

The coevolutionary mosaic of *Betacoronavirus* spillover risk from bat hosts

Norma Forero Rocio Munoz^{1,2} Renata L. Muylaert³ Stephanie N. Seifert⁴ Gregory F. Albery⁵ Colin J. Carlson^{6,7,8} 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, Georgetown University, Washington, DC, ⁷ Center for Global Health Science and Security, Georgetown University Medical Center, Washington, DC, USA ⁸ Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, DC, USA

‡ These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

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

18 In this paper, we examine the biogeographic structure of bats-betacoronaviruses associations, based on a
19 curated dataset of all confirmed bat hosts of betacoronaviruses. By drawing on concepts from the
20 Geographic Mosaic Theory of Coevolution (GMTC; Thompson 2005), we turn these associations into a
21 spatially explicit additive mapping of zoonotic risk components, revealing the extreme heterogeneity of
22 risk at the global scale. Explicitely framing the notion of spillover risk based on propositions from the
23 GMTC (which is to say, based on a framework linking interactions between species to change within
24 species) is a novel idea, that should be relatively general. Indeed, it only assumes the action of well
25 described evolutionary mechanisms. The benefit of this approach is to provide the potential for a more
26 dynamic and nuanced understanding of risk: not only on ecological timescales, but also by providing
27 clues about which areas can change over micro-evolutionary timescales. This provides a way to look at
28 spatial structure by accounting for more notions than species richness/similarity, but also a way to identify
29 spatial areas of higher risk.

30 We identify the Amazon and South-Eastern Asia as hotspots where the phylogenetic distinctiveness of

31 *Betacoronavirus* is the highest (Anthony et al. 2017); surprisingly, current data suggest that viral sharing
32 between hosts is high in the Amazon and low in South-Eastern Asia, which has the potential to result in
33 different evolutionary dynamics between these two regions, hinting at different futures for their viral
34 communities. This work is important both as a description of the bats-betacoronaviruses complex, but
35 also because more broadly, bats are known reservoirs for a variety of emerging viruses and other
36 pathogens (Calisher et al. 2006, Melaun et al. 2014), making balancing the needs for bat conservation and
37 disease prevention most likely very difficult and a source of human-wildlife conflicts, especially in more
38 densely populated areas (Stone et al. 2015, Rego et al. 2015).

39 Methods

40 Known *Betacoronavirus* hosts

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

48 Bats occurrences

49 We downloaded the rangemap of every current bat species that was classified as an empirically
50 documented host of *Betacoronavirus* from the previous step, according to recent IUCN data (IUCN 2021).
51 The range maps were subsequently rasterized using the rasterize function from GDAL (Rouault et al.
52 2022) at a resolution of approximately 100kmx100km. For every pixel in the resulting raster where at least
53 one bat host of *Betacoronavirus* was present, we extract the species pool (list of all competent bat hosts),
54 which was used to calculate the following risk assessment components: bat phylogenetic diversity, bat
55 compositional uniqueness, and predicted viral sharing risk.

56 **Bats phylogenetic diversity**

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

67 **Bats compositional uniqueness**

68 For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres
69 2013); LCBD works from a species-data matrix (traditionally noted as \mathbf{Y}), where species are rows and sites
70 are columns, and a value of 1 indicates occurrence. We extracted the \mathbf{Y} matrix assuming that every pixel
71 represents a unique location, and following best practices (Legendre and Condit 2019) transformed it
72 using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw
73 community data is particularly important for two reasons: first, it prevents the artifact of richer sites
74 having higher importance; second, it removes the effect of overall species richness, which is already
75 incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a
76 community that is on average more dissimilar in species composition than what is expected knowing the
77 entire matrix, i.e. a more unique community. Recent results by Dansereau et al. (2022) shows that LCBD
78 measures are robust with regards to spatial scale, and are therefore applicable at the global scale.

79 **Viral sharing between hosts**

80 For all bat hosts of *Betacoronavirus*, we extracted their predicted viral sharing network, generated from a
81 previously published generalized additive mixed model of virus sharing by a tensor function of
82 phylogenetic distance and geographic range overlap (Albery et al. 2020). This network stores pairwise

83 values of viral community similarity. To project viral sharing values into a single value for every pixel, we
84 averaged the pairwise scores. High values of the average sharing propensity means that this specific extant
85 bat assemblage is likely to be proficient at exchanging viruses.

86 Composite risk map

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

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

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

104 Viral phyogeography and evolutionary diversification

105 To next represent phyogeography of betacoronaviruses in bats, we aggregated and analyzed
106 betacoronavirus sequence data. We used the following query to pull all *Betacoronavirus* sequence data

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

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

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

129 Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the
130 biogeography of their hosts. To test this idea, we loosely adapted a method from (Kreft and Jetz 2007,
131 2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their
132 original method, a distance matrix - where each row or column represents a geographic raster’s grid cell,
133 and the dissimilarity values are the “beta diversity similarity” of their community assemble - undergoes
134 non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected
135 geographically using a four-color bivariate map. Here, we build on this idea with an entirely novel

136 methodology. First, we measure the phylogenetic distance between the different viruses in the
137 betacoronaviruses tree by using the cophenetic function in ape (Paradis and Schliep 2019); subsequently,
138 we take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this
139 case) to project the viral tree into an n-dimensional space. We then take the first two principal
140 components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value
141 and projected them using a four-color bivariate map.

142 Results and discussion

143 Host richness does not predict virus distinctiveness

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

154 [Figure 1 about here.]

155 Nevertheless, locally diverse and virus-rich bat communities could represent an increased risk of spillover
156 under climate change through the creation of novel interactions (Carlson et al. 2022), and therefore the
157 diversity of *Betacoronavirus* strains should similarly be accounted for. In fig. 1 (bottom), we contrast the
158 evolutionary distinctiveness of bats and viruses – this reveals a slightly different portrait than bat richness
159 alone. Chiropterans can be classified, from a macro-evolutionary standpoint, as Yangochiroptera in the
160 “New World,” and Yinpterochiroptera elsewhere (Teeling et al. 2005, Springer 2013). Specifically, we
161 would expect that the so-called “New World” group of bats, being more evolutionary distinct, would also

162 have evolutionary distinct viruses. Indeed fig. 1 (bottom) reveals it to be the case, and this region harbors a
163 distinct bat-betacoronaviruses complex. By contrast, South-Eastern Asia has a lot of non-evolutionary
164 distinct bats, but evolutionary-distinct viruses.

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

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 (Letko et al. 2020,
185 Van Brussel and Holmes 2022). In particular, this should be the case if viruses can circulate among hosts
186 and co-evolve with local hosts communities, making their evolutionary process more than a byproduct of
187 host evolution. High density of hosts sharing the same virus (albeit possibly different strains) can drive or
188 result from evolution of the bat antiviral immune system, resulting in spatially distinct immunological
189 responses, as evidenced in several bat species (Banerjee et al. 2020). Immune characteristics that allow
190 bats to be better adapted to infection by emerging viruses (Gorbunova et al. 2020, Irving et al. 2021), in

addition to being hardcoded in their genome (Jebb et al. 2020), may be related to a wide variety of diets (Banerjee et al. 2020, Moreno Santillán et al. 2021), themselves likely to be driven by spatial effects, especially at the local scale – bats, indeed, occupy a variety of environments, and therefore display a variety of adaptations to these environments (Muylaert et al. 2022).

[Figure 2 about here.]

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

Coevolution-informed spillover risk is different in space

As host richness, joint distinctiveness, or phylogeographic structure suggest that the bat-betacoronaviruses complex is globally fragmented enough to give rise to both different levels of risk (as evidenced by the spatial location of spillover events) and different types of co-evolutionary dynamics, we turn to the Geographic Mosaic Theory of Coevolution (Thompson 2005) to provide a measure of risk accounting for

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

[Figure 3 about here.]

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

[Figure 4 about here.]

From another perspective, areas of high host uniqueness and virus sharing (red-to-pink) could provide

hotspots of *Betacoronavirus* risk through mixing of unique viruses (via codivergence) and in turn recombination. Under our framework, such a hotspot was identified in Madagascar, where most bat species are endemic following evolutionary divergence from sister species in both African and Asian continents (e.g. Shi et al. 2014). Recent surveillance (Kettenburg et al. 2022) has identified a novel *Betacoronavirus* (in the subgenus *Nobecovirus*) in Madagascar-endemic pteropid bat species (*Pteropus rufus*, *Rousettus madagascariensis*), emphasizing strong proof of principle in model predictions.

Human occupancy drives different levels of effective risk globally

Based on the previous result, we extracted the risk component from the composite map (see Methods), to provide a single measure of risk varying between 0 and 1. This measure is presented in fig. 4. As this map represents the potential risk, it must be weighed by the potential for contacts with humans. As a proxy for this measure, we used the proportion of build/urban land from the EarthEnv dataset: this is a reasonable proxy for the density of humans per unit area, which increases the probability of pathogen spread more widely (Hazarie et al. 2021). Since human activity is required to amplify the frequency of virus encounters and thus create areas of viral amplification, mapping the potential risk against measures of land use is required to generate a more actionable assessment of risk. This map is presented in fig. 5. Most of South America and Europe are at comparatively lower risk, as although densely populated, settlements tend to be in areas with lower potential risk. Regions like Malaysia and the North coast of Australia have a high risk component, but should represent a relatively lower effective risk due to low human density. However, this mapping reveals that South-East Asia, the Indian subcontinent, and parts of sub-Saharan Africa, are at high risk due to the overlap between built areas and bat communities representing more opportunities for cross-species transmission of betacoronaviruses. In looking for the origins of SARS in China, Xu et al. (2004) present serological evidence that strongest human-animal contact results in higher risk of virus exposure, regardless of the animal species, but that different types of contact had different impacts. Ideally, finer-grained information about human activity (rather than human presence through anthropisation) could allow to partition this risk further, albeit at the cost of more hypotheses required to estimate the amount of risk represented by each activity. Our map of risk overlays with recent results from Cohen et al. (2022) – areas of purported high risk/diversitification potential (Madagascar, South-America) overlay with sampling gaps for *Betacoronavirus*, stressing the need for spatially targeted monitoring and discovery.

276 **Conclusion**

277 Our study focuses largely on the biogeography of hosts. Yet, we know that viruses with high host
278 plasticity, that is, the ability of a given virus to adapt to various taxonomic orders and ecological groups
279 (Kreuder Johnson et al. 2015), are more likely to amplify viral spillover, followed by secondary
280 human-to-human transmission, and geographical spread (Hazarie et al. 2021). High viral host plasticity is
281 an especially important trait for RNA viruses like betacoronaviruses (Haddad et al. 2021). Indeed, our
282 analysis of viral sequences reveals that Latin America is a hotspot of viral distinctiveness, suggesting that
283 this part of the bats-betacoronaviruses complex may be undergoing independent evolutionary dynamics
284 (related species sharing viruses that are different from the rest of the global pool). The other hotspot of
285 viral distinctiveness is S.E. Asia, in which richness is high but sharing is low; this suggests a different type
286 of evolutionary dynamics (unrelated viruses coevolving with evolutionarily distinct hosts, generating high
287 diversity locally, see e.g. Latinne et al. 2020). Both of these areas should be priority areas for sampling,
288 especially since Becker et al. (2022) advance that they harbor undiscovered hosts of beta-coronaviruses.
289 This diversity of hosts, and the mechanisms by which the exchange of viruses occurs between species, is
290 largely affected by the local environmental conditions and environmental change.

291 Bats are important reservoir hosts for different classes of microorganisms, many of which a threat to
292 human health (Letko et al. 2020, Van Brussel and Holmes 2022). Chiropterans emerged around 64 million
293 years ago and are one of the most diverse mammalian orders, with an estimated richness of more than
294 1400 species (Peixoto et al. 2018, Simmons and Cirranello 2020). They exhibit a broad variety of habitat
295 use, behaviour, and feeding strategies, putting them at key positions in the delivery and provisioning of
296 several ecosystem services, tied to important ecosystem-derived benefits to human (Kasso and
297 Balakrishnan 2013). For example, bats are an essential component of many seed-dispersal networks
298 (Mello et al. 2011). Over two-thirds of bats are known to be either obligate or facultative insectivores,
299 therefore actively contributing for agricultural pest control (Williams-Guillén et al. 2008, Voigt and
300 Kingston 2016), and vectors of pathogens that put a risk on human health (Gonsalves et al. 2013a, b).
301 Because bats are globally distributed and have a long evolutionary history, phylogeographic and
302 biogeographic approaches are required to shed light on the contemporary distribution of coevolutionary

303 processes between bats and the pathogens they host. Not all areas in which bats, viruses, and human are
304 co-occurring are facing a risk of spillover towards human populations, and the areas in which this risk exist
305 may not be facing risks of the same nature and magnitude.

306 There are several factors that drive changes in the diversity of bats (Alves et al. 2018), but human
307 activities' effects on the ecosystem (like modifications of land use) could significantly decrease it.
308 Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their
309 biogeographic variation, and human population density and other anthropogenic factors are decisive
310 moderators for its implications in public health. With the increase of contact between humans and
311 potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al. 2020), as previous
312 studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal
313 (Gryseels et al. 2017). One of these scenarios where interaction between bats and humans can occur can
314 be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats
315 not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and
316 humans (Deshpande et al. 2022). This represents a challenge for conservation strategies and disease
317 ecology since some areas can have both potential for the acquisition of zoonotic viruses and bat-human
318 interactions; in particular, the challenge lies in the fact that actual exposure must then be quantified
319 accounting for several transmission scenarios, including both direct and indirect bat - human interaction.

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

328 **References**

329 Agosta, S. J. et al. 2010. How specialists can be generalists: Resolving the “parasite paradox” and
330 implications for emerging infectious disease. - *Zoologia* (Curitiba) 27: 151–162.

- 331 Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. -
332 Nature Communications 11: 2260.
- 333 Albery, G. F. et al. 2022. Urban-adapted mammal species have more known pathogens. in press.
- 334 Allen, T. et al. 2017. Global hotspots and correlates of emerging zoonotic diseases. - Nature
335 Communications in press.
- 336 Alves, D. M. C. C. et al. 2018. Geographic variation in the relationship between large-scale environmental
337 determinants and bat species richness. - Basic and Applied Ecology 27: 1–8.
- 338 Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - Virus Evolution in press.
- 339 Banerjee, A. et al. 2020. Novel Insights Into Immune Systems of Bats. - Frontiers in Immunology 11: 26.
- 340 Becker, D. J. et al. 2020. Beyond Infection: Integrating Competence into Reservoir Host Prediction. -
341 Trends in Ecology & Evolution 35: 1062–1065.
- 342 Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. -
343 The Lancet Microbe in press.
- 344 Calisher, C. H. et al. 2006. Bats: Important Reservoir Hosts of Emerging Viruses. - Clinical Microbiology
345 Reviews 19: 531–545.
- 346 Carlson, C. J. et al. 2022. Climate change increases cross-species viral transmission risk. - Nature: 1–1.
- 347 Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. - Ecol. Lett.
348 12: 693–715.
- 349 Cohen, L. E. et al. 2022. Sampling strategies and pre-pandemic surveillance gaps for bat coronaviruses.:
350 2022.06.15.496296.
- 351 Dansereau, G. et al. 2022. Evaluating ecological uniqueness over broad spatial extents using species
352 distribution modelling. - Oikos n/a: e09063.
- 353 Deshpande, K. et al. 2022. Forbidden fruits? Ecosystem services from seed dispersal by fruit bats in the
354 context of latent zoonotic risk. - Oikos (Copenhagen, Denmark): oik.08359.
- 355 Drexler, J. F. et al. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of
356 SARS. - Antiviral Research 101: 45–56.

- 357 Engering, A. et al. 2013. Pathogen–host–environment interplay and disease emergence. - Emerging
358 Microbes & Infections 2: e5.
- 359 Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. - Biological Conservation 61: 1–10.
- 360 Gomulkiewicz, R. et al. 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. -
361 The American Naturalist 156: 156–174.
- 362 Gonsalves, L. et al. 2013a. Do mosquitoes influence bat activity in coastal habitats? - Wildlife Research 40:
363 10–24.
- 364 Gonsalves, L. et al. 2013b. Mosquito Consumption by Insectivorous Bats: Does Size Matter? - PLOS ONE
365 8: e77183.
- 366 Gorbunova, V. et al. 2020. The World Goes Bats: Living Longer and Tolerating Viruses. - Cell Metabolism
367 32: 31–43.
- 368 Gryseels, S. et al. 2017. When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure
369 in the Spatial Spread of Arenaviruses (JH Kuhn, Ed.). - PLOS Pathogens 13: e1006073.
- 370 Haddad, D. et al. 2021. SARS-CoV-2: Possible recombination and emergence of potentially more virulent
371 strains (H Attoui, Ed.). - PLOS ONE 16: e0251368.
- 372 Hazarie, S. et al. 2021. Interplay between population density and mobility in determining the spread of
373 epidemics in cities. - Communications Physics 4: 191.
- 374 Hoang, D. T. et al. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. - Molecular Biology
375 and Evolution 35: 518–522.
- 376 Irving, A. T. et al. 2021. Lessons from the host defences of bats, a unique viral reservoir. - Nature 589:
377 363–370.
- 378 Isaac, N. J. B. et al. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.
379 - PLOS ONE 2: e296.
- 380 IUCN 2021. The IUCN Red List of Threatened Species.
- 381 Jebb, D. et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. - Nature 583:
382 578–584.

- 383 Johnson, C. K. et al. 2020. Global shifts in mammalian population trends reveal key predictors of virus
384 spillover risk. - Proceedings of the Royal Society B: Biological Sciences 287: 20192736.
- 385 Kalyaanamoorthy, S. et al. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. -
386 Nature Methods 14: 587–589.
- 387 Kasso, M. and Balakrishnan, M. 2013. Ecological and Economic Importance of Bats (Order Chiroptera). -
388 ISRN Biodiversity 2013: e187415.
- 389 Katoh, K. and Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
390 Improvements in Performance and Usability. - Molecular Biology and Evolution 30: 772–780.
- 391 Keke, S. et al. 2010. Study on skin color image segmentation used by Fuzzy-c-means arithmetic. - 2010
392 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2: 612–615.
- 393 Kettenburg, G. et al. 2022. Full Genome Nobecovirus Sequences From Malagasy Fruit Bats Define a
394 Unique Evolutionary History for This Coronavirus Clade. - Frontiers in Public Health in press.
- 395 Kreft, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. - Proceedings of
396 the National Academy of Sciences 104: 5925–5930.
- 397 Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species
398 distributions. - Journal of Biogeography 37: 2029–2053.
- 399 Kreuder Johnson, C. et al. 2015. Spillover and pandemic properties of zoonotic viruses with high host
400 plasticity. - Scientific Reports 5: 14830.
- 401 Latinne, A. et al. 2020. Origin and cross-species transmission of bat coronaviruses in China. - bioRxiv:
402 The Preprint Server for Biology: 2020.05.31.116061.
- 403 Lednicky, J. A. et al. 2021. Isolation of a Novel Recombinant Canine Coronavirus from a Visitor to Haiti:
404 Further Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. - Clinical
405 Infectious Diseases: An Official Publication of the Infectious Diseases Society of America: ciab924.
- 406 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity
407 coefficients and partitioning (H Morlon, Ed.). - Ecology Letters 16: 951–963.
- 408 Legendre, P. and Condit, R. 2019. Spatial and temporal analysis of beta diversity in the Barro Colorado
409 Island forest dynamics plot, Panama. - Forest Ecosystems 6: 7.

- 410 Letko, M. et al. 2020. Bat-borne virus diversity, spillover and emergence. - *Nature Reviews Microbiology*
411 18: 461–471.
- 412 MacLean, O. A. et al. 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist
413 virus and highly capable human pathogen. - *PLOS Biology* 19: e3001115.
- 414 Melaun, C. et al. 2014. Bats as Potential Reservoir Hosts for Vector-Borne Diseases. - In: Kliment, S. and
415 Mehlhorn, H. (eds), *Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths.*
416 *Parasitology Research Monographs*. Springer, pp. 25–61.
- 417 Mello, M. A. R. et al. 2011. The Missing Part of Seed Dispersal Networks: Structure and Robustness of
418 Bat-Fruit Interactions. - *PLOS ONE* 6: e17395.
- 419 Mollentze, N. and Streicker, D. G. 2020. Viral zoonotic risk is homogenous among taxonomic orders of
420 mammalian and avian reservoir hosts. - *Proceedings of the National Academy of Sciences* 117:
421 9423–9430.
- 422 Moratelli, R. and Calisher, C. H. 2015. Bats and zoonotic viruses: Can we confidently link bats with
423 emerging deadly viruses? - *Memórias do Instituto Oswaldo Cruz* 110: 1–22.
- 424 Moreno Santillán, D. D. et al. 2021. Large-scale genome sampling reveals unique immunity and metabolic
425 adaptations in bats. - *Molecular Ecology*: mec.16027.
- 426 Mull, N. et al. 2022. Virus isolation data improve host predictions for New World rodent
427 orthohantaviruses. - *Journal of Animal Ecology* in press.
- 428 Muylaert, R. L. et al. 2022. Present and future distribution of bat hosts of sarbecoviruses: Implications for
429 conservation and public health. - *Proceedings of the Royal Society B: Biological Sciences* 289:
430 20220397.
- 431 Nguyen, L.-T. et al. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating
432 Maximum-Likelihood Phylogenies. - *Molecular Biology and Evolution* 32: 268–274.
- 433 Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - *Nature* 546:
434 646–650.
- 435 Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary
436 analyses in R. - *Bioinformatics* 35: 526–528.

- 437 Peixoto, F. P. et al. 2018. A synthesis of ecological and evolutionary determinants of bat diversity across
438 spatial scales. - *BMC ecology* 18: 18.
- 439 Plowright, R. K. et al. 2015. Ecological dynamics of emerging bat virus spillover. - *Proceedings of the*
440 *Royal Society B: Biological Sciences* 282: 20142124.
- 441 Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - *Nature Reviews Microbiology* 15: 502–510.
- 442 Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - *The American*
443 *Naturalist* 164: S79–S89.
- 444 Ramshaw, R. E. et al. 2019. A database of geopositioned Middle East Respiratory Syndrome Coronavirus
445 occurrences. - *Scientific Data* 6: 318.
- 446 Rego, K. M. da C. et al. 2015. Assessing human-bat interactions around a protected area in northeastern
447 Brazil. - *Journal of Ethnobiology and Ethnomedicine* 11: 80.
- 448 Rouault, E. et al. 2022. GDAL/OGR Geospatial Data Abstraction software Library. - Zenodo.
- 449 Seekell, D. A. et al. 2018. A geography of lake carbon cycling. - *Limnology and Oceanography Letters* 3:
450 49–56.
- 451 Shi, J. J. et al. 2014. A Deep Divergence Time between Sister Species of Eidolon (Pteropodidae) with
452 Evidence for Widespread Panmixia. - *Acta Chiropterologica* 16: 279–292.
- 453 Simmons, N. B. and Cirranello, A. L. 2020. Bat Species of the World: A taxonomic and geographic
454 database.
- 455 Springer, M. S. 2013. Phylogenetics: Bats United, Microbats Divided. - *Current Biology* 23: R999–R1001.
- 456 Stone, E. et al. 2015. Managing Conflict between Bats and Humans: The Response of Soprano Pipistrelles
457 (*Pipistrellus pygmaeus*) to Exclusion from Roosts in Houses. - *PLoS ONE* 10: e0131825.
- 458 Teeling, E. C. et al. 2005. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record.
459 - *Science (New York, N.Y.)* 307: 580–584.
- 460 Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. - University Of Chicago Press.
- 461 Upham, N. S. et al. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in
462 ecology, evolution, and conservation. - *PLOS Biology* 17: e3000494.

- 463 Van Brussel, K. and Holmes, E. C. 2022. Zoonotic disease and virome diversity in bats. - Current Opinion
464 in Virology 52: 192–202.
- 465 Villalobos, F. and Arita, H. T. 2010. The diversity field of New World leaf-nosed bats (Phyllostomidae). -
466 Global Ecology and Biogeography 19: 200–211.
- 467 Vlasova, A. N. et al. 2022. Animal alphacoronaviruses found in human patients with acute respiratory
468 illness in different countries. - Emerging Microbes & Infections 11: 699–702.
- 469 2016. Bats in the Anthropocene: Conservation of Bats in a Changing World (CC Voigt and T Kingston,
470 Eds.). - Springer International Publishing.
- 471 Wang, N. et al. 2018. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China.
472 - Virologica Sinica 33: 104–107.
- 473 Williams-Guillén, K. et al. 2008. Bats Limit Insects in a Neotropical Agroforestry System. - Science 320:
474 70–70.
- 475 Xu, R.-H. et al. 2004. Epidemiologic Clues to SARS Origin in China. - Emerging Infectious Diseases 10:
476 1030–1037.

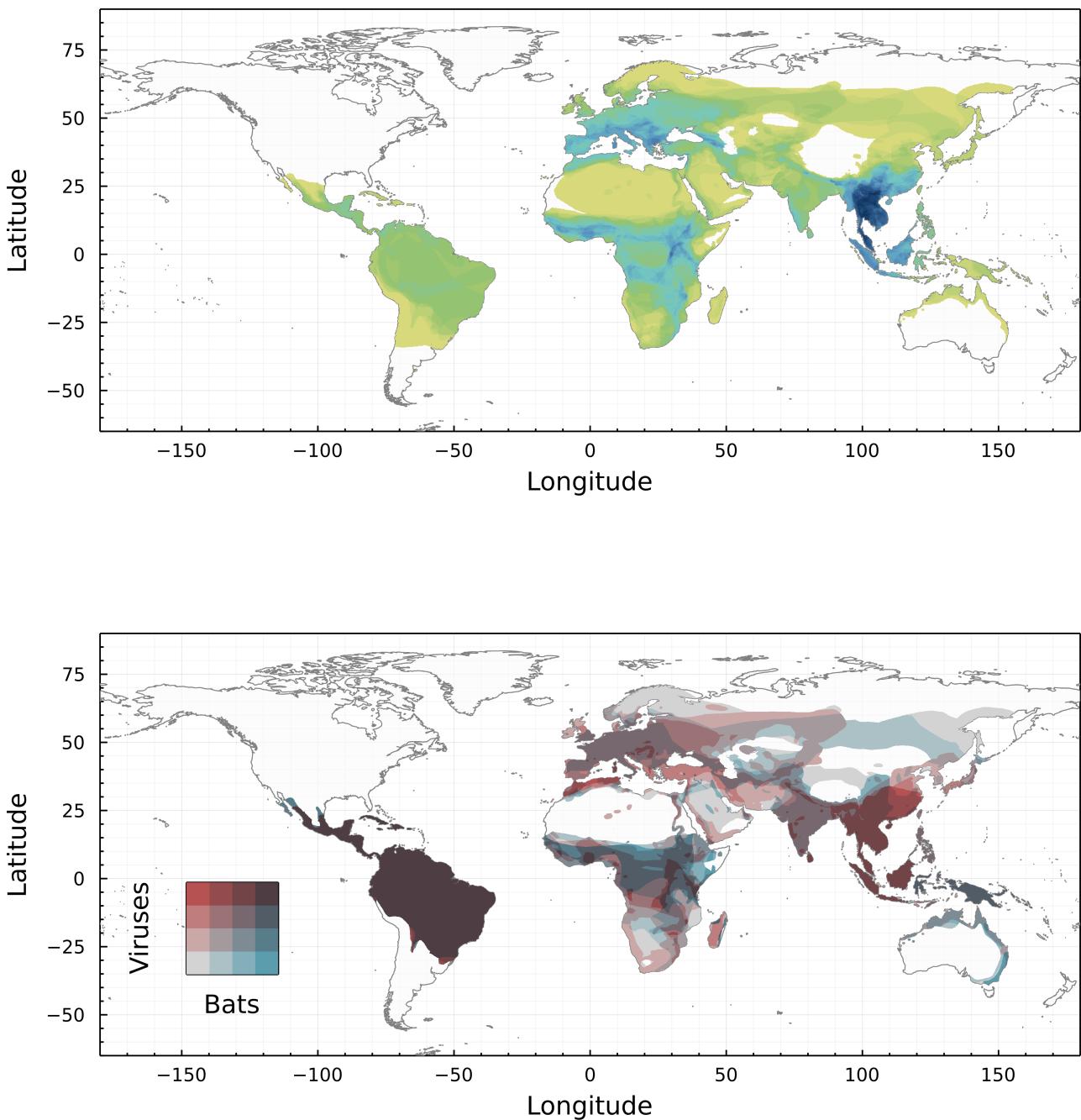


Figure 1: Top panel: relative diversity of known bat hosts of betacoronaviruses. This map shows that the region with the largest number of possible hosts is South-Eastern Asia. Bottom panel: congruence between the evolutionary distinctiveness of the hosts (grey to blue) and the viruses (grey to red).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

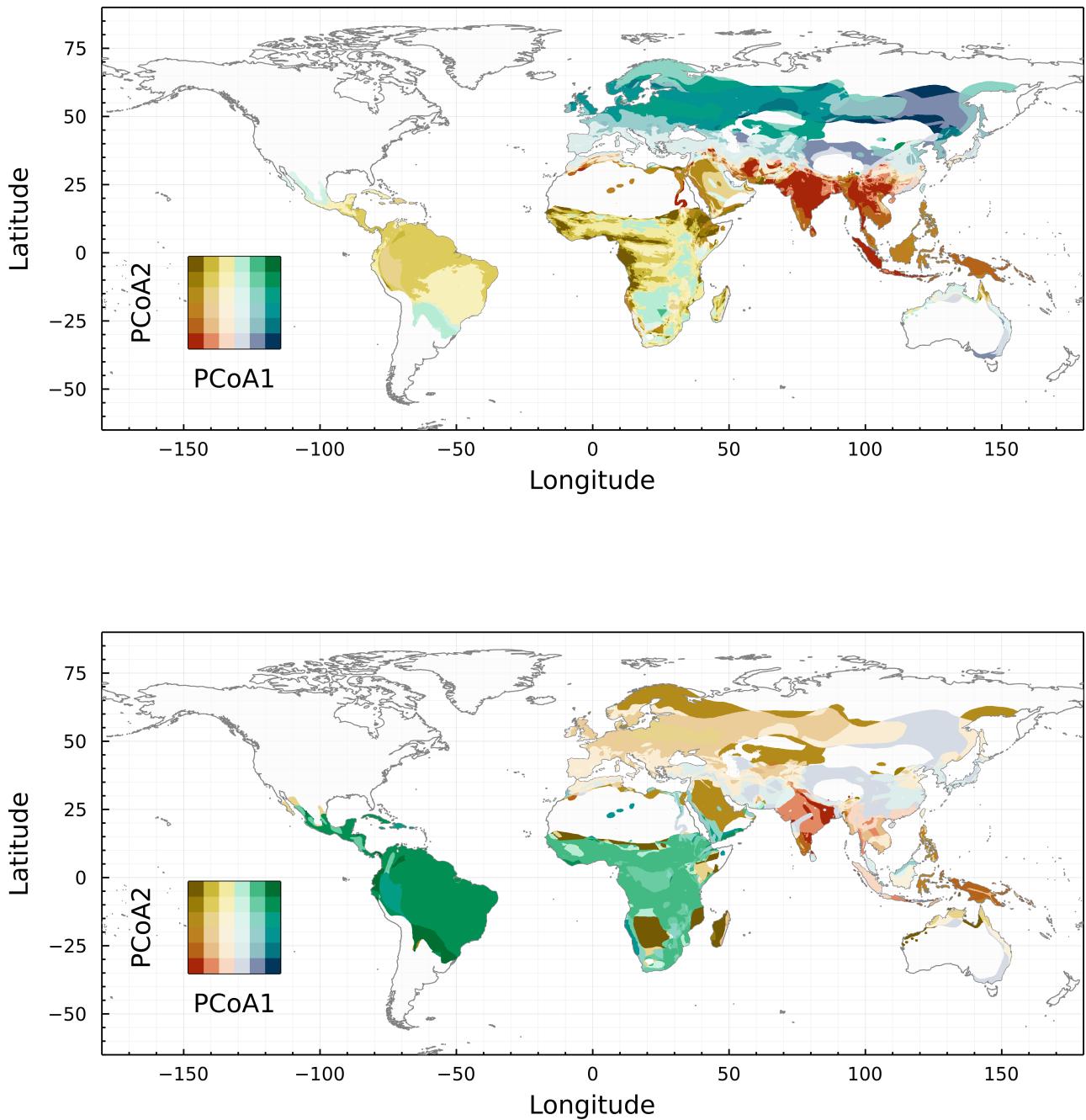


Figure 2: Phylogeographic regions of bats (top) and viruses (bottom) based on the joint analysis of their occurrence and phylogenetic relatedness. The different colors show tendencies to separate alongside the first two components of a PCoA. Note that the PCoA for the bats and viruses are independent,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

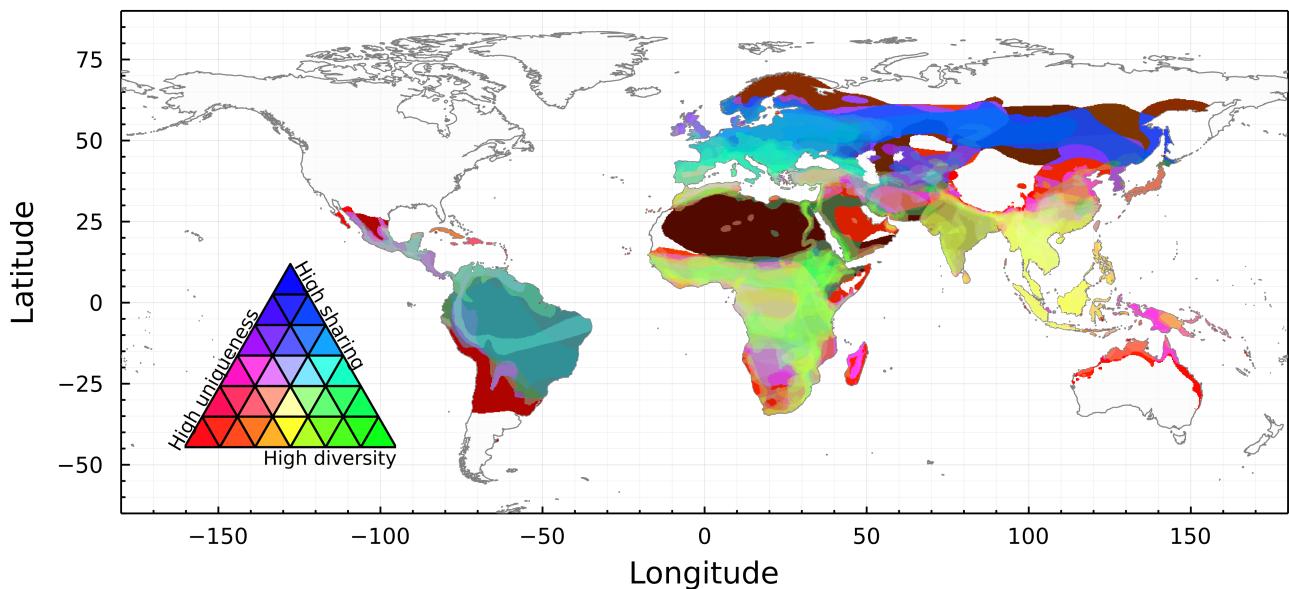


Figure 3: Trivariate additive mapping of the components of risk in the red/green/blue, where high virus sharing is encoded in the blue channel, host phylogenetic diversity in the green channel, and compositional uniqueness in the red channel. A pixel that would maximize all measures (highest possible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

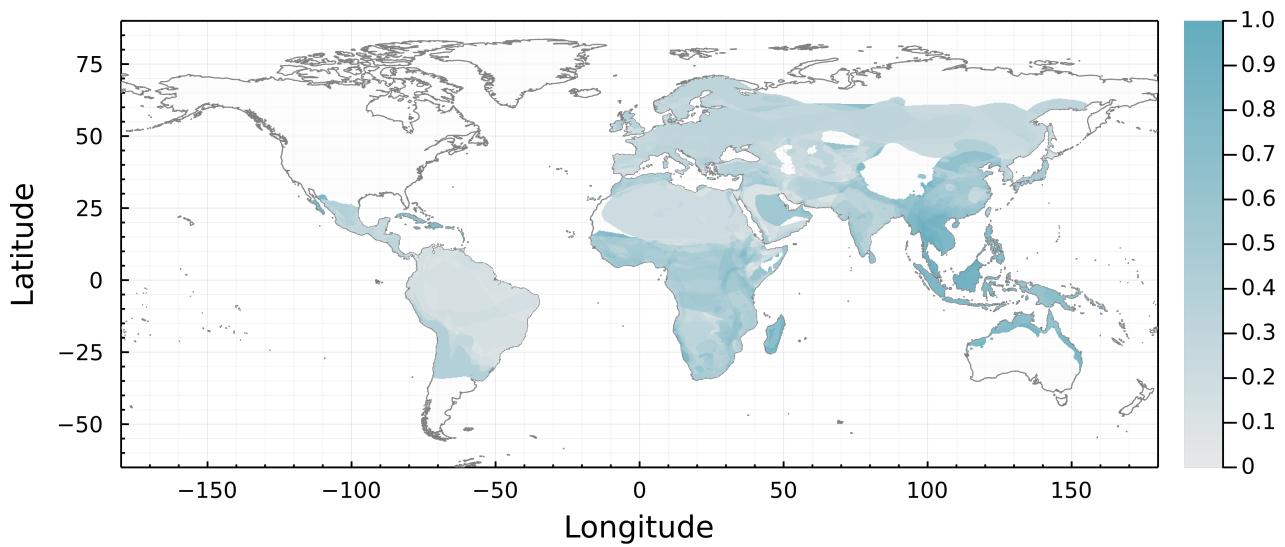


Figure 4: Extraction of a measure of *Betacoronavirus* spillover risk from bat hosts based on the colorimetric space from fig. 3. The risk is a composite measure of the color value and angular distance to the yellow hue, as defined in the methods, ranged in the unit space. Based on this analyses, regions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

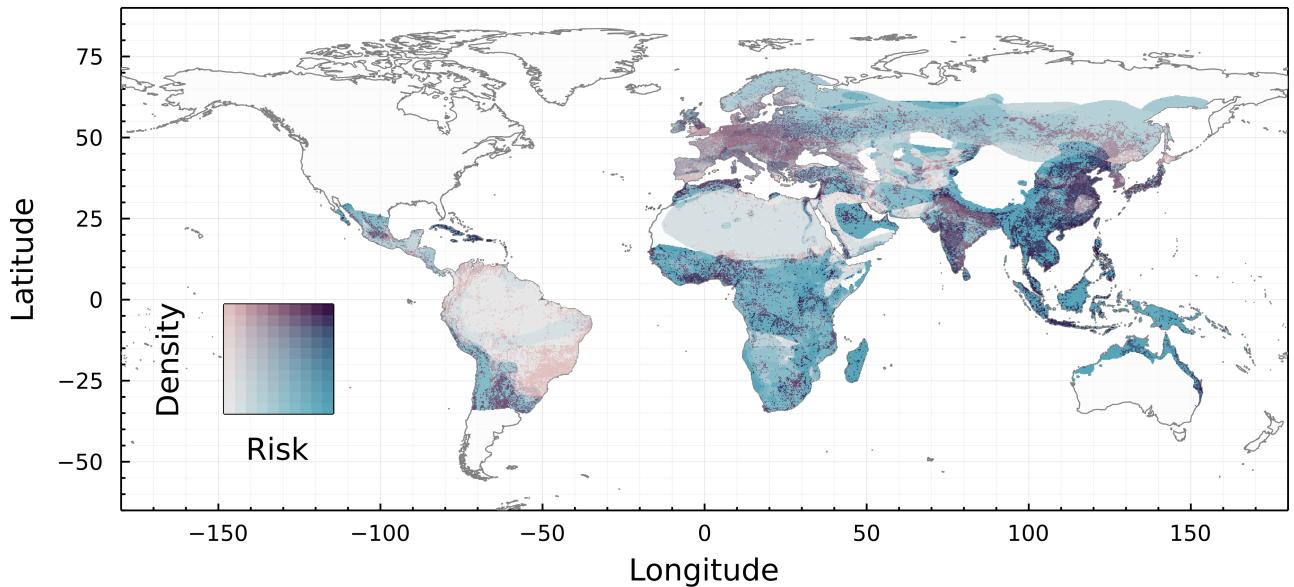


Figure 5: Overlap of the percent of each pixel occupied by urbanized structures, representing the degree of settlement, on the spillover risk map (where the risk comes only from wildlife, and ignores multi-hosts chains of transmissions including non-bats hosts). Darker pixels correspond to more risk,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