The coevolutionary mosaic of bat-betacoronaviruses spillover risk

Timothée Poisot [1,2,‡] Peregrin Took [3,4] Merriadoc Brandybuck [5,4,‡]

[1] Université de Montréal [2] Québec Centre for Biodiversity Sciences [3] Inn of the Prancing Pony

[4] Fellowship of the Ring [5] Green Dragon Inn

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Last revision: *April 6, 2022*

Purpose: This template provides a series of scripts to render a markdown document into an interactive website and a series of PDFs.

Motivation: It makes collaborating on text with GitHub easier, and means that we never need to think about the output.

Internals: GitHub actions and a series of python scritpts. The markdown is handled with pandoc.

Spillover risk is not unidimensional. From the standpoint of an animal community, *i.e.* a pool of suitable hosts, it is driven by a multiplicity of factors (Plowright et al. 2017). The global richness of hosts is one such component commonly mentioned/analysed (see *e.g.* Anthony et al. 2017 for coronaviruses), but there is an argument to be made that species who are not competent (or know) hosts of a specific virus genus may not factor into this (Plowright et al. 2015), or that species who are assumed to share viruses at different rates should be weighted accordingly (Albery et al. 2020). In mammals, key functional traits (for which phylogeny is a reasonable proxy) are determinants of the spillover potential (Olival et al. 2017); these include, notably, body mass and affinity for urban environments (Albery et al. 2022). Finally, especially when the pool of potential hosts spans the entire globe, there may be local host pools that are highly unique; not having been observed in other locations, these can act on the overall risk either by providing novel contact opportunities, reflecting unique host-environment combinations (Engering et al. 2013), or facilitating rapid evolutionary changes in specialism of their pathogens (Agosta et al. 2010). In the specific case of generalist pathogens, there is conceptual and empirical support to the idea that these community- level mechanisms are even more important in driving the overall risk (Power and Mitchell 2004).

Bats are important reservoir hosts for different classes of microorganisms (Chu 2008, Donaldson 2010, Li 2010), some of which can threaten human health. Chiropterans emerged around 64 million years ago and are one of the most diverse mammalian orders, with an estimated richness of more than 12000 species, (Peixoto F et al, 2018) and 14325 known species (Simmons & Cirranello). They exhibit a broad variety of habitat use, behaviour, and feeding strategies, resulting in their playing an essential role in the delivery of several ecosystem services tied to important ecosystem-derived benefits (Kasso 2013). For example, over two-thirds of bats are know to be either obligate or facultative insectivorous mammals, therefore playing an important role in the regulation of insect pests that can affect crops (Williams-Guillen 2011), and vectors of diseases that put a risk on human health (Gonsalves 2013). Because bats are globally distributed and have a long evolutionary history, phylogeographic and biogeographic approaches are required to shed light on the extant distribution of coevolutionary processes between bats and the pathogens they carry. Not all areas in which bats, viruses, and human are co-occuring are facing a risk of spillover towards human populations, and the areas in which this risk exist may not be facing risks of the same nature and magnitude.

In this paper, we examine the biogeographic structure of bats-betacoronaviruses associations, based on a

curated dataset of known and recently discovered hosts. This work is important both as a description of the bats-betacoronavirus complex, but aslo because more broadly, bats are known reservoirs for a variety of emerging viruses (Calisher 2006), making balancing the needs for bat conservation and disease prevention a potentially difficult act and a source of human-wildlife conflicst, especially in more densely populated areas (REF). By drawing on concepts from the Geographic Mosaic Theory of Coevolution (REF), we turn these associations into a spatially explicit additive mapping of zoonotic risk components, which reveals extreme heterogeneity of risk at the global scale; furthermore, we identify the Amazon and South-Eastern Asia as hotspots of phylogenetic distinctiveness of betacoronaviruses; surprisingly, current data suggest that viral sharing between hosts in high in the Amazon and low in South-Eastern Asia, which has the potential to result in different evolutionary dynamics between these two regions.

Methods

Known betacoronavirus hosts

We downloaded the data on bats hosts of betacoronaviruses assembled by Becker et al. (2022) from `https://www.viralemergence.org/betacov` on Aug. 2021, and filtered it to "known" hosts (established before the emergence of SARS-CoV-2) and "novel" hosts (confirmed through sampling since the emergence of SARS-CoV-2). This database was assembled by a combination of data mining and literature surveys.

Bats occurrences

We downloaded the rangemap of every extant bat species that was either classified as an empirically documented or a suspected host of beta-coronaviruses (Becker et al. 2020), according to recent IUCN data (IUCN 2021). The range maps were subsequently rasterized at a resolution of approximately TK TP. For every pixel in the resulting raster where at least one bat host of betacoronavirus was present, we extract the species pool, which was used to calculate the following risk assessment components: phylogenetic diversity, bat compositional uniqueness, and predicted viral sharing risk.

Bats phylogeography

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

Bats compositional uniqueness

For every species pool, we measured its Local Contribution to Beta-Diversity (Legendre and De Cáceres 2013); LCBD works from a species-data matrix (traditionally noted as Y), where species are rows and sites are columns, and a value of 1 indicates occurrence. We extracted the Y matrix assuming that every pixel represents a unique location, and following best practices (Legendre and Condit 2019) transformed it using Hellinger's distance to account for unequal bat richness at different pixels. The correction of raw community data is particularly important for two reasons: first, it prevents the artifact of richer sites having higher importance; second, it removes the effect of overall species richness, which is already incorporated in the phylogenetic diversity component. High values of LCBD indicate that the pixel has a community that is on average more dissimilar in species composition than what is expected knowing the entire matrix, i.e. a more unique community.

Viral sharing between hosts

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

Composite risk map

To visualize the aggregated risk at the global scale, we combine the three individual risk components (phylogenetic diversity, compositional uniqueness, and viral sharing) using an additive color model (Seekell et al. 2018). In this approach, every risk component gets assigned a component in the RGB color model (phylogenetic diversity is green, compositional uniqueness is red, and viral sharing is blue). In order to achieve a valid RGB measure, all components are re-scaled to the [0,1] interval. This additive model conveys both the intensity of the overall risk, but also the nature of the risk as colors diverge towards combinations of values for three risk components.

Viral phylogeography

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

Viral evolutionary diversification

We first tested the hypothesis that hotspots of viral diversification would track hotspots of bat diversification. To do so, we plotted the number of known bat hosts (specifically only those included in the phylogeny, so there was a 1:1 correspondence between data sources) against the "mean evolutionary distinctiveness" of the associated viruses. To calculate this, we derived the fair proportions evolutionary distinctiveness (Isaac et al., 2007) for each of the viruses in the tree, then averaged these at the bat species

level, projected these values onto their geographic distributions, and averaged across every bat found in a given pixel. As such, this can be thought of as a map of the mean evolutionary distinctiveness of the known viral community believed to be associated with a particular subset of bats present.

Co-distribution of hosts and viral hotspots

Subsequently, we tested the hypothesis that the biogeography of bat betacoronaviruses should track the biogeography of their hosts. To test this idea, we loosely adapted a method from Kreft & Jetz (2010), who proposed a phylogenetic method for the delineation of animal biogeographic regions. In their original method, a distance matrix - where each row or column represents a geographic raster's grid cell, and the dissimilarity values are the "beta diversity similarity" of their community assemble - undergoes non-metric multidimensional scaling (NMDS); the first two axes of the NMDS are projected geographically using a four-color bivariate map.

Here, we build on this idea with an entirely novel methodology. First, we measure the phylogenetic distance between the different viruses in the betacoronavirus tree by using the cophenetic function in 'ape'; subsequently, we take a principal components analysis of that distance matrix (readily interchangeable for NMDS in this case) to project the viral tree into an n-dimensional space. We then take the first two principal components and, as with the evolutionary distinctiveness analysis, aggregated these to a mean host value and projected them using a four-color bivariate map.

Outbreaks data geo-referencing

Finally, we provide a summary visualization of what available information describes the spillover of zoonotic betacoronaviruses of bat origin where data was available before and up through the COVID-19 pandemic. The SARS-CoV-2 outbreak was georeferenced to the initial case cluster in Wuhan, China; SARS-CoV was georeferenced based on the cave with the closest known viruses circulating in nature (Hu et al. 2017 PLoS Pathogens), and a nearby location where serological (antibody) evidence has indicated human exposure to SARS-like viruses (Wang et al. 2018 Virologica Sinica). For MERS-CoV, we presented the index cases available from a recently-published compendium of MERS-CoV cases (Ramshaw et al. 2019); these are largely if not all presumed to be camel-to-human transmission, and the precise origin point of MERS-CoV in bats is uncertain. Not shown is a recent case of a recombinant canine coronavirus

that showed the ability to infect humans, both because this study was published after the beginning of the COVID-19 pandemic and because bats' involvement in this cycle of transmission has been marginal to non-existent.

Results

Host distribution

Chiroptera are an hyperdiverse group, distributed in a large part if the world, and are an important reservoir for different strains of betacoronaviruses (Drexler et al., 2014); this has attracted attention to areas where high diversity of bats can be an important issue for human health (Calisher et al., 2006). Accordingly, we collected the IUCN rangemaps for known hosts of betacoronaviruses, to illustrate where hotspots of host diversity are. These results are presented in Fig xx.a. As per our current knowledge of which bats are hosts of betacoronaviruses, these hotspots are primarily South-East Asia, parts of Europe, and to a lesser extent sub-saharan Africa. Even the subset of chiroptera that are hosts of betacoronaviruses fits the evolutionary timeline of the group. Chiropterans can be classified as Microchiroptera and macrochiroptera, where macrochiroptera have an older history from an evolutionary perspective compared to macrochiroptera (Springer, 2013; Teeling et al., 2005). South-East Asia has a high diversity of bats (Kingston, 2010), and our results show that part of that diversity includes betacoronavirus hosts. High density of hosts sharing the same virus (albeit possibly different strains) calls into question the evolution of the bat antiviral immune system and its co-evolution with viruses, which may result in distinct immunological responses in different area, as evidenced in other bat species (Banerjee et al., 2020)

Viral evolutionary distinctiveness

Higher host diversity may not result in a higher viral diversity; for this reason, we quantified and mapped the evolutionary distinctiveness of betacoronaviruses, based on …. Viral evolutionary distinctiveness largely tracks host diversity, particularly in southern China but oddly not throughout the rest of southeast Asia, perhaps indicating that many distinctive viruses remain to be discovered in this region (an idea that is unsurprising given the growing realization, around the emergence of SARS-CoV-2, that a unique lineage of similar viruses are widespread in bats but still mostly undescribed). The most distinct

betacoronaviruses are found in South America, a region with a comparatively lower number of hosts; this suggests that the South American bat-betacoronvirus complex has been more isolated, and is probably undergoing a different co-evolutionary dynamic. Alternatively, this distinctiveness hostpot may be a product of under-sampling: South-America is one of the places where the fewest betacoronaviruses have been discovered (Anthony et al., 2017), and adding more viruses would bring the distinctiveness of known sequences down. Previous work has suggested the Americas may be a hotspot of both undiscovered bat viruses in general (Olival) and coronavirus specifically (Anthony), though not necessarily betacoronaviruses, and particularly not those in clades with notable zoonotic potential (c.f. Anthony).

Geographic Mosaic of bat-betacoronavirus risk

In order to turn the hypotheses based on the Geographic Mosaic Theory of Coevolution into a measure of risk, we overlapped three components: 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, the phylogenetic diversity, i.e. how dispersed the bats in a location are within the tree of life. These results are presented using an additive color mapping in Figure xx, and lead to the definition of broad biogeographic regions of risk, where the same color represents the same type of risk. Pairwise maps of the three components are present in supplementary materials.

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 betacoronavirus 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. xx, 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.

Available data on bat betacoronavirus spillover into humans (TP overlay on the figure) is limited and circumstantial at best for these purposes, but our risk maps suggest that the areas predicted by prior expectations about host biogeography correspond loosely to those where previous emergence events have been recorded. 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 - while not an actual risk factor for spillover itself - likely increases the random chance of a virus with the raw genomic components required for the potential to infect humans.

Global distribution of spillover risk

Based on the previous result, we extracted the yellow component of the risk map (TP add methods), to provide a single measure of risk varying between 0 and 1. This measure is presented in Fig. xxA. However, this maps the potential risk, which 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. xxB. Most of South America and Europe are at low risk, as although densely populated, settlements tend to be in areas with lower potential risk. 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.

Discussion

Driven by the need to understand the ecological factors involved in the emergence of viral pathogens, we spatially mapped bat-betacoronavirus interactions worldwide, using (i) a database of known betacov hosts(Becker et al., 2020), and (ii) range maps for the hosts according to IUCN (IUCN 2021). To reflect the fact that the risk posed by viruses has many ecological origins, we quantified the phylogenetic diversity of hosts, their compositional uniqueness, and the expected viral sharing. Because these components of risk matter when contrasted to human density, we compared them to a proxy, namely the proportion of each pixel that is covered by urban or built land. This provides a synthetic risk map, allowing to identifying of hotspots where the bat-betacoronavirus system may originate viruses in humans. SE Asia is one of the

regions with the highest risk since, according to our results, several of its conditions could increase the risk of transmission of the virus.

Bats are found worldwide and are one of the most diverse groups among mammals (Moratelli & Calisher, 2015). Previous research (Anthony et al., 2017; Mollentze & Streicker, 2020) states that locally diverse bat communities could maintain more viruses and hence, a higher probability of having a pathogen that could represent a risk for human health. This probability involves multiple factors, among which the relatedness of hosts (which can make the jumps easier (Longdon et al., 2011; Mollentze et al., 2020; Wolfe et al., 2007), and the overall tendency of hosts within a locality to share viruses, which may limit viral diversity because of within-host competition (Leeks et al., 2018; Sallinen et al., 2020). Species richness, therefore, is not a sufficient measure of viral risk. This is exemplified in our results, where both South America and South-Eastern Asia have a high species richness of betacov hosts, but only the latter region has a high risk. Specifically, because previous studies propose that Asia is important when it comes to understanding the evolutionary origin of various mammalian taxa (Beard C K, 1988). Including bats (Yu et al., 2014), which could support the relationship between evolutionary time and the development of an immune system with characteristics that allow them to be better adapted to infection by emerging viruses (Gorbunova et al., 2020; Irving et al., 2021) may be related to a wide variety of diets (Jones et al., 2022; Moreno Santillán et al., 2021; Banerjee et al., 2020; Schneeberger et al., 2013).

Our study focuses largely on the biogeography of hosts. Yet, we know that viruses with high host plasticity, that is, the ability of a given virus to adapt to various taxonomic orders and ecological groups (Kreuder Johnson et al., 2015); are more likely to amplify viral spillover, followed by secondary human-to-human transmission, and geographical spread (Hazarie et al., 2021). High viral host plasticity is an especially important trait for RNA viruses such as betacov (Kreuder Johnson et al., 2015; Haddad et al., 2021). Indeed, our analysis of viral sequences reveals that Latin America is a hotspot of viral distinctiveness, suggesting that this part of the bats-betacov system may be undergoing independent evolutionary dynamics (related species sharing viruses that are different from the rest of the global pool). The other hotspot of viral distinctiveness is S.E. Asia, in which richness is high but sharing is low; this suggests a different type of evolutionary dynamics (unrelated viruses coevolving with evolutionarily distinct hosts, generating high diversity locally).

This diversity of hosts and how the exchange of viruses occurs between species, is largely affected by the different environmental changes, as the case of sarbecovirus bats reservoirs (Muylaert et al., 2021) where

they are affected by the area of the cave or the alteration of the forest, which could result in modifications of host distribution. Additionally, our results highlight the importance of Asia as a betacov hotspot, which is consistent with recent studies (Muylaert et al., 2021), where projections on this area suggest that new future events of sarbecovirus viral exchange might be easily spread among species or humans.

There are several factors that drive changes in the diversity of bats (Alves et al., 2018), but human activities' effects on the ecosystem (like modifications of land use) could significantly decrease it. Therefore, it can be suggested that changes in the diversity of betacovs in bats are linked to their biogeographic variation, and human population density and other anthropogenic factors are decisive moderators for its implications in public health. With the increase of contact between humans and potential hosts, we also increase the risk of emergence of novel diseases (Johnson et al., 2020), as previous studies on RNA viruses suggest the importance of host phylogeography at the time of virus dispersal (Gryseels et al., 2017).

One of these scenarios where interaction between bats and humans can occur can be seed dispersal in tropical agroecosystems. It opens the discussion of whether the fruits thrown by bats not only disperse seeds but could also be a source of indirect interaction between viruses of bat origin and humans (Deshpande et al., 2022) . This represents a challenge for conservation strategies and disease ecology since we have areas with potential zoonotic viruses and bat-human interaction. However, it must still be taken into account the quantification of real exposure from several scenarios, where there can be directly or indirectly bat - human interaction.

Comparing scenarios of high viral exchange vs low viral exchange, open the discussion to consider if the best scenario is where viruses easily adapted to multiple hosts but with low virulence or easily ignored by the immune system of the host, or where we have viruses specialized to a specific host, but highly virulent when invade a new host. Accordingly, the understanding of viral-host interactions from a taxonomic and phylogenetic contributes to improving zoonoses surveillance programs.

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

BIOS^2 CREATE program. TP and NF are funded by the Courtois Foundation. RLM was supported by Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation.

References

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

Albery, G. F. et al. 2020. Predicting the global mammalian viral sharing network using phylogeography. - Nature Communications 11: 2260.

Albery, G. F. et al. 2022. Urban-adapted mammal species have more known pathogens. in press.

Anthony, S. J. et al. 2017. Global patterns in coronavirus diversity. - Virus Evolution in press.

Becker, D. J. et al. 2022. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. - The Lancet Microbe in press.

Engering, A. et al. 2013. Pathogen–host–environment interplay and disease emergence. - Emerging Microbes & Infections 2: e5.

Olival, K. J. et al. 2017. Host and viral traits predict zoonotic spillover from mammals. - Nature 546: 646–650.

Plowright, R. K. et al. 2015. Ecological dynamics of emerging bat virus spillover. - Proceedings of the Royal Society B: Biological Sciences 282: 20142124.

Plowright, R. K. et al. 2017. Pathways to zoonotic spillover. - Nature Reviews Microbiology 15: 502–510.

Power, A. G. and Mitchell, C. E. 2004. Pathogen Spillover in Disease Epidemics. - The American Naturalist 164: S79–S89.