**2. Project Description**

**Response to Reviewers**

We appreciate the constructive comments on our re-submitted proposal, and its classification as ‘meritorious’, within the top 20% of proposals submitted. The panel commented that the proposal “does an excellent job of addressing concerns”, on our “serious efforts...to enhance its integrative component”, and that it is an “exciting proposal”. **In this** **2nd resubmission**, we have addressed the panel’s comments by: **1)** publishing papers that directly demonstrate the feasibility of our approach[1-5](#_ENREF_1), a conceptual model linking establishing a metacommunity approach linking phylogenetic diversity and viral persistence and circulation[6](#_ENREF_6), and submitting a paper based on our earlier sampling and discovery at this field site (REF) *[pp. x, x, x]*; **2)** piloting a field study that demonstrates that we will be able to collect sufficient samples from at least 6 bat species and modifying our proposed sampling regime to reflect this *[pp. x, x, x]*; **3)** clearly explaining how we define viral operational taxonomic unit for the purposes of this study, based on ICTV rules (KING REF) and on papers we have published in the last 12 months that use this approach[1](#_ENREF_1),[2](#_ENREF_2) *[pp. x, x, x]*; **4)** demonstrating our capacity to work together as a team by publishing co-authored papers on viral discovery in bats[3](#_ENREF_3),[4](#_ENREF_4),[7](#_ENREF_7),[8](#_ENREF_8) and other taxa[2](#_ENREF_2), on biodiversity theory[6](#_ENREF_6),[9](#_ENREF_9), and on a range of other biodiversity and health issues[10-15](#_ENREF_10) *[See Biographies]*; **5)** providing a detailed explanation of how we deal with biosafety issues *[pp. x, x, x]*; and, **6)** providing detailed training and mentoring plans for undergraduate students, masters students, and postdoctoral researchers, from both the USA and Mexico *[pp. x, x, x]*.

**Introduction**

Viruses likely represent the most abundant source of genetic diversity on the planet. They drive major ecological processes like carbon cycling and shape the evolutionary trajectories of the host taxa they infect[16](#_ENREF_16),[17](#_ENREF_17). Yet despite their critical importance, and more than 100 years of research, only ~2,600 viral species have been recognized by the International Committee for the Taxonomy of Viruses (ICTV)[18](#_ENREF_18). This is ~0.1% of the nearly 2 million species of animals, plants and algae that have been described so far, whereas when properly assessed viruses will certainly outnumber their hosts. The number of described viruses includes only ~600 known from mammals, one third of which are from humans and many others from our domesticated animals. This exposes a clear taxonomic and functional bias in our discovery efforts to date[19](#_ENREF_19), tending towards viruses of public health or economic importance. Indeed, much of what we know of viruses in wildlife has arisen from research, surveillance and control efforts for emerging infectious diseases in humans, the majority of which come from animals[20](#_ENREF_20),[21](#_ENREF_21).

Yet understanding the true extent of global viral diversity has intrinsic ecological value, requiring research that is independent of our anthropocentric motivations. The challenge of systematically measuring the diversity of viruses (here termed *‘virodiversity’*, referring to viral richness and abundance patterns within and among hosts) is complicated by several factors, including the large number of host species (e.g., 5,486 species of mammals alone) and the expense of sampling and viral discovery[22](#_ENREF_22),[23](#_ENREF_23). Until our recently published study[24](#_ENREF_24), no robust estimates of the number of viruses had been achieved for even a single wild mammalian host species, despite the potential utility of such estimates for understanding the roles of viruses in global ecology and public health.

Bats have been widely recognized as important mammalian reservoirs for viruses, and a growing body of disease research has focused on this taxonomic group[13](#_ENREF_13),[25](#_ENREF_25). In a pilot effort to estimate mammalian virodiversity we initiated a survey of *Pteropus* *giganteus* inBangladesh[24](#_ENREF_24) (see Preliminary Data). Our survey suggested that the pool of unknown viruses in mammals is several orders of magnitude greater than the number currently recognized. Nonetheless, it represents only the first step in a challenging program to estimate global virodiversity. **First**, our preliminary extrapolation on the size of the viral pool in mammals[24](#_ENREF_24) was based on a working model wherein we assumed that all host species harbor a similar number of viruses. This assumption is almost certainly inaccurate and requires in depth sampling of additional mammalian host species for refinement and clarification. **Second**, our extrapolation of total viral richness in mammals did not factor in the observation that many viruses infect multiple hosts. **Third**, the use of consensus PCR alone means that only viruses within the targeted viral taxonomic level (e.g., family) are considered and our results dependent on the broad reactivity and sensitivity of these assays[24](#_ENREF_24). Broad scale screening across known viral families must be combined with unbiased high-throughput sequencing (HST) techniques to fully characterize viral richness in a subset of species**. We address these issues in the current proposal.**

The primary aims of this proposal are thus to 1) obtain more robust virodiversity estimates with maximum genetic and taxonomic detail across multiple species in a model group of wildlife hosts, and 2) test specific hypotheses regarding the **functional drivers of virodiversity** patterns, and forecast what we will need to know in the future to better characterize the full dimension of the virosphere. In doing so, we aim to **identify the limits of mammalian virodiversity and begin to unravel the functional, phylogenetic and genetic factors implicit in viral community assemblages among mammalian host species.**

We have developed a cohesive, multi-disciplinary research plan to realize these aims. Our proposed methods are **readily transferrable to other taxonomic and geographic systems to ensure maximum impact of this research**. Specifically, we will: **1) intensively sample a diverse, but geographically restricted, bat community in Mexico that our team has worked with for the last 3 years – El Volcán de los Murciélagos ("The Volcano of Bats").** This cave system makes taxonomic sampling manageable, adequate sample sizes possible, and eliminates many confounding factors. We have determined that it is not feasible given budget constraints to simultaneously and rigorously characterize both geographic beta diversity (change in viral communities through space) *and* inter-specific beta diversity (change in viral communities among species) – our study design prioritizes the latter by controlling for the former, with the advantage that we will also be able to characterize temporal beta diversity (change in viral communities through time)**; 2)** employ state-of-the-art laboratory methods to **characterize and quantify the genetic and taxonomic dimensions of known and novel viruses;** and **3)** use ecological and phylogenetic analytical approaches to **explore the functional drivers of patterns of viral diversity within and among host species (functional alpha and inter-species beta diversity).** We will examine the following primary aims and hypotheses with this study system:

**AIM 1. Establish whether virodiversity patterns vary among mammalian host species (descriptive).**

With the exception of our preliminary virodiversity study[24](#_ENREF_24) (see Preliminary Data) virtually nothing is known about the potential viral richness ( = number of viral operational taxonomic units, or ‘species’) or diversity (functional alpha diversity = richness~abundance relationship)in *any* wild mammal species. Thus it remains unknown whether viral richness varies dramatically among host species or how viral communities change among host species (inter-specific beta diversity). This fundamentally limits our ability to evaluate the genetic, taxonomic and functional depth of the mammalian viral pool.

For Aim 1, we will fill this gap by detecting known, discovering novel, and estimating the number of detectable (but undetected) viruses in several bat species across multiple viral families. From this dataset, we will be able to 1) consistently characterize the viral community structure (including its genetic and taxonomic dimensions, and functional alpha diversity i.e., viral richness, and richness~abundance relationships as expressed using richness-independent diversity indices commonly employed in biodiversity studies)[26](#_ENREF_26) in each host species, which will be a first for any group of wildlife hosts; 2) accurately describe patterns of viral co-occurrence (positive or negative) within samples, individuals and species; and 3) accurately describe patterns of inter-specific beta diversity (viral assemblage similarity among hosts, both standard and phylogenetic, with compositional and phylogenetic dissimilarity indices[26-30](#_ENREF_26)). We employ simple null hypotheses to test for departure from equal or random patterns in intra- and inter-specific viral community structure (see below). These data, which describe the genetic and taxonomic dimensions of viruses in hosts, will be the foundation for Aim 2, which focuses on the functional drivers of varying patterns of virodiversity within and among host species (see below).

**Hyp 1: ‘Host-specific viral richness hypothesis’: Viral richness (number of species) varies among host species.** The null hypothesis is that all species host a similar number of viruses, as assumed in several previous studies[20](#_ENREF_20),[24](#_ENREF_24).

**Hyp 2**: **“Host-specific viral diversity hypothesis”: Viral diversity (functional alpha diversity) varies among host species.** The null hypothesis is that viral diversity is uniform among host species.

**Hyp 3. “Viral community structure hypothesis”: Viral assemblages within species are non-randomly structured.** The null hypothesis is random viral assembly. We will test for significant departures from random co-occurrences (both positive and negative) that might indicate competitive exclusion or viral clustering and other structuring processes.

**Hyp 4: “Shared viral assemblage hypothesis”: Host species share virus species.** The null hypothesis is that each host has a completely unique viral assemblage (i.e., similarity measures = 0). This forms the basis of quantifying inter-species beta diversity, the functional drivers of which are explored in Aim 2.

**AIM 2: Understand the drivers of genetic, taxonomic and functional virodiversity among mammalian host species (predictive).**

Aim 1 will provide the baseline data needed for Aim 2. In addition to the description of the virodiversity of each host species and quantification of the differences between species, our sampling design is sufficient to probe which factors contribute to these differences, using a range of analytical tools (see Methods)*.* Differences in viral assemblages between pairs of host species will first be quantified (Aim 1, Hyp 4) with compositional and phylogenetic dissimilarity indices[26-30](#_ENREF_26). These indices will then be used in Aim 2 as **response variables** in our analyses (see Methods). As **explanatory variables**, we will explore the potential roles of host **phylogeny** (see Hyp 5) and functional ***intrinsic***host and viral traits as well as ***extrinsic*** factors. Functional **intrinsic** host traits refer to fixed characteristics of species (e.g., ecological and key life-history traits) (see Hyp 6). Functional **intrinsic** viral traits refer to fixed characteristics of the viruses we detect (see Hyp 8). **Extrinsic** factors are external to the host and viral species themselves and here include only metrics of inter-specific ‘contact potential’ (Hyp 7) (since most other variation in external factors are purposefully controlled for in our sampling design). In Aim 2, we will test the following specific hypotheses:

**Hyp 5: ‘Phylogenetic proximity hypothesis’: Viral community turnover is a function of phylogenetic distance among host species.** Previous work suggests that closely related host species are more likely to share viruses. For example, phylogenetic distance is a predictor of pathogen sharing for: plants[31](#_ENREF_31); amphibian viruses[32](#_ENREF_32); fish macroparasites[33](#_ENREF_33" \o "Poulin, 2010 #15); primate parasites[34](#_ENREF_34),[35](#_ENREF_35); bat rabies[36](#_ENREF_36),[37](#_ENREF_37); carnivores[38](#_ENREF_38); and viral sharing in bats (see Preliminary Data). However, while phylogeny may be good for prediction, it may not be *mechanistically* informative. Rather, it may just be a *proxy* for the variation (or similarity) in evolutionarily derived traits across species that are more mechanistically influential in driving virodiversity patterns (e.g., more closely related species may share similar immune system characteristics, conserved host cell receptors[39](#_ENREF_39), cause phylogenetic clustering across different spatial scales[40](#_ENREF_40), and/or evolutionary vicariance whereby viruses are maintained from a common ancestor). Shared evolutionary history may thus be a confounding factor that first needs to be detected in our study, and then controlled for if present in order to parse out the effects of other explanatory predictors[41](#_ENREF_41). We previously published[42](#_ENREF_42) a Bayesian phylogeny for 58% of the world’s bats (n = 648 species) based on the CytB gene, which is the most widely available marker for most mammals and shown to be a particularly reliable phylogenetic marker[43](#_ENREF_43). We will update and produce a more robust phylogenetic reconstruction using multiple individuals of each species and a combination of mtDNA (cytB, CO1) and nDNA (RAG2 and various introns) markers during the course of this study. The updated phylogenetic tree will be used to calculate phylogenetic distances among species for use in our analyses (see Methods).

**Hyp 6: ‘Functional intrinsic host trait hypothesis’: Viral community turnover is a function of intrinsic ecological and/or life-history trait similarity among host species**. Chiropterans have a very wide-range of functional **intrinsic** ecological, morphological, and life-history traits that could be important determinants of their viral assemblages[44-47](#_ENREF_44). To explore the **functional** dimension of virodiversity in our model system, we have identified a range of key **intrinsic functional** traits in our host species. We will use these to test the hypothesis that species with similar life-history or ecological traits host viral assemblages that are more similar in composition or structure after controlling for phylogenetic non-independence (see Hyp 5). The selection of each host trait is linked to a specific sub-hypothesis that suggests a potential **functional** link with virodiversity. For example, body size could influence a species’ viral diversity by representing varying ‘patch sizes’ for viruses – larger bodied species may, for example, be more capable of hosting more viral species or higher viral abundance or sustain more stable populations of viruses[48](#_ENREF_48), consistent with the species-area relationship[49](#_ENREF_49), theory of island biogeography[50](#_ENREF_50), and critical community sizemodels[51](#_ENREF_51),[52](#_ENREF_52). Similarly, host population size, rather than host body size, may determine total ‘patch size’ for viruses and be linked to viral genetic diversity[53](#_ENREF_53). Other traits showing variation across our host assemblage that we will investigate (and a brief, referenced justification for its inclusion) include host population density (important epidemiological factor in models of disease dynamics)[54](#_ENREF_54), basal metabolic rate (metabolism~immunity trade-off)[55](#_ENREF_55), dietary niche (e.g., frugivore vs nectivore vs insectivore; food may be a source for viral transmission)[56](#_ENREF_56), ecological niche (e.g., cave vs non-cave, roost or foraging sites may be source for viral transmission)[57](#_ENREF_57), geographic range size (exposure to larger viral pool)[58](#_ENREF_58), wing morphology (dispersal ability, inter-species transmission, long-distance transmission, foraging area, exposure to larger viral pool)[36](#_ENREF_36),[59](#_ENREF_59), and longevity (varying temporal opportunity to accumulate viruses and for their coevolution)[59](#_ENREF_59). We have begun developing a database of functional trait characteristics for each species from published literature and from our ongoing work in this region for use in our analyses (see Methods).

**Hyp 7: ‘Extrinsic trait hypothesis’: Viral richness and community turnover is a function of extrinsic factors (e.g., environmental conditions, inter-specific contact).** An alternative to the intrinsic host trait hypothesis is that some **extrinsic factor(s)** are more important for explaining variation in virodiversity among host species[60](#_ENREF_60). Our sampling design has intentionally eliminated variation in many extrinsic factors (e.g., geography and environmental conditions, since all of our species are sampled in the same place (at the cave) and are effectively experiencing the same or similar environments) that we might hypothesize to be related to virodiversity. However, several sources of **extrinsic** variation remain. For example, inter-specific host contact may be independent of phylogeny, behavior or level of sociality and be variable depending on specific circumstances (rather than a trait of the species itself). This could strongly affect the opportunity for viral sharing among species. Previous studies have shown that overlapping distributions among host species, a likely proxy for contact potential, is a significant determinant of pathogen sharing in primates and carnivores[34](#_ENREF_34),[61](#_ENREF_61). We will test for a potential effect of inter-specific contact in our study at **two distinct spatial scales**: At a **macro-geographic scale** we have created a ‘potential contact’ metric for each host species based on the number of other mammalian hosts that are present within its geographic distribution. At this scale, the contact hypothesis would suggest that the more species occurring within the range of the target species, the greater the opportunity for accumulating viruses. Hence, the difference in the *number* of species that each target species overlaps, in addition to the *similarity* of those species (i.e., as measured by similarity metrics, such as Horn index of overlap), may be related to the *differences* we observe in the viral assemblages among species. At a **micro-geographic scale** (i.e., at our study cave), inter-specific **spatial proximity in roosting location** may also facilitate viral sharing between species. Although pairwise proximity between species may ultimately be governed by the intrinsic traits of each species (e.g., roosting behaviour, thermal preferences and subsequent distribution within the cave, preferred roosting surfaces, level of sociality etc.) we will derive a general ‘proximity’ metric between each pair of species at the cave as another extrinsic explanatory variable. Species pairs that include one of the two species that do not roost in the cave will have proximity scores of 0 (i.e., do not roost together), while species pairs that include only the cave roosting species will have higher proximity scores dependent on their association within the cave (e.g., species pairs with communal roosts will tend towards 1, whereas species occupying different niches within the cave will have intermediate scores).

**Hyp 8: ‘Functional intrinsic viral trait hypothesis’: Virodiversity is explained by functional intrinsic viral traits.** Our preliminary analysis of the literature and our preliminary data from bats in Mexico and Bangladesh suggest that some viral families are much more common and diverse than others, and that there are observed differences in host breadth among different viruses. Fundamental questions remain as to how viral traits might also influence both the degree of diversification and potential host range[37](#_ENREF_37),[62](#_ENREF_62), contributing to differences in virodiversity patterns among hosts. In our preliminary analyses, we found that herpes-, adeno-, paramyxo- and astroviruses were all highly diverse, while corona- and polyomaviruses appeared to be less so. We hypothesize that some viral families will be more predictably ‘biodiverse’ and/or prevalent within and among hosts than others in the proposed study, and that this is an important **functional** **driver** of virodiversity patterns. **Functional** **intrinsic viral traits** of interest for their potential to influence broad patterns of virodiversity include genome type (RNA/DNA; segmented/non-segmented), site of replication (nuclear, cytoplasmic), structure (capsid arrangement, presence/absence of an envelope), and transmission strategy. Target viral families were selected to capture this variability in intrinsic virus traits, in addition to those that will likely vary in prevalence, strain diversity, and the potential or propensity for either intra or interfamily co-occurrence. Viral family-level phylogenies will be used to quantify virus diversity at various scales, within and across host taxa.

We believe this study has the potential to shape the future direction of the field by generating new testable hypotheses that explore specific mechanisms behind the **functional** **associations** we identify. For example, in the event we find a significant host association for a particular virus, we might hypothesize that this host restriction is driven by the exclusive expression of a specific cell-entry receptor in that host. Equally, if we were to observe that significant patterns of co-occurrence correlated with increasing genetic distance we might hypothesize that this is driven by strong immune recognition in the host. These examples represent just two of many possible hypotheses that could be tested in each case, but both are also examples of how this study could lead to targeted future research that continues to refine our understanding of the factors that control viral diversity and community structure.

**Preliminary Data**

**Viral discovery in Mexico (pilot study)**

We collected non-invasive samples from 606 bats of 42 different species. These individuals were screened by consensus PCR (cPCR) for the presence of coronaviruses, and **8 novel viruses were identified**. Four were α-CoVs, and four were β-CoVs. Two of the eight CoVs were discovered in the ‘Volcano of bats’ (our proposed study site), and one of these was shown to be a β-CoV with 96.5% amino acid identity to the emerging MERS-CoV associated with human disease in the Middle East[63](#_ENREF_63" \o "Zaki, 2012 #38). These data have now been published in the Journal of General Virology (REF).

Using these 8 viruses (plus other published CoV sequences from Latin America), we then performed a preliminary study to look for evidence of phylogenetic congruence with particular host species, as host range is believed to be an important risk factor for zoonotic emergence. Viruses that can infect multiple hosts are considered more likely to emerge as pathogens, compared with those that are restricted to a single host species or to a group of closely related hosts. We tested for congruence between α-CoVs and β-CoVs and their bat hosts using global co-speciation tests, ParaFit and PACo and found evidence for significant congruence between host and viral phylogenies (ParaFitGlobal = 48.32226, P≤0.001; m2XY = 2.016884, P≤0.0001), suggesting that some degree of co-speciation has occurred in this region (Author et al., in review). Similar studies in China did not show this same pattern, suggesting that more host-switching is occurring in Asia, compared with Latin America. We suggest therefore that there may be geographical variation in the degree of host specificity for CoVs, which may subsequently reflect regional differences in pathogen emergence risk. **The current proposal will build on these preliminary data, by looking at whether host-restrictions are maintained in co-roosting species (i.e., the degree to which proximity might lead to virus sharing), and whether there are specific host traits that are associated with infections of particular viruses**.

**Mammal-virus host range and viral discovery from the literature**

We created an extensive database from the last 70 years of peer-reviewed literature to explore patterns of viral richness and viral sharing among mammals. The database includes over 2700 mammal-virus associations for 770 unique mammal species and ~600 ICTV recognized viruses. Preliminary analyses of this dataset found that contact potential, as represented by geographic overlap (z = 3.210, p<0.001), phylogenetic distance between host species (z = -15.676, p<0.001), viral transmission strategy (z = -7.072, p<0.001, vector-borne) and viral genome type (z = -3.762, p<0.001, segmented) However, we also observed that the number of known viruses from any given host species is strongly dependent on research effort as indicated by the number of publications per host (z = 6.528, p<0.001) and virus (z = 6.161, p<0.001). This strong research bias clearly and inappropriately defines what we currently know about the mammalian virosphere. We also discovered that methodologies for viral discovery vary considerably across studies, limiting our ability to draw general conclusions about patterns of virodiversity from the literature. Our analyses of these data have strongly shaped the specific hypotheses and the methods proposed here (specifically Hyp 5, Hyp 7 and Hyp 8). We overcome the primary limitations of the literature study here by designing a standardized approach that includes both *depth* and *consistency* in viral discovery as well as *breadth* across multiple viral and host taxonomic groups. .

Virodiversity **in *Pteropus giganteus***

**Fig. 1.** Non-parametric viral richness estimates for *P. giganteus*. Red line = collector curve; Black line = accumulation curve; Blue line = Chao2 richness estimator. The number of samples (vertical dashed lines) needed to detect given % of total viral richness estimated is shown.

Our pilot study describing the viral diversity of Mexican bats (presented above) was limited to a single viral family, but targeted a large number of host species. In a separate study we used a modified approach by limiting the number of host species to one (*Pteropus giganteus* from Bangladesh), but expanding both the number of samples (n=1700) and the number of viral families screened for (n=9)[24](#_ENREF_24" \o "Anthony, 2013 #97). By combining virological and ecological techniques, we asked whether it was possible to saturate the discovery of viruses in this single species in order to estimate just how many viruses mammalian hosts can sustain. Using cPCR, we identified 5 known and 50 novel viruses from nine viral families (total n=55 viruses), including coronaviruses (n=4), paramyxoviruses (n=11), astroviruses (n=8), bocaviruses (n=2), adenoviruses (n=14), herpesviruses (n=13), polyomaviruses (n=3) hantavirus (n=0) and influenza virus (n=0). We then used **v** toe the, (Fig 1). provided a consistent estimate of total viral richnessaestimate given the diminishing returns on viral discovery our approach to additional sample types, by expanding our cPCR testing to include additional iesand with the additionDespite these limitations, this study provided the first ever estimate of **viral richness** in a wildlife species and the sampling effort required to detect any proportion of it, a **phylogenetic description** of 55 viruses (50 of which were previously unknown), and a **functional ecological description of all viral co-occurrences that were observed in this system**.

**T**he identification of co-existing microbes is important to a description of virodiversity because of the positive and negative associations that can occur between them[66-71](#_ENREF_66). In our virodiversity study[24](#_ENREF_24), we reported a large number of intra and inter-familial co-occurrences in *P. giganteus* and showed that as many as five different viruses can exist in a single sample. This revealed information about the carrying capacity and composition of discrete viral niches within an individual bat, and also the number of different viruses that could potentially spill over to a new host from a single exposure event. *S*ignificant intraspecific co-detections were identified in the Herpesviridae (HV) and Adenoviridae(AdV) families. Patterns of HV co-occurrence were non-random (*p*<0.001 with C-score[72](#_ENREF_72),[73](#_ENREF_73)), and positive pairwise associations were observed between certain virus pairs, particularly within the HVs. It is unknown why these specific viruses should so readily co-exist, though ecological mechanisms such as simultaneous transmission (co-dispersal), the availability of requisite resources and/or shared benefits associated with host-immunomodulation by one or more of these viruses may explain the observed co-occurrence. Recombination is a common feature in the ecology and evolution of HVs and can present a beneficial consequence of co-infection[74-80](#_ENREF_74). We detected three HVs in the same sample type (throat), multiple times suggesting that true co-infection does occur, albeit with unknown frequency. A negative association was also observed between viruses PgHV-13 and 12, where the observed frequency of co-occurrence was below that expected by chance, given the prevalence of both viruses (*p*<0.001). These two viruses are very closely related and we speculate that co-occurrence may offer little benefit to the viral population because of increased competition for resources coupled with minimal potential for fitness gains via recombination. Even though previous studies showed a lack of immune recognition in betaherpesviruses[81](#_ENREF_81" \o "Boppana, 2001 #64), we suggest this might act as an effective mechanism for reducing co-existence of closely related viruses by preventing sequential infections. Studies on co-infection are thus highly relevant because they explore community structure and the functional relationship between viral community structure and the evolution of new viral diversity. We have used coevolutionary tools to quantify the degree of host-switching between bat species[82](#_ENREF_82), and have generated preliminary analyses to understand these patterns in bat CoVs. **These preliminary results have strongly guided the development of specific tests and hypotheses in the current proposal (Hyp 3, Hyp 8).**

**Viral discovery in the Rhesus macaque (*Macaca mulatta***)

Most recently, we expanded our viral discovery efforts even further to include both cPCR and HTS in order to investigate how much viral diversity is missed when cPCR is the only discovery tool used. The samples included feces from 458 individual macaques collected from 9 different urban sites across Bangladesh. They were processed using a discovery strategy that included cPCR for 28 different viral families as well as Illumina HiSeq HTS (with 5 million reads allocated per sample), and we identified a total of 184 viruses from 14 families. We identified 37/184 by cPCR, and 147/184 by HTS – highlighting the usefulness of combining the high sensitivity of PCR with the broad reactivity of HTS. We make particular note of one virus (currently designated MmMNG-1) which is distantly related to members of the *Ebolavirus* genus, and likely to represent a new family within the order *Mononegavirales.* These data demonstrate that our discovery platform is well designed to discover significant viral diversity with which to explore the hypotheses presented in this proposal. These data are currently in review at Nature Communications.

**RESEARCH PLAN**

**Our team**: We have assembled a unique, experienced, multidisciplinary team of ecologists, virologists, evolutionary biologists, mammalogists, and veterinarians. Our team includes leading disease ecologists, evolutionary biologists, and wildlife veterinarians (Daszak, Murray, Olival, Rostal, Zambrana-Torrelio) from EcoHealth Alliance, a research institution that specializes in understanding the ecological links between pathogen and wildlife diversity; lead scientists (Suzan, Rico, Ojeda) from Mexico’s premier university, Universidad Nacional Autónoma de México (UNAM) who specialize in chiropteran research and who manage the field site research; and the Director (Lipkin), and virologist (Anthony) of the world’s leading pathogen discovery laboratory, the Center for Infection and Immunity (CII) at Columbia University, which specializes in identifying novel viruses from wildlife and humans. **Our team has been collaborating effectively for over a decade**, conducting ecological fieldwork in Mexico, Brazil, Bolivia and >20 other countries, collecting and importing samples into the USA from a wide diversity of species, including bats from Mexico, analyzing and modeling viral traits and diversity in bats, and characterizing hundreds of new viruses.

**Field methods**

***Field site characterization:***Our primary field site, known locally as “El Volcán de los Murciélagos” ("Volcano of Bats"), is located in Campeche, Mexico on the Yucatan Peninsula (**Fig. 2)**. The cave is located in the Calakmul Reserve, the most expansive protected area in Mexico and a recognized biodiversity hotspot. The Calakmul region contains the largest forest area in the Mexican tropics, and is an extremely important site for bat conservation with the largest known population of bats on the peninsula[83](#_ENREF_83).

**Fig 2.** Cave map and location of “Volcano of Bats” showing distinct roosting locations for two species.

***Study species:*** Eight bat species from five families co-occur in the cave. In previous studies, including our own pilot study, all bat specieswerereadily detectable via hand nets or harp traps[84](#_ENREF_84),[85](#_ENREF_85), although their abundances vary. We will thus target the six most detectable species within the cave (*Mormoops megalophylla, Pteronotus parnellii, P.davyi, Natalus laticaudatus, N. mexicanus* *and Glossophaga soricina*) and an additional two species that are readily detectable outside the cave (*Artibeus jamaicensis* and *A.lituratus*) (n=8 species total) (see below). The total population of the cave is estimated to be 0.8-2.24 million individuals (population size varies seasonally after the influx of pups, prior to dispersal and subject to the survival rate of juveniles)[64](#_ENREF_64). The eight species selected are phylogenetically diverse and vary in the functional traits considered in this proposal (see Hyp 5, 6).

***Sample size:*** Our target sample size is at least **400 individuals per bat species**. Our preliminary work on *P. giganteus* indicates that sampling 4 sample types from each of 400 individuals from each host species should be sufficient to identify a large fraction (~75%) of the viruses expected to occur within the target viral families and determine their co-occurrence patterns. Prevalence of each virus in each host species will be estimable with a 5% maximum tolerable error at almost any background prevalence (except for the extremes close to 0 and 100%)[86](#_ENREF_86" \o "Digiacomo, 1986 #119). Together, these data will allow an accurate characterization of virodiversity metrics and their confidence intervals (depending to some extent on the number of unseen species)[87](#_ENREF_87" \o "Chao, 2003 #120) within and among host species, and facilitate our analyses with multiple regression on distance matrices (MRM) (**n = 8C2 to 10C2 = 28 to 45 pairwise values** for each variable, dependent on the final number of host species adequately sampled, see above and Methods).

Our previous experience with bats in Mexico gives us confidence that we can capture the large number of bats proposed in this study for at least 6 of the 8 species that occur in the cave and 2 additional species outside the cave. To better estimate the sampling requirements for the proposed study, **we conducted a targeted pilot study at the proposed field site** (the ‘Volcano of Bats’). In total we captured 2,120 individuals in a single harp trap (2 x 3 m) placed near the entrance of the cave for 3 hours per night over 4 nights (total sampling effort = 24 m2/12 hours). These captures comprised all 8 species previously known to inhabit the cave, which occur at varying abundances: *M. megalophylla* (1,142 individuals, nightly average = 285)*, P. parnellii* (474 ind, na = 118)*, P. davyi* (324 ind, na = 81), *N. laticaudatus* (80 ind, na = 20)*, N. mexicanus* (65 ind, na = 16)*, G. soricina* (30 ind, na = 7.5)*, P. personatus* (4 ind, na = 1)and *Myotis keaysi* (1 ind, na = 0.25)..Based on these species-specific capture rates, we estimate that we will need to conduct sampling for approximately 54 nights in total to ensure that we meet the target sample sizes for all target species (*M. megalophylla* 400/285 = 1.4 nights; *P. parnellii* 400/118 = 3.4 nights*; P. davyi* 400/81 = 4.9 nights; *N. laticaudatus 400/*20 = 20 nights; *N. mexicanus* 400/16 = 25 nights; *G. soricina* 400/7.5 = 53.3 nights)*.* Based on these numbers, we will impose nightly quotas (n=~10-15 individuals per species per night, depending on total sampling capacity per night) on sampling the more common species in order to ensure that their sampling takes place consistently over the full course of the study. Once this nightly quota has been reached, all subsequent captures of these species will be released immediately without processing. The two remaining, lower abundance species in the cave (*P. personatus* and *M. keaysi*)will also be sampled opportunistically. Outside the cave we will target the two most abundant species in the region (*A. lituratus* and *A.jamaicensis*), we estimated that will need a minimum of 25 nights of sampling to capture 400 animals of each species (400/20 = 20 nights and 400/15 = 25 nights).

***Sampling Plan*:** A total of approximately 54 sampling nights will be required to achieve target sample sizes for all species (see above). To account for potential seasonality and inter-annual differences in host and viral detectability, two sampling trips will be conducted per year, one in each of the wet and dry seasons, for two years. Bat sampling trips will last two weeks each during the Y1 and Y2 with traps open for 14 nights in each (=14x2x2 = 56 trap nights in total). In Yr1, two weeks of cave mapping will also follow bat sampling.

***Bat capture:*** We will use two harp traps 2 x 3m in size placed 10m apart to capture cave bats. Harp traps decrease the risk of injury to and from captured animals and enable greater speed in processing. For the non-cave bats we will open four mist nets per night within 100m of the cave entrance to avoid capturing bats exiting or entering out of the cave. Traps and nets will be opened 40 minutes prior to sunset for four hours and checked every 10 minutes. The total sample effort in the cave will be 240m2/160 hours per year (6 m2/4 hours x 20 nights x 2 harp traps) and for outside the cave 1350m2/100 hours (18 m2/4 hours x 25 nights x 3 mistnets).

***Animal handling and sampling procedure:*** All animals will be identified and non-invasively sampled to collect blood and oral, urine and rectal swabs using standardized techniques[24](#_ENREF_24). Bats will be placed in cotton bags with drawstring mouths and kept in a cool, dry place until sampling; holding time will not exceed six hours. All bats will be identified by local experts and with field guides for Mexican bats[88](#_ENREF_88). Genus, species, age class, sex, and morphometrics (e.g., forearm length, weight) will be recorded[89](#_ENREF_89). Fine polyester swabs will be used to collect samples. We will use a venipuncture system to collect blood at a ratio of no greater than 10µl per 1g of body weight (equivalent to 1% of body weight)[90](#_ENREF_90). Swabs will be placed in 200uL of virus transport medium (VTM). Sealed, labeled vials with samples will be placed in a liquid nitrogen tank in the field to maintain cold-chain during transport to the laboratory for viral analysis.

***Vertebrate Animals:***All vertebrate research methods proposed here are covered by IACUCs issued to EHA from Tufts Univ. (G2011-106) and UC Davis (16048). We have valid permits from the Ministry of the Environment (SEMARNAT) to capture and collect specimens from bats (Suzán, FAUT- 0250).

***Biohazard and Personnel Safety:*** We will minimize the risk to our field and laboratory personnel through the use of appropriate personal protective equipment (PPE) and training. All field personnel and students collecting bat samples will wear full PPE (dedicated clothing, gloves, safety glasses and respirators). An experienced field veterinarian will supervise all bat handling, sample collection, biohazard disposal, and personnel safety. The current field team members have extensive experience working with bats at the proposed site and throughout Mexico and are highly experienced with animal safety and biohazard protocols. All students and new field personnel added to the project will have training in bat handling, zoonotic diseases of bats, sample collection, handling biohazardous waste, and sharps and the proper use of PPE. Additionally, all field staff will be fully vaccinated for rabies and demonstrate a protective titer before handling bats. All materials from bat sampling (needles, syringes, tubes, etc.) will be treated as potentially biohazardous material, and will be properly incinerated at UNAM. Laboratory personnel at CII have been thoroughly trained in the use of PPE, handling infectious materials, blood-borne pathogens, and the disposal of biohazardous wastes. CII staff have significant experience in working with infectious samples from both humans and animals for the detection of known and novel viruses, and have a biosafety level 3 laboratory where infectious samples will be processed until inactivated (at which point they will be processed at BSL2). CII staff undergoes annual fit testing (for respiratory protection) and medical surveillance in order to be cleared for work in the BSL3. .

***Sample export and import:*** Our collaborative group has over a decade of experience successfully importing samples into the USA from a range of high profile countries (including India, Indonesia, Brazil and Mexico) from bats and other wildlife species including those protected by CITES. This includes specifically importing samples from UNAM into CII[4](#_ENREF_4),[91](#_ENREF_91). Centers for Disease Control and Prevention (CDC) permits for the importation of bats from Mexico have been successfully obtained for four consecutive years (through 2015).

***Mapping roosts within the cave:*** Using methods similar to Rodríguez-Durán[92](#_ENREF_92" \o "Rodríguez-Durán, 1998 #123) we will complete a roost map of the cave. Current work by our collaborators at the Universidad Autónoma de Campeche have mapped the inclination and length of the cave (**Fig. 2**), and the roosting locations for two species, *N. laticaudatus* near the mouth of the cave and *N. mexicanus* near the back of the cave. We will make several expeditions within the cave and use visual observation and hand nets to map the roosting locations of the various species within the cave. We will map these sites using a laser distance meter. We will also characterize the cave microclimate (temperature and relative humidity) at 5-10 sites across the cave transect using iButton data loggers. These data will contribute to our analysis of potential interspecies contact via pairwise ‘proximity’ scores between species pairs within the cave (see Hyp 7).

**Laboratory methods**

***Overview:*** Viral discovery will be performed using a combination of consensus PCR (cPCR) and high throughput sequencing (HTS) at the CII. We have developed this strategic approach to combine high sensitivity with broad reactivity, and have used it in the discovery of novel agents from many different hosts and sample types [24](#_ENREF_24),[91](#_ENREF_91),[93-97](#_ENREF_93). Consensus PCR uses degenerate primers that bind to short conserved domains flanking stretches of variable sequence. This allows for the ‘universal’ amplification of viruses within a given viral family, and for strains (both known and novel) to be discerned from the variable region within. This method is inexpensive and has been used successfully in our lab to identify hundreds of novel viruses. It is however limited by the ability of the degenerate primers to bind successfully to the target (conserved) domains and therefore has the potential to miss viruses that show divergence in this region. By contrast, HTS is more expensive and less sensitive (sequencing depth dependant), but is able to capture a very broad diversity because it does not rely on degenerate primers. Instead, HTS uses adaptors that are ligated onto dsDNA for unbiased amplification of all nucleic acids present in the sample. Following unbiased amplification, viral sequences are parsed out *in silico* and identified based on genetic homology to known viruses. We plan to use cPCR as our main discovery tool and apply HTS to a subset of samples as a secondary means of capturing any diversity that falls outside the detection limits of our PCR assays.

***Sample Preparation and Extraction:*** All samples (max **n=12,800** if all sample types are successfully collected from each individual) will be subject to a process of viral particle enrichment through centrifugation, filtration and nuclease treatment in order to remove bacteria, host cells, and naked/unencapsulated nucleic acids. This process selects for the retention of intact viral capsids and the removal of the abundant bacteria and host cell debris that would otherwise dominate the sample. Viral RNA/DNA will then be extracted using the MagnaPure® (Roche) platform for cPCR or HTS testing.

***Consensus PCR:*** Following particle enrichment, cDNA synthesis will be performed using SuperScript® III (Invitrogen). All cDNAs for each animal (maximum of four) will then be pooled (**n=3,200** pools) for cPCR testing. When positive cDNA pools are identified, cDNA from the individual samples (throat, faecal, urine and blood) will be subsequently tested to identify the positive sample(s). Positive PCR products will then be cloned into Strataclone™ PCR cloning vector and 12 white colonies sequenced to: 1) generate high quality sequence data for analysis (the highly degenerate nature of our PCR primers often preclude efficient direct sequencing), and 2) to identify any co-occurring viruses that might have been simultaneously amplified in the PCR (this occurs because the degenerate primers will bind to the conserved sequence in multiple strains if present in that sample). A total of 21 broadly reactive cPCR assays will be used, targeting 15 different viral families/genera, including: corona-; paramyxo-; hanta-; influenza-; herpes-; adeno-; astro-; flavi-; hepaci-; pegi-; polyoma-; alpha-, rota-; adeno-associated; and picobirna- viruses. These assays represent a combination of published and unpublished (but publicly available) assays that were either developed or tested in our lab as part of a global viral discovery effort (funded under the USAID PREDICT program) and were selected for this proposal because of their demonstrated ability to detect divergent and novel viruses in many parts of the world - including Mexico. These assays were also selected to account for viruses with variable characteristics (see Hyp 8). Based on our previous work[4](#_ENREF_4),[24](#_ENREF_24),[91](#_ENREF_91) we anticipate ~30% of samples to test positive for at least one virus, and for families such as herpesvirus, adenovirus, adeno-associated viruses and picobirnavirus to show frequent co-occurrence (Hyp 3 and Hyp 8).

***High Throughput Sequencing (HTS):*** HTS will be used to complement our cPCR strategy outlined above by 1) generating additional genome sequence for viruses identified by cPCR, 2) identifying viruses that belong to one of the 15 target families, but that were not successfully detected because of sequence divergence in the primer binding site, and 3) to identify viruses from additional viral families that might be present and might contribute to our comparisons of viral diversity in each species. This approach is supported by our preliminary data on discovery in rhesus macaques. A subset of **1000 pools** will be processed by HTS to satisfy these objectives, which will be selected randomly (but weighted to ensure each species is represented relative to its abundance). We will also allocate resources to deep sequence a further 120 pools, to ensure additional genome sequencing for any viruses of interest that were detected by cPCR in pools that were not within the main 1000 pool subset (e.g those for which insufficient sequence was generated to obtain an accurate phylogenetic placement). Enriched viral nucleic acids will be subjected to first and second strand cDNA synthesis then fragmentation in the Covaris E210 System. Libraries will be constructed by ligating dsDNA adaptors (with 3’-dTMP overhangs) to the fragmented double-stranded cDNA, followed by PCR amplification targeting these specific adaptors. The final yield will be analyzed and quantified on a bioanalyzer. Sequencing will be performed on the Illumina Hiseq 2500 platform, and 5 million reads allocated to each pool (32 lanes of Illumuna Hiseq). Following de-multiplexing of sequence data, raw reads will be pre-processed by quality filtering and computational subtraction against host reference databases to remove host sequences using the bowtie2 mapper. Host subtracted reads will be used to generate quality control reports. Quality control filtered reads will be assembled using MIRA (3.9.15) assembler and all contigs and unique singletons will be evaluated by MegaBlast against the entire Genbank nucleotide database. Sequences that do not generate hits with high similarity will be reassessed against a translated nucleotide viral database using BlastX. Viral hits from BlastX will be re-blasted against the entire Genbank protein database to correct for biased e-values. CII has 5 dedicated bioinformaticians; thus, we anticipate no difficulty in completing these analyses. We note that despite our current plan to use Illumina HiSeq,

***Testing Strategy:*** To summarise, we will test all four samples from 400 individuals from 8 different species (=3,200 pooled samples) by cPCR, for 15 viral families. We will also test a subset of 1000 pools by HTS (evenly distributed among the 8 species). We will reserve resources for an additional 120 HTS pools should we need them to fully characterise viruses identified by cPCR.

**AIM 1 METHODS: Establish whether virodiversity patterns vary among mammalian host species (descriptive).**

**Quantifying viral richness, diversity and inter-specific beta diversity**

***Viral identity:*** Viral sequences identified by cPCR or HTS will be analyzed phylogenetically (by family) and delineated based on distinct monophyletic clustering. This approach is concordant with the recently ratified definition of a virus ‘species’[98](#_ENREF_98), and obviates the wide variations in criteria previously used to demarcate individual viral species. We qualify that it is not our intention to propose the recognition of new species unless we have sufficient data to do so. Rather, it is our intention to adopt the central principal of monophyly in order to align with the general philosophy ratified by the International Committee for the Taxonomy of Viruses (ICTV). . We recognize that monophyly often has a nested structure, so we clarify our ‘operational unit’ to be the smallest cluster of overlapping genetic identities. We include a 2% ‘error threshold’ in our delineation of clusters, to account for any PCR or sequencing errors that may artificially separate groups of sequences. Thus, a monophyletic cluster is defined as the smallest group of overlapping sequence identities +/- 2%. This is the approach we have now used in several of our studies, including i) the virodiveristy of *P.giganteus* and ii) the viral community structure of macaques. The use of monophyletic clusters (MC) as our operational unit offers several advantages. First, this approach can be applied consistently to all viral families. Second, it only considers phylogenetic relatedness between viruses, and is not biased by the host from which a given sequence was identified, which is important to our considerations of host specificity. This allows us to investigate whether closely related sequences can be present in the same host (potentially revealing insights into the relative contribution of mechanisms such as competitive exclusion or immune recognition) or in different hosts (potentially representing dispersal). Third, MCs do not rely on specific measures of genetic distance (or cut-offs) to discern groups, and therefore allow a range of genetic diversity to be considered by any given cluster (i.e clusters do not necessarily all have equal diversity within). This presumably better reflects any biological pressures that might be acting to either constrain diversity within a MC (again, strong immune-recognition) or allow the accumulation of greater genetic diversity. Despite our definition of a viral ‘operating unit of analysis’, we also intend to investigate the effect of scale on our analyses in order to understand the degree to which our definition has affected our conclusions. This may include increasing the ‘nestedness’ of our clusters (i.e making them bigger), or investigating cluster-independent methods as they become available (the literature on phylogenetic measures of b-diversity is rapidly expanding). We assert that it is not our intention to make any assertions of novelty for the viruses we identify because of the implicit association with an official species designation. Rather, we intend to cluster viruses into MCs for analytical purposes (to satisfy the objectives outlined in this proposal), and then separately report the level of sequence identity these clusters have to known viruses. This study will generate a rich database of viral genomes that will be suitable for additional genetic and phylogenetic studies. We plan to make full use of these data to describe the molecular epidemiology of these viruses in a global context, examine how evolutionary mechanisms such as recombination and sequence duplication/deletion have shaped the observed diversity, and design assays for viral discovery of pathogens affecting public and domestic animal health.

***Sequence data***: All viral sequence data together with individual level data (e.g., specimen ID, sex, age, species) will be entered into our central database (see Data Management Plan). We will aggregate data at different hierarchical levels (e.g., sample, individual, roost group, species, community) using different resolutions (e.g., prevalence and binary data) relevant to each hypothesis.

***Statistical methods overview:***We consider the detection and discovery of viruses akin to the problem of detection and discovery of biodiversity in ecological studies. The basic mechanism of species detection occurs from drawing samples by collection from some larger assemblage[99](#_ENREF_99). In this context, our samples will be compiled as described above: urine, throat, fecal, or blood taken from an individual bat, which represents the ‘biomes’ for our assemblage of interest. Each of these sample types targets a unique viral ‘habitat’ within the host species, each with potentially differing efficacy for detecting any given virus. We are interested in this variation and will include sample types as one unit for our analyses. However, our primary interest will be at the level of the individual, therefore we will also pool sample types, providing a complete picture of the detectable viruses in each individual. For the purposes of our analyses we will consider the sampling of each animal a random and equivalent draw from the larger assemblage of viruses associated with this host community. We recognize that viral PCR assays may generate false negatives due to stochastic variation, inhibitory compounds, low viral load, and other factors. Our discovery approach has been designed to combat this as best as possible by using multiple assays per viral family (where available) and HTS to account for any bias or differences in efficiency that may exist in the cPCR.

***Viral richness*** (**Hyp 1**) will be assessed in each host species by both evaluating the detected virus lists (composition) and by using non-parametric species richness estimators (e.g., Chao2, ICE) commonly used in biodiversity studies[99](#_ENREF_99). For the latter, we will implement models that utilize both incidence (i.e., presence/absence of each virus in each host species) and abundance (i.e., number of detections of each virus in each host species) distributions to estimate viral richness given incomplete sampling (and hence to also estimate the number of undetected viruses in each assemblage). From our samples, we will first construct virus accumulation and rarefaction curves for visualization. The asymptote of the rarefaction curve provides the estimate of the total number of viruses in the assemblage (Fig. 1). We will use the non-parametric estimators Chao2, ICE, and Jackknife[65](#_ENREF_65),[100](#_ENREF_100) statistics to provide three alternative estimates of the asymptote. Unlike conventional curve fitting procedures, the non-parametric estimators make no assumptions on an underlying abundance distribution, do not require *ad hoc* or *a priori* model fitting, are relatively robust to spatial autocorrelation and scale and frequently outperform other methods of richness estimation[99](#_ENREF_99). In addition, we fill follow Chao *et al.*[65](#_ENREF_65) to calculate how many additional samples would be required to detect any proportion of the estimated virus richness (*see* Preliminary Data[65](#_ENREF_65),[101](#_ENREF_101)) and inform us about our sampling completeness.

***Viral diversity*** (**Hyp 2**) in each host species (or other grouping unit of interest e.g., sample type) will be evaluated with richness-independent species diversity indices that capture information about richness~abundance relationships (e.g., Shannon’s H)[26](#_ENREF_26). We will use Chao and Shen’s (2003) non-parametric estimation of Shannon’s H to account for undetected species in the sampled viral assemblages of each host species[87](#_ENREF_87).

***Co-occurrence (Hyp 3)***patterns of viral association / disassociation within samples / individuals / species will be explored with the Fortran software program PAIRS[73](#_ENREF_73), a null model approach utilizing the C-score statistic as our measure of species co-occurrence. PAIRS implements a Bayesian approach (Bayes M criterion) to detect non-random associations between pairs of species[72](#_ENREF_72).

***Functional inter-specific beta diversity (viral turnover between host species)* (Hyp 4)** will be quantified with compositional, functional and phylogenetic dissimilarity indices. Our final choice of indices will somewhat depend on the characteristics of the dataset, but we anticipate using indices that are richness-independent[26](#_ENREF_26), additive in that beta-diversity can be decomposed into its two constituent components (turnover and nestedness, which may reflect contrasting functional mechanisms underlying differences in beta diversity)[28](#_ENREF_28),[29](#_ENREF_29),[102](#_ENREF_102), and adjustable for the non-detection of species[30](#_ENREF_30). Both incidence based metrics (utilizing presence/absence of each virus in each host species; e.g., Jaccard or Sorenson) and abundance based metrics (utilizing number of detections of each virus in each host species; e.g., Horn index of overlap) will be explored[27](#_ENREF_27). Utilizing abundance data provides more information about differences in diversity than presence/absence data, and interpreting indices in terms of ‘numbers equivalents’ together result in more power to detect statistically and biologically meaningful differences between two or more communities[26](#_ENREF_26). To compare whether viral assemblages among host species are more or less similar than would be expected by chance, we will take a null model approach[27](#_ENREF_27),[103-105](#_ENREF_103). This will involve calculating the similarity metric (e.g., Horn index of overlap (H), or similar) for the viral assemblages between pairs of species and testing for deviations from that expected by random chance. Deviation from expected similarity (Hdev) will be calculated by randomizing the observed viral occurrence/abundance matrix using Monte Carlo randomizations. Randomizations will be constrained to ensure fixed row and column totals, equating to maintaining observed viral richness in each host species and maintaining observed viral occupancy/abundance across host species (commonly referred to in the biodiversity literature as a fixed-fixed model)[106](#_ENREF_106). For each randomized matrix, we will calculate all pairwise H scores. Our null model will assume the mean H of these randomizations (Hnull). Deviation from the null model will be calculated as the difference between the mean H observed (Hobs) in the data and the mean H expected, such that Hdev = Hobs - Hnull. Positive values of Hdev will thus indicate that viral community assemblages between host species are more similar than would be expected by random chance, while negative values would indicate greater dissimilarity in the viral assemblages than would be expected by chance. For a practical example of this method (using Jaccard’s similarity index), see Smith et al. (2009)[107](#_ENREF_107).

**AIM 2 METHODS: Understand the drivers of genetic, taxonomic and functional virodiversity among mammalian host species (predictive).**

Our objective is to identify what factors (e.g., phylogeny, functional intrinsic factors e.g., ecological and life-history traits, extrinsic factors e.g., contact potential) contribute to non-random patterns of viral community similarity and structure among hosts. We will use models (e.g., multiple regression on distance matrices (MRM)[108](#_ENREF_108) and multi-model averaging and inference to account for model and parameter estimation uncertainty)[109](#_ENREF_109) to explain the pairwise dissimilarity of viral assemblages among host species (quantified as described above; Hyp 4). In this framework, **response variables are matrices of viral assemblage dissimilarity values** (i.e., functional beta diversity, phylogenetic beta diversity). **Explanatory variables are matrices of ‘distances’ in factors of interest**. We will explore both intrinsic host (Hyp 5, 6) and viral traits (Hyp 8) and extrinsic factors (Hyp 7). Distance values for explanatory variables can be binomial (e.g., a pair of frugivores will have a dietary niche similarity value of 1 i.e., the same, while a pair of host species comprising a frugivore and an insectivore will have dietary similarity value of 0 i.e., not the same), ordinal (e.g., each pair of species will have a proximity score ranging between 1 and 5, where 1 means low proximity and 5 means high proximity), or continuous (e.g., the raw difference in measured variables for each species, such as mean body mass). All possible combinations of our 8 study species gives a sample size of **8C2 = 28 pairwise** **values** for analyses (up to 45 if we adequately sample our two lower abundance species). Multiple regression power analyses (conducted in the ‘pwr’ package in R) suggest we will have adequate power (0.75) with this sample size (n=28) to detect moderate effect sizes (f2>0.25) at a standard significance level (0.05) using a forward stepwise approach (starting with df=1). MRM is conceptually similar to Mantel’s procedure[110](#_ENREF_110), and is similarly used to account for non-independence in the elements of distance matrices and where statistical tests of significance and the effects of individual predictors (e.g., variation explained) are performed via permutation[41](#_ENREF_41),[108](#_ENREF_108),[111](#_ENREF_111). The effect of phylogenetic non-independence among species units will be included in this framework by including a matrix on phylogenetic distance between species pairs (e.g., patristic distance) as another explanatory variable (See Hyp5). In this way, we will be able to assess the proportion of variance in the response variable across all species pairs that is attributable to each predictor variable. To further quantify the degree of viral sharing among bat hosts, we will use statistical measures of cophylogenetic congruence that explicitly include host and virus phylogenetic data using “global fit” statistics, e.g. ParaFit[112](#_ENREF_112" \o "Legendre, 2002 #135) and Procrustean Approach to Cophylogeny (PACo)[113](#_ENREF_113), as well as event-based methods, e.g., Jane 4[114](#_ENREF_114), techniques we recently used to quantify the cophylogeny of bat-bacterial associations[82](#_ENREF_82). All statistical analyses will be conducted in R with packages for community ecology and species diversity (e.g., vegan, fossil, ecodist), and phylogenetic modeling (e.g., picante, geiger, ade4, ape, phytools).

**Project Management and Timeline**

This project will be managed by a PI (Daszak) with two decades experience managing international, multidisciplinary disease ecology research, including successful multi-year projects funded by NSF, NIH and USAID. Co-PIs Suzán, Murray, Olival and senior personnel Zambrana-Torrelio will lead analysis of data. Co-investigator Suzán and senior personnel Rostal and Rico, veterinarians, will lead the field teams, and each build on over a decade of experience working internationally and 4 years together in Mexico, where Suzán was the Mexico lead and Rostal the Latin America Coordinator for EHA’s USAID PREDICT project. Co-investigator Anthony and senior personnel Lipkin will oversee the laboratory analyses. As director of the CII, Lipkin brings 4 decades of experience in pathogen discovery, and manages the world’s premier

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Timeline** | **Year 1** | **Year 2** | **Year 3** |
| Field work/sample collection (2 trips p.a.) | X | X |  |
| Laboratory work | X | X | X |
| Viral Sequencing/HTS |  | X | X |
| Phylogenetic analysis; bioinformatics |  | X | X |
| Post-doctoral fellow training | X | X |  |
| Masters' student training | X | X | X |
| REU student training | X | X |  |
| Publishing results |  | X | X |
| Annual PI meeting | X | X | X |
| Domestic and international conferences | X | X | X |
| Monthly team video calls | X | X | X |

viral discovery lab. EHA and CII have a decade long partnership, shared staff, and extremely close collaboration (offices are within a 30 minute metro ride of each other). To ensure efficient progress among the three institutions, monthly conference calls will be held with all project staff. In addition, an annual, in-person management meeting will be held at EHA’s offices in New York. Joint UNAM-EHA-CII teams will conduct field research and travel between the US and Mexico will occur for both US and Mexican staff to ensure a mutual understanding of techniques and better collaboration.

**BROADER IMPACTS**

This study will have significant implications for biodiversity science, conservation and public health. For the first time, we will identify critical **intrinsic host and viral traits** and **extrinsic factors** that drive viral diversity within an ecologically critical group of mammals, and in so doing allow us to better understand the **functional** link between host and microbial diversity. We will discover many novel viruses from species of bats that are known reservoirs of zoonotic diseases, and thereby acquire information that might be used in public health programs. We will also contribute improved estimations of total viral diversity to help justify, refine, target and budget for global surveillance efforts for the discovery of novel, potentially zoonotic, viruses in wildlife. This study will also help to shape the future direction of this field by generating new testable hypotheses that explore specific mechanisms behind the **functional** **associations** we identify. Our approach will also generate rich datasets (many full genome sequences) that will be shared publicly and can be used to address questions beyond those of this specific proposal.

We will share the results of this research with scientific peers via manuscripts and presentations at international conferences, and will contribute to the professional development of postdoctoral scholars, graduate students, field assistants in-country, and undergraduates. PI Suzán has been working with SEMARNAT (the Mexican Department of the Environment and Natural Resources) to develop a method for reporting the results of research conducted on Mexican wildlife, he has begun this initiative for reporting the USAID PREDICT data and we will follow this protocol with the results of this study as well. We will train four Research Experience for Undergraduates (REUs; two per year) from Columbia University. We also offer a **unique multidisciplinary post-doctoral fellowship** (see Post-Doctoral Mentoring Plan) that includes field experience, laboratory experience, training in ecological and functional modelling and phylogenetic and taxonomic techniques. In addition, **three Masters’ students** (see Student Training Plan) will be co-advised with Columbia professors and EHA’s adjunct research faculty or PI Suzán (UNAM students) and EHA investigators/senior personnel; EHA and UNAM has additional funding to cover any costs required for the Masters’ students beyond the budgeted amount for this proposal. Opportunities will also be available informally for students of PI Suzán and senior personnel Rico and Ojeda to join the team in the field to gain experience in bat identification and handling, sampling and biodiversity assessments (for example senior personnel Sotomayor and Martinez-Duque). Surveillance and environmental data collected in Mexico will be disseminated to government and public health agencies there, and in the USA, to help inform and protect the public against the risk of viral spillover.

**Prior NSF and Federal Support**

**P. Daszak, S.J. Anthony, K.J. Olival:** NIH-NIAID 5R01AI079231; $2,687,394; 2009-2013; ‘Risk of Viral Emergence from Bats’. Published over 50 peer-reviewed papers, including in *Nature*, *PNAS*, *Lancet*, *MBio*, *Emerging Infectious Diseases*, *PLoS Pathogens*, *J. Virol.* Discovered >100 new viruses from bats, produced first ever predictive map of bat zoonoses under climate change, and first estimate of unknown viral diversity in a mammal. Trained >250 participants from 8 countries in non-invasive sampling protocols for optimal bat virus sample collection.

**W. Ian Lipkin**: NIH AI057158; $96,000,000; 2003-2014; ‘Northeast Biodefense Center’. As Director and Principal Investigator, Lipkin coordinated the activities of >250 investigators at 28 institutions in NY, NJ and CT in diagnostics, bacterial and viral pathogenesis, immunology, and vaccine research resulting in close to 800 publications and 68 patent applications in the field of emerging infectious diseases.

**Kevin J. Olival:** NIH Fogarty Global Health Fellow; $204,668; 2009-2011; Ecology of Nipah virus in Bangladesh; population genetic structure of *P. giganteus*; and pathogen discovery from a diverse range of bats. Over the course of this 2-year award, >10 papers were published including 4 in the high-impact journal, *Emerging Infectious Diseases*; presented at >20 national and int’l conferences; and media coverage in the *New York Times* Science section.

CHECK if need to add others? – What about the EHN RCN?

1 Anthony, S. J. *et al.* A strategy to estimate unknown viral diversity in mammals. *mBio* **4**, e00598-00513 (2013).

2 Anthony, S. J. *et al.* Predictable patterns of viral diversity in a non-human primate. *Nature Communications* (In Review).

3 Quan, P.-L. *et al.* Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8194-8199, doi:10.1073/pnas.1303037110 (2013).

4 Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *Journal of General Virology* **Online Early** (2013).

5 Rico Chávez, O. *et al.* Viral diversity of bat communities in human-dominated landscapes in Mexico. *Veterinaria México OA* **2** (2015).

6 Suzan, G. *et al.* Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. *Ecology and Evolution* **5**, 865-873 (2015).

7 Sotomayor-Bonilla, J. *et al.* Short Report: Dengue Virus in Bats from Southeastern Mexico. *American Journal of Tropical Medicine and Hygiene* **91**, 129-131, doi:10.4269/ajtmh.13-0524 (2014).

8 Olival, K. J. *et al.* Ebolavirus Antibodies in Fruit Bats, Bangladesh. *Emerging Infectious Diseases* **19**, doi:DOI: 10.3201/eid1902.120524 (In Press).

9 Murray, K. A. & Daszak, P. Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. *Current Opinion in Virology* **3**, 79-83, doi:10.1016/j.coviro.2013.01.006 (2013).

10 Allen, T. *et al.* EIGHT CRITICAL QUESTIONS FOR PANDEMIC PREDICTION. *Influence of Global Environmental Change on Infectious Disease Dynamics: Workshop Summary*, 182-193 (2014).

11 Schwind, J. S. *et al.* Capacity building efforts and perceptions for wildlife surveillance to detect zoonotic pathogens: comparing stakeholder perspectives. *Bmc Public Health* **14**, doi:10.1186/1471-2458-14-684 (2014).

12 Levinson, J. *et al.* Targeting surveillancefor zoonotic virus discovery *Emerging Infectious Diseases* **19**, doi:10.3201/eid1905.121042 (2013).

13 Olival, K. J., Epstein, J. H., Wang, L. F., Field, H. E. & Daszak, P. in *New Directions in Conservation Medicine: Applied Cases of Ecological Health* (eds A.A. Aguirre, R.S. Ostfeld, & P. Daszak) Ch. 14, 195-212 (Oxford University Press, 2012).

14 Morse, S. F. *et al.* Global distribution and genetic diversity of Bartonella in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). *Infection, Genetics and Evolution* **12**, 1717-1723 (2012).

15 Bogich, T. L. *et al.* in *New directions in conservation medicine* (eds A. A. Aguirre, R.S. Ostfeld, & P. Daszak) 607-618 (Oxford University Press, 2012).

16 Suttle, C. A. Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5**, 801-812, doi:10.1038/nrmicro1750 (2007).

17 Suttle, C. A. Viruses: a vast reservoir of genetic diversity and driver of global processes. *Retrovirology* **6**, doi:I7 (2009).

18 King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. *Virus taxonomy: Classification and noimenclature of viruses*. (Elsevier Academic Press, 2012).

19 Woolhouse, M., Scott, F., Hudson, Z., Howey, R. & Chase-Topping, M. Human viruses: discovery and emergence. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 2864-2871, doi:10.1098/rstb.2011.0354 (2012).

20 Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990-993 (2008).

21 Levinson, J. *et al.* Targeting surveillancefor zoonotic virus discovery *Emerging Infectious Diseases* **19**, 743-747, doi:10.3201/eid1905.121042 (2013).

22 Lipkin, W. I. The changing faces of pathogen discovery and surveillance. *Nature Reviews Microbiology* **11**, 133-141 (2013).

23 Daszak, P. & Lipkin, W. I. The search for meaning in virus discovery. *Current opinion in virology* **1**, 620-623 (2011).

24 Anthony, S. J. *et al.* A strategy to estimate unknown viral diversity in mammals. *mBio* **4**, doi:10.1128/mBio.00598-13 (2013).

25 Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. Bats: Important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews* **19**, 531-545 (2006).

26 Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427-2439, doi:10.1890/06-1736.1 (2007).

27 Magurran, A. E. *Measuring biological diversity*. (Blackwell Publishing, 2004).

28 Baselga, A. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**, 134-143, doi:10.1111/j.1466-8238.2009.00490.x (2010).

29 Leprieur, F. *et al.* Quantifying phylogenetic beta diversity: distinguishing between ‘true’ turnover of lineages and phylogenetic diversity gradients. *PLoS ONE* **7**, e42760, doi:10.1371/journal.pone.0042760 (2012).

30 Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T.-J. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62**, 361-371, doi:10.1111/j.1541-0420.2005.00489.x (2006).

31 Gilbert, G. S. & Webb, C. O. Phylogenetic signal in plant pathogen-host range. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **104**, 4979-4983, doi:10.1073/pnas.0607968104 (2007).

32 Hoverman, J. T., Gray, M. J., Haislip, N. A. & Miller, D. L. Phylogeny, life history, and ecology contribute to differences in amphibian suceptibility to Ranaviruses. *EcoHealth* **8**, 301-319 (2011).

33 Poulin, R. Decay of similarity with host phylogenetic distance in parasite faunas. *Parasitology* **137**, 733-741, doi:10.1017/s0031182009991491 (2010).

34 Pedersen, A. B. & Davies, T. J. Cross-Species Pathogen Transmission and Disease Emergence in Primates. *EcoHealth* **6**, 496-508, doi:10.1007/s10393-010-0284-3 (2009).

35 Cooper, N., Griffin, R., Franz, M., Omotayo, M. & Nunn, C. L. Phylogenetic host specificity and understanding parasite sharing in primates. *Ecology Letters* (2012).

36 Streicker, D. G. *et al.* Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. *Science* **329**, 676-679, doi:10.1126/science.1188836 (2010).

37 Faria, N. R., Suchard, M. A., Rambaut, A., Streicker, D. G. & Lemey, P. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philosophical Transactions of the Royal Society B-Biological Sciences* **368**, doi:10.1098/rstb.2012.0196 (2013).

38 Huang, S., Bininda-Emonds, O. R., Stephens, P. R., Gittleman, J. L. & Altizer, S. Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *The Journal of animal ecology*, doi:10.1111/1365-2656.12160 (2013).

39 Parrish, C. R. *et al.* Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* **72**, 457-470, doi:10.1128/mmbr.00004-08 (2008).

40 Cavender-Bares, J., Kozak, K. H., Fine, P. V. A. & Kembel, S. W. The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**, 693-715, doi:10.1111/j.1461-0248.2009.01314.x (2009).

41 Poulin, R., Krasnov, B. R., Mouillot, D. & Thieltges, D. W. The comparative ecology and biogeography of parasites. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2379-2390, doi:10.1098/rstb.2011.0048 (2011).

42 Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP & LJ., M.-C. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Currents: Tree of Life* **1**, doi:10.1371/currents.RRN1212. (2011).

43 Tobe, S. S., Kitchener, A. C. & Linacre, A. M. T. Reconstructing mammalian phylogenies: A detailed comparison of the cytochrome *b* and cytochrome oxidase subunit I mitochondrial genes. *PLoS ONE* **5**, e14156, doi:10.1371/journal.pone.0014156 (2010).

44 Patterson, B. D., Dick, C. W. & Dittmar, K. Roosting habits of bats affect their parasitism by bat flies (Diptera: Streblidae). *Journal Of Tropical Ecology* **23**, 177-189 (2007).

45 Boyles, J. G. & Willis, C. K. R. Could localized warm areas inside cold caves reduce mortality of hibernating bats affected by white-nose syndrome? *Frontiers in Ecology and the Environment* **8**, 92-98, doi:10.1890/080187 (2010).

46 Turmelle, A. S. & Olival, K. J. Correlates of viral richness in bats (Order Chiroptera). *EcoHealth* **6**, 522-539, doi:10.1007/s10393-009-0263-8 (2009).

47 Luis, A. D. *et al.* A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society B-Biological Sciences* **280**, doi:10.1098/rspb.2012.2753 (2013).

48 Loh, E. H. *et al.* in *Microbiol Spectrum* Vol. 1 OH-0009-2012 (American Society for Microbiology, 2013).

49 Lomolino & Weiser. Towards a more general species–area relationship: diversity on all islands, great and small. *Journal of Biogeography* **28**, 431-445, doi:10.1046/j.1365-2699.2001.00550.x (2001).

50 MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography*. (Princeton University Press, 1967).

51 Lloyd-Smith, J. O. *et al.* Should we expect population thresholds for wildlife disease? *Trends in Ecology & Evolution* **20**, 511-519, doi:10.1016/j.tree.2005.07.004 (2005).

52 Grenfell, B. & Harwood, J. Metapopulation dynamics of infectious diseases. *Trends in Ecology & Evolution* **12**, 395-399 (1997).

53 Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327-332, doi:10.1126/science.1090727 (2004).

54 Anderson, R. M. & May, R. M. Population biology of infectious-diseases .1. *Nature* **280**, 361-367 (1979).

55 Martin, A. P. & Palumbi, S. R. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences* **90**, 4087-4091, doi:10.1073/pnas.90.9.4087 (1993).

56 Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-848, doi:<http://dx.doi.org/10.1016/j.cell.2006.02.017> (2006).

57 Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459**, 193-199 (2009).

58 Nunn, C. L., Altizer, S., Jones, K. E. & Sechrest, W. Comparative tests of parasite species richness in primates. *The American Naturalist* **162**, 597-614 (2003).

59 Wang, L. F., Walker, P. J. & Poon, L. L. M. Mass extinctions, biodiversity and mitochondrial function: are bats 'special' as reservoirs for emerging viruses? *Current Opinion in Virology* **1**, 649-657, doi:10.1016/j.coviro.2011.10.013 (2011).

60 Krasnov, B. R. *et al.* Similarity in ectoparasite faunas of Palaearctic rodents as a function of host phylogenetic, geographic or environmental distances: Which matters the most? *International Journal for Parasitology* **40**, 807-817, doi:10.1016/j.ijpara.2009.12.002 (2010).

61 Davies, T. J. & Pedersen, A. B. Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings Of The Royal Society B-Biological Sciences* **275**, 1695-1701, doi:10.1098/rspb.2008.0284 (2008).

62 Pulliam, J. R. C. & Dushoff, J. Ability to Replicate in the Cytoplasm Predicts Zoonotic Transmission of Livestock Viruses. *Journal of Infectious Diseases* **199**, 565-568, doi:10.1086/596510 (2009).

63 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).

64 Arroyo Cabrales, J. *et al.* Evaluación de las poblaciones de murciélagos durante la construcción de la línea eléctrica., (Comisión Federal de Electricidad-Universidad Veracruzana. Convenio RRCPTTP-UV- 001/07, 2011).

65 Chao, A., Colwell, R. K., Lin, C. W. & Gotelli, N. J. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* **90**, 1125-1133, doi:10.1890/07-2147.1 (2009).

66 Telfer, S. *et al.* Species interactions in a parasite community drive infection risk in a wildlife population. *Science* **330**, 243-246, doi:10.1126/science.1190333 (2010).

67 Joseph, S. B., Hanley, K. A., Chao, L. & Burch, C. L. Coinfection rates in Φ6 bacteriophage are enhanced by virus-induced changes in host cells. *Evolutionary Applications* **2**, 24-31 (2009).

68 Lawn, S. D., Bekker, L. G., Middelkoop, K., Myer, L. & Wood, R. Impact of HIV infection on the epidemiology of tuberculosis in a peri-urban community in South Africa: the need for age-specific interventions. *Clinical Infectious Disease* **42**, 1040-1047, doi:10.1086/501018 (2006).

69 Newman, C. M. *et al.* Culex flavivirus and West Nile virus mosquito coinfection and positive ecological association in Chicago, United States. *Vector Borne and Zoonotic Diseases* **11**, 1099-1105, doi:10.1089/vbz.2010.0144 (2011).

70 Stapleton, J. T., Williams, C. F. & Xiang, J. GB virus type C: a beneficial infection? *Journal of clinical microbiology* **42**, 3915-3919, doi:10.1128/JCM.42.9.3915-3919.2004 (2004).

71 Techasaensiri, B., Techasaensiri, C., Mejias, A., McCracken, G. H., Jr. & Ramilo, O. Viral coinfections in children with invasive pneumococcal disease. *Pediatric Infectious Disease Journal* **29**, 519-523, doi:10.1097/INF.0b013e3181cdafc2 (2010).

72 Gotelli, N. J. & Ulrich, W. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* **162**, 463-477 (2010).

73 Pairs - a FORTRAN program for studying pair-wise species associations in ecological matrices [http://www.uni.torun.pl/\*ulrichw](http://www.uni.torun.pl/*ulrichw) (2008).

74 Haberland, M., Meyer-Konig, U. & Hufert, F. T. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *Journal of General Virology* **80**, 1495-1500 (1999).

75 Meyer-Konig, U., Ebert, K., Schrage, B., Pollak, S. & Huefert, F. T. Simultaneous infection of healthy people with multiple human cytomegalovirus strains. *Lancet* **352**, 1280-1281, doi:Doi 10.1016/S0140-6736(05)70487-6 (1998).

76 Norberg, P., Kasubi, M. J., Haarr, L., Bergstrom, T. & Liljeqvist, J. A. Divergence and recombination of clinical herpes simplex virus type 2 isolates. *J Virol* **81**, 13158-13167, doi:10.1128/Jvi.01310-07 (2007).

77 Smith, L., McWhorter, A., Shellam, G. R. & Redwood, A. J. The genome of murine cytomegalovirus is shaped by purifying selection and extensive recombination. *Virology* **In Press: Corrected Proof** (2012).

78 Greenwood, A. D. *et al.* A Potentially Fatal Mix of Herpes in Zoos. *Curr Biol* **22**, 1727-1731, doi:Doi 10.1016/J.Cub.2012.07.035 (2012).

79 Roycroft, E., Rose, L., Scallan, M. F. & Crowley, B. Molecular characterization of varicella-zoster virus clinical isolates from 2006 to 2008 in a tertiary care hospital, Dublin, Ireland, using different genotyping methods. *J Med Virol* **84**, 1672-1679, doi:Doi 10.1002/Jmv.23344 (2012).

80 Thiry, E. *et al.* Recombination in alphaherpesviruses. *Rev Med Virol* **15**, 89-103, doi:Doi 10.1002/Rmv.451 (2005).

81 Boppana, S. B., Rivera, L. B., Fowler, K. B., Mach, M. & Britt, W. J. Intrauterine transmission of cytomegalovirus to infants of women with preconceptional immunity. *New England Journal of Medicine* **344**, 1366-1371, doi:10.1056/NEJM200105033441804 (2001).

82 Lei, B. R. & Olival, K. J. Contrasting patterns in mammal–bacteria coevolution *Bartonella* and *Leptospira* in bats and rodents. *PLoS Neglected Tropical Diseases* **8**, e2738, doi:10.1371/journal.pntd.0002738 (2014).

83 Vargas-Contreras, J., Escalona-Segura, G., Arroyo-Cabrales, J., Rendón, J. & Navarro, L. Conservación de Murciélagos en Campeche. *THERYA* **3**, 53-66 (2012).

84 Escalona-Segura, G., Vargas-Contreras, J. A. & Interián-sosa, L. Registros importantes de mamíferos para Campeche, México. *Revista Mexicana de Mastozollogía* **6**, 166-170 (2002).

85 Pozo, C. & Calmé, S. in *Uso y monitoreo de los recursos naturales en el Corredor Biológico Mesoamericano (áreas focales Xpujil-Zoh Laguna y Carrillo Puerto)* Vol. SNIB-CONABIO BJ002 (eds M. Ponzo-de-la-Tijera & S. Calmé) (El Colegio de la Frontera Sur, Unidad Chetumal, 2005).

86 Digiacomo, R. F. & Koepsell, T. D. Sampling for detection of infection or disease in animal populations. *J Am Vet Med Assoc* **189**, 22-23 (1986).

87 Chao, A. & Shen, T.-J. Nonparametric estimation of Shannon’s index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics* **10**, 429-443, doi:10.1023/A:1026096204727 (2003).

88 Medellín, R. A., Arita, H. T. & Sanchez, H. *Identificación de los Murciélagos de México, Clave de Campo*. 2nd edn, (2008).

89 Kunz, T. H., Tidemann, C. R. & Richards, G. C. in *Measuring and Monitoring Biological Diversity Standard Methods for Mammals* (eds Don E. Wilson *et al.*) 115-156 (Smithsonian Institution Press, 1996).

90 Smith, C. S., De Jong, C. & Field, H. Sampling small quantities of blood from microbats. *Acta Chiropterologica* **12**, 255-258 (2010).

91 Quan, P. L. *et al.* Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences U S A* **110**, 8194-8199, doi:10.1073/pnas.1303037110 (2013).

92 Rodríguez-Durán, A. Non-random aggregations and distribution of the cave bats of Puerto Rico. *Journal of Mammalogy* **79**, 141-146 (1998).

93 Alagaili, A. N. *et al.* Middle East respiratory syndrome coronavirus infection in dromedary camels in saudi arabia. *mBio* **5**, doi:10.1128/mBio.00884-14 (2014).

94 Anthony, S. J. *et al.* West Nile Virus in the British Virgin Islands. *Ecohealth*, doi:10.1007/s10393-014-0910-6 (2014).

95 Anthony, S. J. *et al.* Emergence of fatal avian influenza in New England harbor seals. *mBio* **3**, e00166-00112, doi:10.1128/mBio.00166-12 (2012).

96 Anthony, S. J. *et al.* Identification of a novel cetacean polyomavirus from a common dolphin (*Delphinus delphis*) with tracheobronchitis. *Plos One* **8(7): e68239. doi:10.1371/journal.pone.0068239** (2013).

97 Tokarz, R., Sameroff, S., Leon, M. S., Jain, K. & Lipkin, W. I. Genome characterization of Long Island tick rhabdovirus, a new virus identified in Amblyomma americanum ticks. *Virology Journal* **11**, 26, doi:10.1186/1743-422X-11-26 (2014).

98 Adams, M. J., Lefkowitz, E. J., King, A. M. Q. & Carstens, E. B. Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Arch Virol* **158**, 2633-2639, doi:10.1007/s00705-013-1749-9 (2013).

99 Gotelli, N. J. & Colwell, R. K. in *Frontiers in measuring biodiversity* (eds A.E. Magurren & B.J. McGill) Ch. 4, 39-54 (Oxford University Press, 2011).

100 Chao, A. in *Encyclopedia of statistical sciences* Vol. Volume 12 (eds N Balakrishnan, C. B Read, & B Vidakovic) (Wiley, 2005).

101 Colwell, R. K. *et al.* Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology-Uk* **5**, 3-21, doi:10.1093/jpe/rtr044 (2012).

102 Bryant, J. A. *et al.* Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences* **105**, 11505-11511, doi:10.1073/pnas.0801920105 (2008).

103 Diamond, J. & Gilpin, M. Examination of the “null” model of connor and simberloff for species co-occurrences on Islands. *Oecologia* **52**, 64-74, doi:10.1007/bf00349013 (1982).

104 Connor, E. F. & Simberloff, D. Species Number and Compositional Similarity of the Galápagos Flora and Avifauna. *Ecological Monographs* **48**, 219-248 (1978).

105 Raup, D. M. & Crick, R. E. Measurement of Faunal Similarity in Paleontology. *Journal of Paleontology* **53**, 1213-1227 (1979).

106 Gotelli, N. J. NULL MODEL ANALYSIS OF SPECIES CO-OCCURRENCE PATTERNS. *Ecology* **81**, 2606-2621, doi:10.1890/0012-9658(2000)081[2606:nmaosc]2.0.co;2 (2000).

107 Smith, K. G., Lips, K. R. & Chase, J. M. Selecting for extinction: nonrandom disease-associated extinction homogenizes amphibian biotas. *Ecology Letters* **12**, 1069-1078, doi:10.1111/j.1461-0248.2009.01363.x (2009).

108 Lichstein, J. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol* **188**, 117-131, doi:10.1007/s11258-006-9126-3 (2007).

109 Burnham, K. P. & Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*. (Springer-Verlag, 2002).

110 Legendre, P. & Legendre, L. *Numerical Ecology*. (1998).

111 Zapala, M. A. & Schork, N. J. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences* **103**, 19430-19435, doi:10.1073/pnas.0609333103 (2006).

112 Legendre, P., Desdevises, Y. & Bazin, E. A statistical test for host–parasite coevolution. *Systematic Biology* **51**, 217-234, doi:10.1080/10635150252899734 (2002).

113 Balbuena, J. A., Míguez-Lozano, R. & Blasco-Costa, I. PACo: a Novel procrustes application to cophylogenetic analysis. *PloS one* **8**, e61048 (2013).

114 Conow, C., Fielder, D., Ovadia, Y. & Libeskind-Hadas, R. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms for Molecular Biology* **5** (2010).