. 2022). Finally, especially
11 when the pool of potential hosts spans the entire globe, there may be local host pools that are highly
12 unique; not having been observed in other locations, these can act on the overall risk either by providing
13 novel contact opportunities, reflecting unique host-environment combinations (Engering et al. 2013), or
14 facilitating rapid evolutionary changes in specialism of their pathogens (Agosta et al. 2010). In the specific
15 case of generalist pathogens (which is the case many viruses in the betacoronavirus genus, see *e.g.*
16 MacLean et al. 2021), there is conceptual and empirical support to the idea that these community-level
17 mechanisms are even more important in driving the overall risk (Power and Mitchell 2004).

18 Bats are important reservoir hosts for different classes of microorganisms (Chu 2008, Donaldson 2010, Li
19 2010), many of which can threaten human health (Letko et al. 2020, Van Brussel and Holmes 2022).
20 Chiropterans emerged around 64 million years ago and are one of the most diverse mammalian orders,
21 with an estimated richness of more than 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020).
22 They exhibit a broad variety of habitat use, behaviour, and feeding strategies, putting them at key positions
23 in the delivery and provisioning of several ecosystem services, tied to important ecosystem-derived
24 benefits to human (Kasso and Balakrishnan 2013). For example, bats are an essential component of many
25 seed-dispersal networks (Mello et al. 2011). Over two-thirds of bats are known to be either obligate or
26 facultative insectivores, therefore playing an important role in the regulation of insect pests that can affect
27 crops (Williams-Guillén et al. 2008, Voigt and Kingston 2016), and vectors of pathogens that put a risk on
28 human health (Gonsalves et al. 2013a, b). Because bats are globally distributed and have a long
29 evolutionary history, phylogeographic and biogeographic approaches are required to shed light on the
30 contemporary distribution of coevolutionary processes between bats and the pathogens they host. Not all

31 areas in which bats, viruses, and human are co-occurring are facing a risk of spillover towards human
32 populations, and the areas in which this risk exist may not be facing risks of the same nature and
33 magnitude.

34 In this paper, we examine the biogeographic structure of bats-betacoronaviruses associations, based on a
35 curated dataset of all confirmed bat hosts of betacoronaviruses. This work is important both as a
36 description of the bats-betacoronavirus complex, but also because more broadly, bats are known reservoirs
37 for a variety of emerging viruses and other pathogens (Calisher et al. 2006, Melaun et al. 2014), making
38 balancing the needs for bat conservation and disease prevention most likely very difficult and a source of
39 human-wildlife conflicts, especially in more densely populated areas (Stone et al. 2015, Rego et al. 2015).
40 By drawing on concepts from the Geographic Mosaic Theory of Coevolution (Thompson 2005), we turn
41 these associations into a spatially explicit additive mapping of zoonotic risk components, revealing the
42 extreme heterogeneity of risk at the global scale. We identify the Amazon and South-Eastern Asia as
43 hotspots where the phylogenetic distinctiveness of betacoronaviruses is the highest (Anthony et al. 2017);
44 surprisingly, current data suggest that viral sharing between hosts is high in the Amazon and low in
45 South-Eastern Asia, which has the potential to result in different evolutionary dynamics between these
46 two regions.

47 **Methods**

48 **Known betacoronavirus hosts**

49 We downloaded the data on bats hosts of betacoronaviruses assembled by Becker et al. (2022) from
50 <https://www.viralemergence.org/betacov> on Apr. 2022, and filtered it to “known” hosts (established
51 before the emergence of SARS-CoV-2) and “novel” hosts (confirmed through sampling since the
52 emergence of SARS-CoV-2). The original database was assembled by a combination of data mining and
53 literature surveys, including automated alerts on the “bats” and “coronavirus” keywords to identify novel
54 empirical evidence of bats-betacoronaviruses associations.

55 **Bats occurrences**

56 We downloaded the rangemap of every extant bat species that was either classified as an empirically
57 documented host of beta-coronaviruses from the previous step, according to recent IUCN data (IUCN
58 2021). The range maps were subsequently rasterized using the `rasterize` function from GDAL (Rouault et
59 al. 2022) at a resolution of approximately **TK TP**. For every pixel in the resulting raster where at least one
60 bat host of betacoronavirus was present, we extract the species pool (list of all bat species), which was used
61 to calculate the following risk assessment components: phylogenetic diversity, bat compositional
62 uniqueness, and predicted viral sharing risk.

63 **Bats phyogeography**

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

74 **Bats compositional uniqueness**

75 For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres
76 2013); LCBD works from a species-data matrix (traditionally noted as \mathbf{Y}), where species are rows and sites
77 are columns, and a value of 1 indicates occurrence. We extracted the \mathbf{Y} matrix assuming that every pixel
78 represents a unique location, and following best practices (Legendre and Condit 2019) transformed it
79 using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw
80 community data is particularly important for two reasons: first, it prevents the artifact of richer sites
81 having higher importance; second, it removes the effect of overall species richness, which is already

82 incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a
83 community that is on average more dissimilar in species composition than what is expected knowing the
84 entire matrix, i.e. a more unique community. Recent results by Dansereau et al. (2022) shows that LCBD
85 measures are robust with regards to spatial scale, and are therefore applicable at the global scale.

86 **Viral sharing between hosts**

87 For all bat hosts of betacoronaviruses, we extracted their predicted viral sharing network (Albery et al.
88 2020). This network stores pairwise values of viral community similarity. To project viral sharing values
89 into a single value for every pixel, we averaged the pairwise scores. High values of the average sharing
90 propensity means that this specific extant bat assemblage is likely to be proficient at exchanging viruses.

91 **Composite risk map**

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

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

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

109 **Viral phylogeography and evolutionary diversification**

110 We used the following query to pull all betacoronavirus sequence data from the GenBank Nucleotide
 111 database except SARS-CoV-2; (“Betacoronavirus”[Organism] OR betacoronavirus[All Fields]) NOT
 112 (“Severe acute respiratory syndrome coronavirus 2”[Organism] OR sars-cov-2[All Fields]). We added a
 113 single representative sequence for SARS-CoV-2 and manually curated to remove sequences without the
 114 RNA-dependent RNA polymerase (RdRp) sequence or that contained words indicating recombinant or
 115 laboratory strains including “patent,” “mutant,” “GFP,” and “recombinant.” We filtered over-represented
 116 taxa including betacoronavirus 1, hCoV-OC43, Middle East respiratory syndrome coronavirus, Murine
 117 hepatitis virus, and hCoV-HKU1. Curated betacoronavirus RdRp sequences were then aligned using
 118 MAFFT v 1.4.0 (**Katoh and Standley 2013**, parameters in text?) and a maximum likelihood tree
 119 reconstructed in IQ-TREE v 1.6.12 (Nguyen et al. 2015) with ModelFinder (Kalyaanamoorthy et al. 2017)
 120 ultrafast bootstrap approximation (Hoang et al. 2018) and the following parameters (**STEPH WILL ADD**,
 121 parameters in text?).

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

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

131 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the
132 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,
133 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their
134 original method, a distance matrix - where each row or column represents a geographic raster's grid cell,
135 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes
136 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected
137 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel
138 methodology. First, we measure the phylogenetic distance between the different viruses in the
139 betacoronavirus tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently, we
140 take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this
141 case) to project the viral tree into an n-dimensional space. We then take the first two principal
142 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value
143 and projected them using a four-color bivariate map.

144 **Results and discussion**

145 **Host richness does not predict virus distinctiveness**

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

157 Nevertheless, locally diverse and virus-rich bat communities could represent an increased risk of spillover
158 under climate change through the creation of novel interactions (**Ice ice berg berg**), and therefore the
159 diversity of betacoronavirus strains should similarly be accounted for. In fig. 1 (bottom), we contrast the
160 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness
161 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as microchiroptera and
162 macrochiroptera, where macrochiroptera have an older history from an evolutionary perspective
163 compared to microchiroptera (Teeling et al. 2005, Springer 2013). Specifically, we would expect that the
164 so-called “New World” group of bats, being more evolutionary distinct, would also have evolutionary
165 distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a distinct
166 bat-betacoronavirus complex. By contrast, South-Eastern Asia has a lot of non-evolutionary distinct bats,
167 but evolutionary-distinct viruses.

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

182 **The phylogeographic regions of hosts and their viruses overlap**

183 Despite the difference in evolutionary distinctiveness globally, there are reasons to expect that the
184 phylogeography of bats and betacoronaviruses should show some degree of congruence. High density of
185 hosts sharing the same virus (albeit possibly different strains) can drive or result from evolution of the bat

186 antiviral immune system, resulting in spatially distinct immunological responses, as evidenced in several
187 bat species (Banerjee et al. 2020). Immune characteristics that allow bats to be better adapted to infection
188 by emerging viruses (Gorbunova et al., 2020; Irving et al., 2021), in addition to being hardcoded in their
189 genome (**Jebb et al. six...**), may be related to a wide variety of diets (Jones et al., 2022; Moreno Santillán
190 et al., 2021; Banerjee et al., 2020; Schneeberger et al., 2013; Muylaert et al., 2021), themselves likely to be
191 driven by spatial effects, especially at the local scale – bats, indeed, occupy a variety of environments, and
192 therefore display a variety of adaptations to these environments.

193 [Figure 2 about here.]

194 In fig. 2, we show a projection of the phylogeographic signal of bats (top) and viruses (bottom) in space;
195 the distinct groupings (represented by different colors symbolizing positions in the subspace formed by
196 the first two axes of the PCoA) are essentially equivalent between the two groups, and can be coarsely
197 delineated as southeast Asia, Eurasia above a northing of 25, and Africa and south America. These results
198 suggest that, although the evolutionary distinctiveness of the bat/betacoronavirus complex varies spatially,
199 the system shows an important degree of spatial consistency, with a reduced number of bioregions.
200 Available information describing the spillover of zoonotic betacoronaviruses of bat origin where data was
201 available before and up through the COVID-19 pandemic puts spillover events of SARS-CoV-2 in Wuhan,
202 China; SARS-CoV in XXX based on the presence of closest known viruses circulating in nature, and a
203 nearby location where serological (antibody) evidence has indicated human exposure to SARS-like viruses
204 (Wang et al. 2018 *Virologica Sinica*); MERS-CoV in XXX based on index cases available from a
205 recently-published compendium of cases (Ramshaw et al. 2019). For the latest event, most if not all index
206 cases are presumed to be camel-to-human transmission, and the precise origin point (if it exists) of
207 MERS-CoV in bats is uncertain. Recent recombinant canine coronavirus spillover events in Haiti (**ref**)
208 and Europe (**ref**) are not relevant here, as bats' involvement in these cycles of transmission have been
209 supposed to be non-existent.

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

211 As host richness, joint distinctiveness, or phylogeographic structure suggest that the bat/betacoronavirus
212 complex is globally fragmented enough to give rise to both different levels of risk (as evidenced by the
213 spatial location of spillover events) and different types of co-evolutionary dynamics, we turn to the

214 Geographic Mosaic Theory of Coevolution **REF** to provide a measure of risk accounting for multiple
215 processes. In fig. 3, we overlapped three components of spillover risk: viral sharing, *i.e.* the chance that
216 two bats will share viruses overall; Local Contribution to Beta Diversity, *i.e.* the fact that a bat community
217 is compositionally unique compared to the average compositional similarity across the entire system;
218 finally, host phylogenetic diversity, *i.e.* how dispersed the bats in a location are within the tree of life. This
219 approach leads to the definition of broad biogeographic regions of risk, where the same color represents
220 the same type of risk. By way of contrast to figures fig. 1 and fig. 2, these regions do not necessarily
221 overlap with previous spatial partitions of the bat/betacoronavirus complex.

222 [Figure 3 about here.]

223 From the perspective of spillover risk, the most important combination of factors is a high phylogenetic
224 diversity of hosts with low viral sharing; this, essentially, means that very different betacoronavirus could
225 co-exist within the same place. This is particularly the case given that betacoronaviruses often evolve and
226 even achieve host shifts through recombination, which requires the co-occurrence of sufficiently distinct
227 viruses to be a major driver of emergence. In fig. 3, this corresponds to yellow to pale green areas, which
228 are essentially limited to South-Eastern Asia, and to some part of Sub-Saharan Africa. Adopting a
229 geographic mosaic theory perspective on risk, other regions of the world are of lesser concern fig. 4. Our
230 risk decomposition does not account for viral diversity or distinctiveness. The simple rationale behind it is
231 that the acquisition of viral data is rarely disconnected from the acquisition of host data; furthermore, there
232 are more sources of information on hosts than on viruses, allowing to develop a host-centric perspective
233 on risk. Areas with high bat diversity and high turnover *may* facilitate the evolutionary radiation of
234 viruses, matching previous findings that the diversification of bat coronaviruses is driven largely by host
235 shifts (inter-genus or higher levels of cross-species transmission) and, to a lesser degree, cospeciation and
236 sharing (intra-genus cross-species transmission; Anthony et al. 2017). This diversification is not an actual
237 risk factor for spillover itself, but acts downstream of a spillover event by increasing the random chance of
238 the emergence of a virus with the raw genomic components required for the potential to infect humans.

239 [Figure 4 about here.]

240 From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide
241 hotspots of betacoronavirus risk through mixing of unique viruses (via codivergence) and in turn

242 recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat
243 species are endemic following evolutionary divergence from sister species in both African and Asian
244 continents (e.g. Shi et al. 2014). Recent surveillance (Kettenburg et al. 2022) has identified a novel
245 betacoronavirus (in the subgenus *Nobecovirus*) in Madagascar-endemic pteropid bat species (*Pteropus*
246 *rufus*, *Rousettus madagascariensis*), emphasizing strong proof of principle in model predictions.

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

248 Based on the previous result, we extracted the yellow component of the risk map (TP add methods), to
249 provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. However,
250 this maps the potential risk, which must be weighed by the potential for contacts with humans. As a proxy
251 for this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a
252 reasonable proxy for the density of humans per unit area, which increases the probability of pathogen
253 spread more widely (Hazarie et al., 2021). Since human activity is required to amplify the frequency of
254 virus encounters and thus create areas of viral amplification, mapping the potential risk against measures
255 of land use is required to generate a more actionable assessment of risk. This map is presented in fig. 5.
256 Most of South America and Europe are at low risk, as although densely populated, settlements tend to be
257 in areas with lower potential risk. However, this mapping reveals that South-East Asia, the Indian
258 subcontinent, and parts of sub-Saharan Africa, are at high risk due to the overlap between built areas and
259 bat communities representing more opportunities for cross-species transmission of betacoronaviruses.

260 [Figure 5 about here.]

261 **Conclusion**

262 Our study focuses largely on the biogeography of hosts. Yet, we know that viruses with high host
263 plasticity, that is, the ability of a given virus to adapt to various taxonomic orders and ecological groups
264 (Kreuder Johnson et al., 2015); are more likely to amplify viral spillover, followed by secondary
265 human-to-human transmission, and geographical spread (Hazarie et al., 2021). High viral host plasticity is
266 an especially important trait for RNA viruses such as betacov (Kreuder Johnson et al., 2015; Haddad et al.,
267 2021). Indeed, our analysis of viral sequences reveals that Latin America is a hotspot of viral

268 distinctiveness, suggesting that this part of the bats-betacov system may be undergoing independent
269 evolutionary dynamics (related species sharing viruses that are different from the rest of the global pool).
270 The other hotspot of viral distinctiveness is S.E. Asia, in which richness is high but sharing is low; this
271 suggests a different type of evolutionary dynamics (unrelated viruses coevolving with evolutionarily
272 distinct hosts, generating high diversity locally). Both of these areas should be priority areas for sampling,
273 especially since (**Becker?**) advance that they harbor undiscovered hosts of beta-coronaviruses. This
274 diversity of hosts, and the mechanisms by which the exchange of viruses occurs between species, is largely
275 affected by the local environmental conditions and environmental change.

276 There are several factors that drive changes in the diversity of bats (Alves et al., 2018), but human
277 activities' effects on the ecosystem (like modifications of land use) could significantly decrease it.
278 Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their
279 biogeographic variation, and human population density and other anthropogenic factors are decisive
280 moderators for its implications in public health. With the increase of contact between humans and
281 potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al., 2020), as previous
282 studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal
283 (Gryseels et al., 2017). One of these scenarios where interaction between bats and humans can occur can
284 be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats
285 not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and
286 humans (Deshpande et al., XX). This represents a challenge for conservation strategies and disease
287 ecology since we have areas with both potential for the acquisition of zoonotic viruses and bat-human
288 interaction; in particular, the challenge lies in the fact that actual exposure must be quantified from
289 several scenarios, including both direct and indirect bat - human interaction.

290 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional
291 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and
292 Omàmiwininiwak nations. This work was supported by funding to the Viral Emergence Research
293 Initiative (VERENA) consortium including NSF BII 2021909 and a grant from Institut de Valorisation des
294 Données (IVADO). This research was enabled in part by support provided by Calcul Québec
295 (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). NF is funded by the NSERC
296 BIOS² CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by
297 Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation.

298 **References**

- 299 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and
300 implications for emerging infectious disease. - *Zoologia (Curitiba)* 27: 151–162.
- 301 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -
302 *Nature Communications* 11: 2260.
- 303 Albery, G. F. et al. 2022. Urban-adapted mammal species have more known pathogens. *in press*.
- 304 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - *Nature*
305 *Communications in press*.
- 306 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - *Virus Evolution* *in press*.
- 307 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - *Frontiers in Immunology* 11: 26.
- 308 Becker, D. J. et al. 2020. Beyond Infection: Integrating Competence into Reservoir Host Prediction. -
309 *Trends in Ecology & Evolution* 35: 1062–1065.
- 310 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -
311 *The Lancet Microbe* *in press*.
- 312 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - *Clinical Microbiology*
313 *Reviews* 19: 531–545.
- 314 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - *Ecol. Lett.*
315 12: 693–715.
- 316 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species
317 distribution modelling. - *Oikos* n/a: e09063.
- 318 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of
319 SARS. - *Antiviral Research* 101: 45–56.
- 320 Engering, A. et al. 2013. Pathogen–host–environment interplay and disease emergence. - *Emerging*
321 *Microbes & Infections* 2: e5.
- 322 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - *Biological Conservation* 61: 1–10.

- 323 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -
324 The American Naturalist 156: 156–174.
- 325 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:
326 10–24.
- 327 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE
328 8: e77183.
- 329 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.
330 - PLOS ONE 2: e296.
- 331 IUCN 2021. The IUCN Red List of Threatened Species.
- 332 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -
333 ISRN Biodiversity 2013: e187415.
- 334 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010
335 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.
- 336 Kettenburg, G. et al. 2022. Full Genome Nobcovirus Sequences From Malagasy Fruit Bats Define a
337 Unique Evolutionary History for This Coronavirus Clade. - Frontiers in Public Health in press.
- 338 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - Proceedings of
339 the National Academy of Sciences 104: 5925–5930.
- 340 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species
341 distributions. - Journal of Biogeography 37: 2029–2053.
- 342 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity
343 coefficients and partitioning (H Morlon, Ed.). - Ecology Letters 16: 951–963.
- 344 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado
345 Island forest dynamics plot, Panama. - Forest Ecosystems 6: 7.
- 346 Letko, M. et al. 2020. Bat-borne virus diversity, spillover and emergence. - Nature Reviews Microbiology
347 18: 461–471.
- 348 MacLean, O. A. et al. 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist
349 virus and highly capable human pathogen. - PLOS Biology 19: e3001115.

- 350 Melaun, C. et al. 2014. Bats as Potential Reservoir Hosts for Vector-Borne Diseases. - In: Kliment, S. and
351 Mehlhorn, H. (eds), *Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths*.
352 *Parasitology Research Monographs*. Springer, pp. 25–61.
- 353 Mello, M. A. R. et al. 2011. The Missing Part of Seed Dispersal Networks: Structure and Robustness of
354 Bat-Fruit Interactions. - *PLOS ONE* 6: e17395.
- 355 Mollentze, N. and Streicker, D. G. 2020. Viral zoonotic risk is homogenous among taxonomic orders of
356 mammalian and avian reservoir hosts. - *Proceedings of the National Academy of Sciences* 117:
357 9423–9430.
- 358 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with
359 emerging deadly viruses? - *Memórias do Instituto Oswaldo Cruz* 110: 1–22.
- 360 Mull, N. et al. 2022. Virus isolation data improve host predictions for New World rodent
361 orthohantaviruses. - *Journal of Animal Ecology* in press.
- 362 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - *Nature* 546:
363 646–650.
- 364 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary
365 analyses in R. - *Bioinformatics* 35: 526–528.
- 366 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across
367 spatial scales. - *BMC ecology* 18: 18.
- 368 Plowright, R. K. et al. 2015. Ecological dynamics of emerging bat virus spillover. - *Proceedings of the
369 Royal Society B: Biological Sciences* 282: 20142124.
- 370 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - *Nature Reviews Microbiology* 15: 502–510.
- 371 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - *The American
372 Naturalist* 164: S79–S89.
- 373 Rego, K. M. da C. et al. 2015. Assessing human-bat interactions around a protected area in northeastern
374 Brazil. - *Journal of Ethnobiology and Ethnomedicine* 11: 80.
- 375 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 376 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - *Limnology and Oceanography Letters* 3:

- 377 49–56.
- 378 Shi, J. J. et al. 2014. A Deep Divergence Time between Sister Species of *Eidolon* (Pteropodidae) with
379 Evidence for Widespread Panmixia. - *Acta Chiropterologica* 16: 279–292.
- 380 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic
381 database.
- 382 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - *Current Biology* 23: R999–R1001.
- 383 Stone, E. et al. 2015. Managing Conflict between Bats and Humans: The Response of Soprano Pipistrelles
384 (*Pipistrellus pygmaeus*) to Exclusion from Roosts in Houses. - *PLoS ONE* 10: e0131825.
- 385 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.
386 - *Science (New York, N.Y.)* 307: 580–584.
- 387 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.
- 388 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in
389 ecology, evolution, and conservation. - *PLOS Biology* 17: e3000494.
- 390 Van Brussel, K. and Holmes, E. C. 2022. Zoonotic disease and virome diversity in bats. - *Current Opinion
391 in Virology* 52: 192–202.
- 392 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,
393 Eds.). - Springer International Publishing.
- 394 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - *Science* 320:
395 70–70.

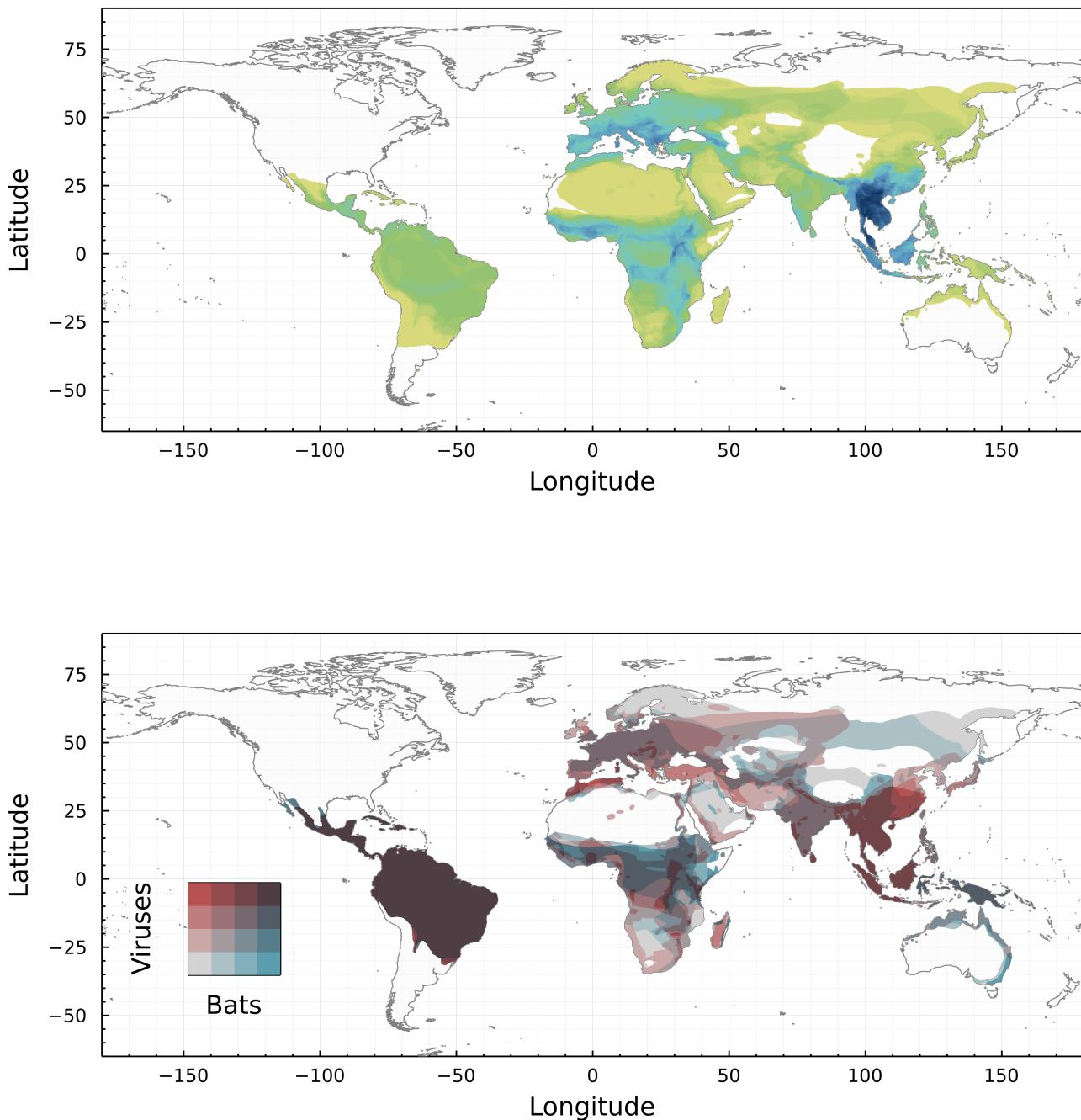


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).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

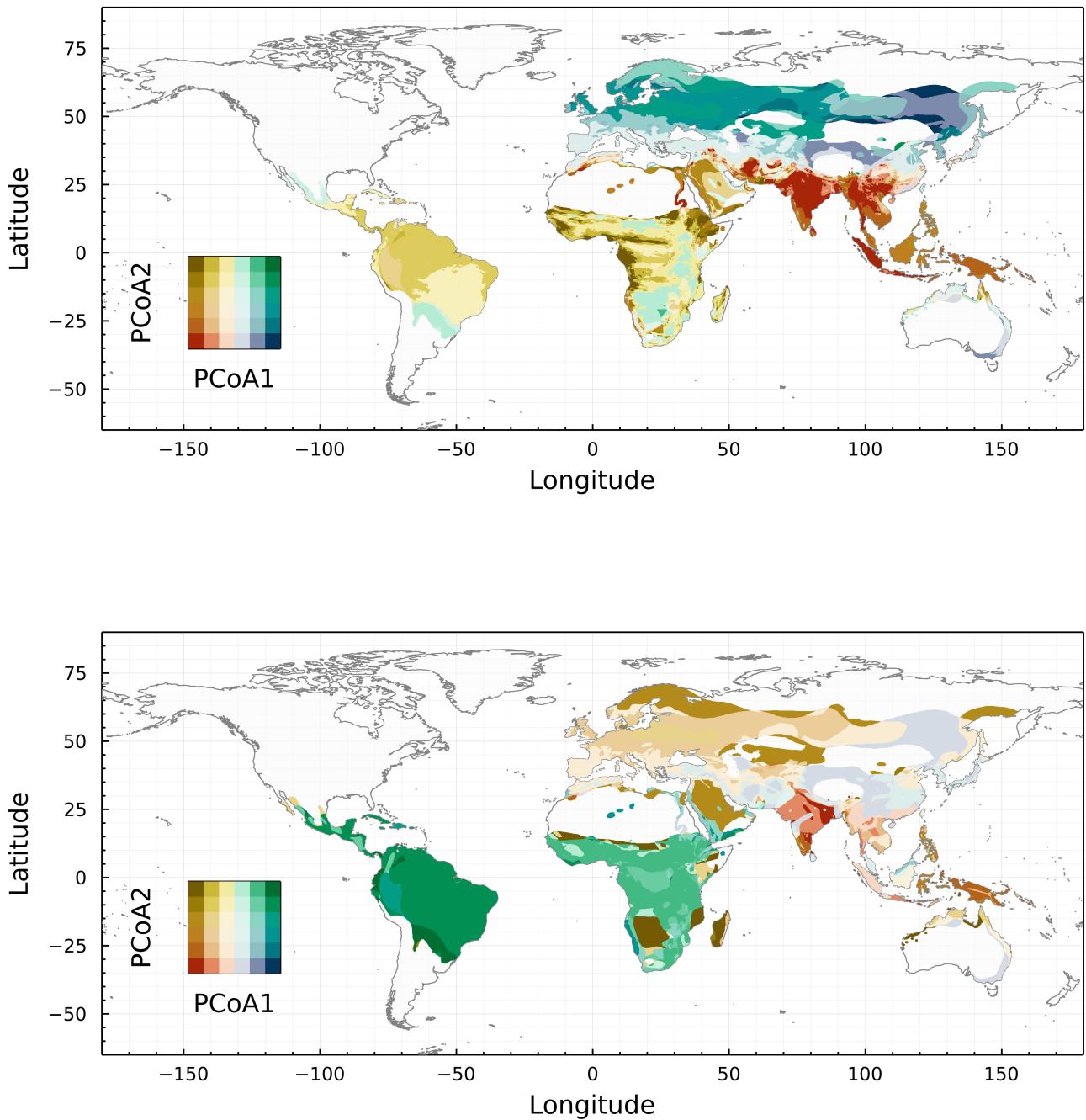


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

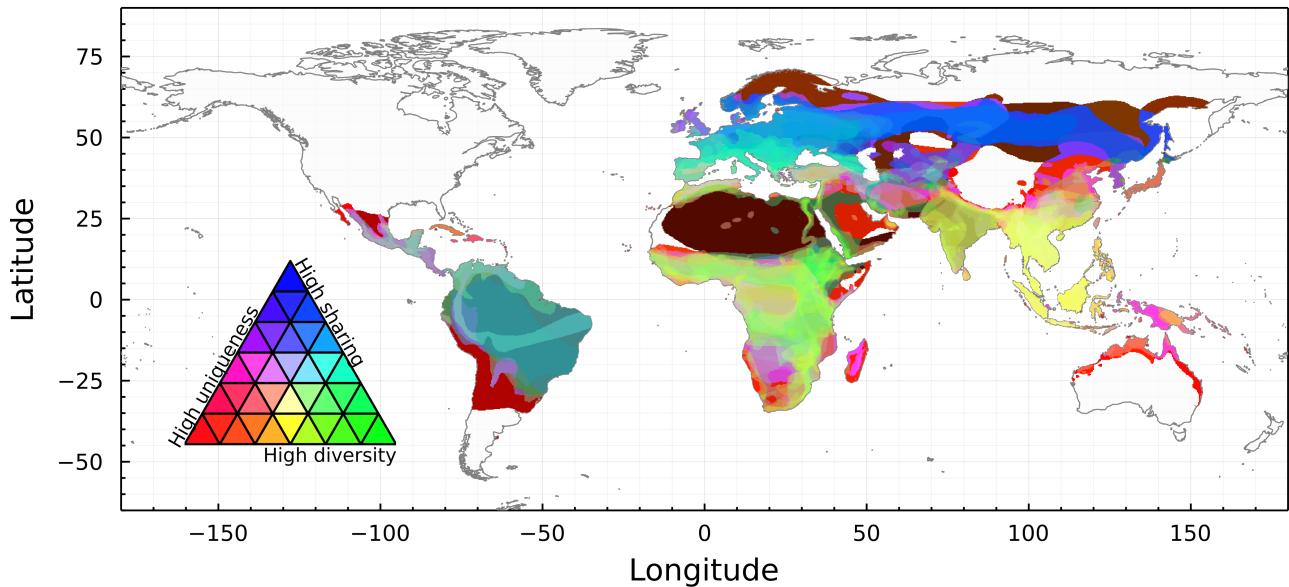


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures would be a pure white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

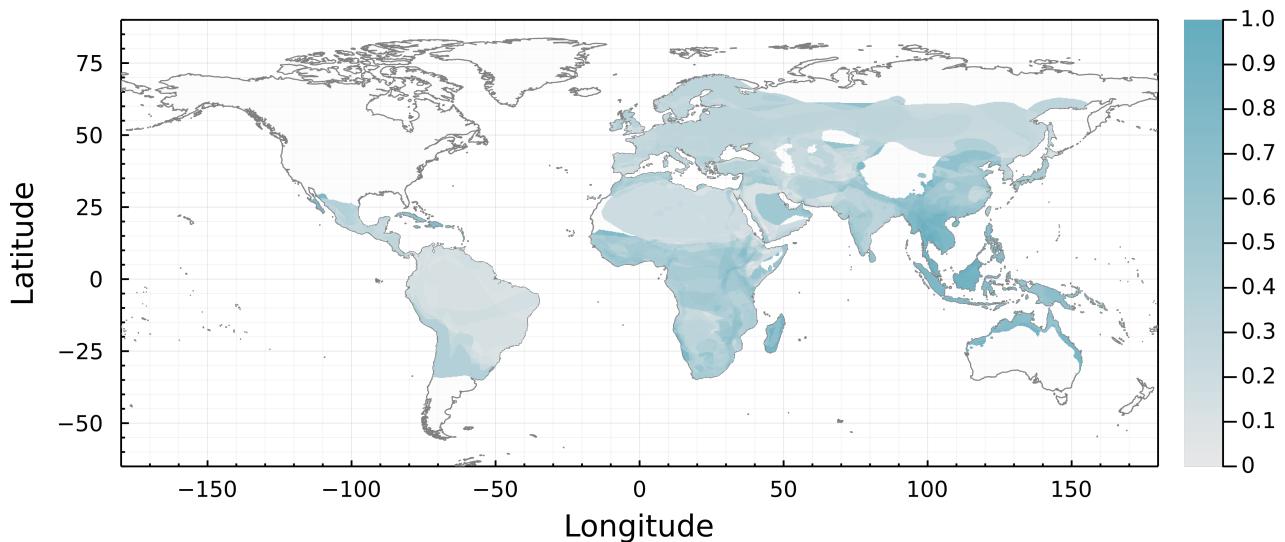


Figure 4: Extraction of a measure of risk based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions at high risk of spillover are southeast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

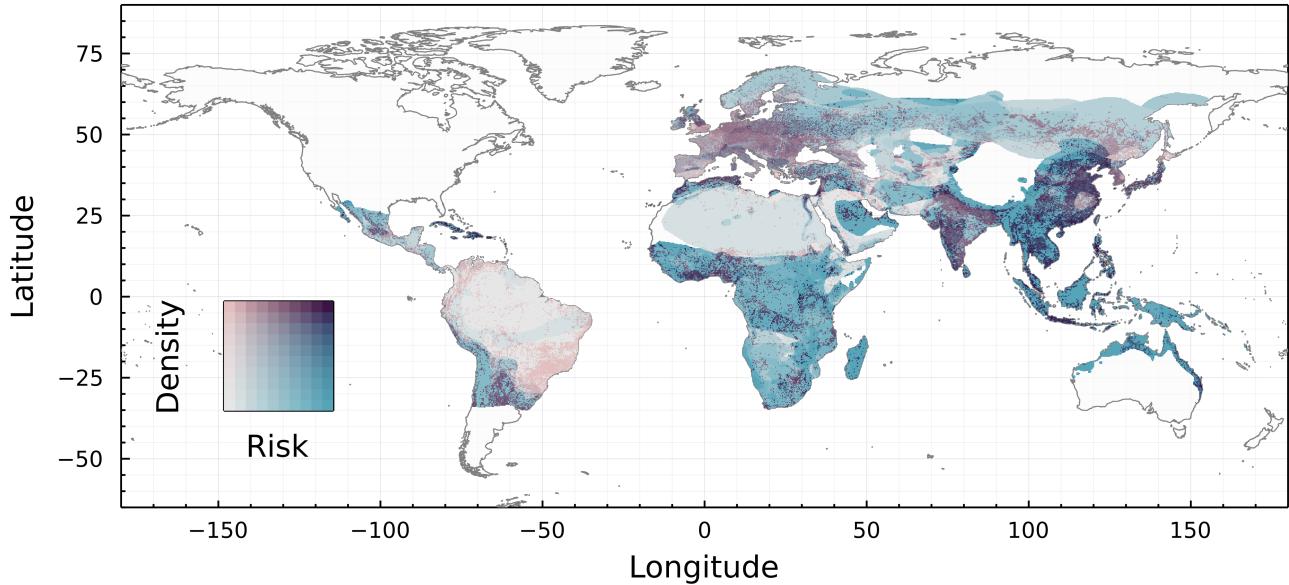


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the risk map. Darker pixels correspond to more risk, in that the GMTC-derived risk of fig. 4 is high *and* the pixel is densely occupied by human populations. This approach increases the relative risk of several regions in Africa, and highlights the risk in India, southeast China, and the arabic peninsula where areas of high to moderate risk overlap with areas of denser population.