SPECIFIC AIMS:

Zoonotic coronaviruses are a significant threat to global health, as demonstrated with the emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002, and the recent emergence Middle East Respiratory Syndrome (MERS-CoV). The wildlife reservoirs of SARS-CoV were identified by our group as bat species, and since then hundreds of novel bat-CoVs have been discovered (including >260 by our group). These, and other wildlife species, are hunted, traded, butchered and consumed across Asia, creating a largescale human-wildlife interface, and high risk of future emergence of novel CoVs.

To understand the risk of zoonotic CoV emergence, we propose to examine **1)** the transmission dynamics of bat-CoVs across the human-wildlife interface, and **2)** how this process is affected by CoV evolutionary potential, and how it might force CoV evolution. We will assess the nature and frequency of contact among animals and people in two critical human-animal interfaces: live animal markets in China and people who are highly exposed to bats in rural China. In the markets we hypothesize that viral emergence may be accelerated by heightened mixing of host species leading to viral evolution, and high potential for contact with humans. In this study, we propose three specific aims and will screen free ranging and captive bats in China for known and novel coronaviruses; screen people who have high occupational exposure to bats and other wildlife; and examine the genetics and receptor binding properties of novel bat-CoVs we have already identified and those we will discover. We will then use ecological and evolutionary analyses and predictive mathematical models to examine the risk of future bat-CoV spillover to humans. This work will follow 3 specific aims:

Specific Aim 1: Assessment of CoV spillover potential at high risk human-wildlife interfaces. We will examine if: **1)** wildlife markets in China provide enhanced capacity for bat-CoVs to infect other hosts, either via evolutionary adaptation or recombination; **2)** the import of animals from throughout Southeast Asia introduces a higher genetic diversity of mammalian CoVs in market systems compared to within intact ecosystems of China and Southeast Asia; **We will interview people about the nature and frequency of contact with bats** and other wildlife**; collect blood samples from people highly exposed to wildlife;** and **collect a full range of clinical samples from bats and other mammals in the wild and in wetmarkets;** and screen these for CoVs using serological and molecular assays.

Specific Aim 2: Receptor evolution, host range and predictive modeling of bat-CoV emergence risk. We propose two competing hypotheses: **1)** CoV host-range in bats and other mammals is limited by the phylogenetic relatedness of bats and evolutionary conservation of CoV receptors; **2)** CoV host-range is limited by geographic and ecological opportunity for contact between species so that the wildlife trade disrupts the 'natural' co-phylogeny, facilitates spillover and promotes viral evolution. We will develop CoV phylogenies from sequence data collected previously by our group, and in the proposed study, as well as from Genbank. We will examine co-evolutionary congruence of bat-CoVs and their hosts using both functional (receptor) and neutral genes. We will predict host-range in unsampled species using a generalizable model of host and viral ecological and phylogenetic traits to explain patterns of viral sharing between species. We will test for positive selection in market vs. wild-sampled viruses, and use data to parameterize mathematical models that predict CoV evolutionary and transmission dynamics. We will then examine scenarios of how CoVs with different transmissibility would likely emerge in wildlife markets.

Specific Aim 3: Testing predictions of CoV inter-species transmission. We will test our models of host range (i.e. emergence potential) experimentally using reverse genetics, pseudovirus and receptor binding assays, and virus infection experiments in cell culture and humanized mice. With bat-CoVs that we've isolated or sequenced, and using live virus or pseudovirus infection in cells of different origin or expressing different receptor molecules, we will assess potential for each isolated virus and those with receptor binding site sequence, to spill over. We will do this by sequencing the spike (or other receptor binding/fusion) protein genes from all our bat-CoVs, creating mutants to identify how significantly each would need to evolve to use ACE2, CD26/DPP4 (MERS-CoV receptor) or other potential CoV receptors. We will then use receptor-mutant pseudovirus binding assays, *in vitro* studies in bat, primate, human and other species' cell lines, and with humanized mice where particularly interesting viruses are identified phylogenetically, or isolated. These tests will provide public health-relevant data, and also iteratively improve our predictive model to better target bat species and CoVs during our field studies to obtain bat-CoV strains of the greatest interest for understanding the mechanisms of cross-species transmission.

RESEARCH STRATEGY

A. SIGNIFICANCE:

General: Severe Acute Respiratory Syndrome, like many other emerging human pathogens (*1*), originated in a wildlife reservoir host, initially thought to be terrestrial mammals (*2*), and later shown by our group to be bats (*3*). Bats harbor the most closely-related viruses to SARS-CoV, and are traded widely for food in the wildlife markets of China (*4*). The diversity of bat-CoVs is very high, and some studies even suggest that the *Coronaviridae* originated within bats (*3, 5-9*). Recently a novel CoV emerged in the Middle East (MERS-CoV) (*10*) and available data (including from our group) suggest that MERS-CoV also has bat origins (*11-13*). Given that hunting and eating of bats continues across Asia, future spillover of bat-CoVs is likely. Yet *salient questions remain: How diverse are bat-CoVs? Can the conditions in these markets enhance bat-CoV evolution and spillover of bat-CoVs?* **The proposed work** addresses these issues and examines viral diversity in a critical zoonotic reservoir (bats), at sites of high risk for emergence (wildlife markets) in an EID hotspot (China) (*14*).

SARS and bat-CoVs: Coronaviruses are found in a wide range of animal species (*15*). Before the SARS epidemic, only two human coronaviruses (HCoVs) had been characterized (HCoV-229E and HCoV-OC43) (*16, 17*). Since then three more human coronaviruses (HCoV-NL63, HCoV HKU-1, and MERS-CoV), in addition to SARS-CoV, have been identified in individuals with respiratory infections (*16, 18, 19*). One of these, HCoV-NL63, is thought to be zoonotic and of bat origin (*6*). Our group recently identified a CoV from bats in Bangladesh closely-related and likely ancestral to HCoV-OC43 (*20*) and is currently characterizing CoVs from bats in Saudi Arabia. The animal origins of SARS-CoV were first suspected due to the association among index cases and the trade in wildlife for food (*21*). Initially, civets and other mammals consumed in restaurants in southern China were implicated (*2*), however these species did not exhibit the high seroprevalence and low viral (PCR) prevalence expected from a natural wildlife reservoir of a zoonotic virus (*21*). **In 2004, our group discovered SARS-like (SL) CoVs in free-living wild bat species in China and demonstrated that human SARS-CoV nestled phylogenetically within this group (*4*)**. However, SARS-CoV uses the angiotensin-converting enzyme 2 (ACE2) receptor to gain entry to human cells (*22*), and bat SL-CoVs appeared unable to bind to ACE2. A large number of novel *Alpha-* and *Betacoronaviruses* have since been discovered in Old and New World bats, but few have been isolated (*8, 11-13, 23-27*). In 2012, we isolated and characterized two bat SL-CoVs from *Rhinolophus sinicus* from Yunnan Province, China that use the ACE2 receptor and are closely related to SARS-CoV (**Fig. 1**) (*28*). We found a seasonal shedding pattern for this SL-CoV, with peak prevalence of 30-50%. Bats from this population are hunted for human consumption, posing two crucial questions: **1) What is the risk of these CoVs emerging in humans? 2) Will the conditions that exist in live animal markets in Asia promote further emergence of bat-CoVs in human populations?**

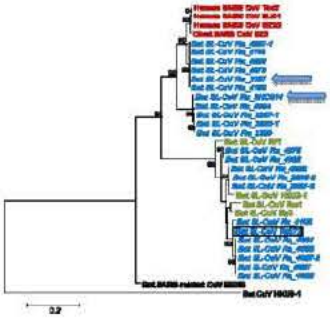


Figure 1. Phylogenetic tree of receptor binding domain sequences of SARS-CoVs (Red), bat SARS-like CoVs discovered by our group in the last 2 years (Blue), and bat SL-CoVs that we published previously in our paper proposing a bat-origin for SARS in 2005 (Green) (*3*). In 2012, we isolated two novel SL-CoVs (SL-CoV-SHC014 and 3367, blue arrows) and have shown for the first time that a bat SL-CoV use the ACE2 receptor which SARS-CoV uses to infect human cells. Unpublished data from Ge *et al.* (in review) (*28*).

Evolution, host-virus co-phylogeny and risk of CoV emergence: There is wide variation in the propensity of viruses for cross-species transmission, within and among viral genera and families (*29*). Coronaviruses undergo genetic recombination by a genomic template-switching mechanism and generate point mutations at a rate similar to that of other RNA viruses, perhaps explaining their capacity for host switching and zoonotic transmission (*15, 30*). This capacity is heightened by the ecology of host species, opportunity for contact, characteristics of the pathogen, and evolutionary (phylogenetic) relationships between hosts (*31-33*). Bats (Order Chiroptera) are the second most diverse group of mammals (~1,200 species) with a wide range ecological and life-history traits that affect their ability to share viruses (*34, 35*) and may explain variation in viral diversity (*36, 37*). Phylogenetic relationships may determine limits to viral binding at receptor sites and to cross-species transmission (*31, 33*), and these factors could be used to predict the risk of spillover (see **Specific Aim 2**). Apart from our own work (see **Section C2b, Fig. 7**), bat and CoV co-evolutionary patterns haven't been rigorously examined. Recent work suggests that most bat-CoV clades correspond to specific bat species or genera (*38, 39*), with little evidence of bat-CoV spillover among species roosting together in the same cave (*40*). There is also evidence for geographically distributed, but related, bat taxa sharing related CoV strains (*8, 38*). In contrast, other studies of wild-caught bats did not find strict co-evolutionary congruence in bat-CoVs for host

species, genera or families (*41-43*). Thus, the same CoV strains may circulate in different bat genera (*41*), and multiple diverse CoV lineages can be found in the same bat species and even individuals (*7, 40, 44, 45*). This, and density of some bat species populations, suggests that viral recombination may be possible in these hosts (*6*). Forced contact in wildlife markets could also facilitate recombination, and may explain divergent Gammacoronavirus strains ancestral to those in birds, in two mammals species in Southern Chinese wetmarkets (*46*). **In this proposal**, we will look for generalizable patterns among bat species and the CoV genotypes they harbor, and use this to examine how phylogeny and contact affect CoV spillover risk.

Host-CoV interactions: an evolutionary approach: The interaction between CoV receptor binding domains (RBDs) and host receptors, e.g. ACE2 for SARS-CoV; dipeptidyl peptidase 4 (DPP4) for MERS-CoV; carcinoembryonic antigen-related cell adhesion molecules (CEACAM) for mouse hepatitis virus; and aminopeptidase N (APN) for hCoV-229E, is critical to understanding limits to host species range (*47-52*). Bats have highly diverse ACE2 receptors at a nucleotide and especially protein level (**Fig. 2**). This is in contrast to other viral receptors in bats, e.g. Ephrin-B2 receptors for henipaviruses (*53, 54*), and DPP4 for MERS-CoV appear to be highly conserved (*51*). Several different genera of bats (e.g. *Myotis*, *Rhinolophus*, and *Rousettus*) have receptors that support viral mediated entry by the SARS-CoV Spike protein (*52, 55*).

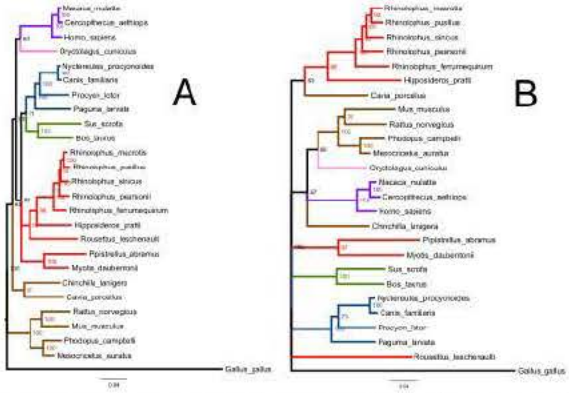


Figure 2. Mammal ACE2 phylogeny using: A) nucleotide data (~2400bp) of ACE2 gene; B) translated protein sequences of same ACE2 genes. All mammal species with available data, including primates (purple), lagomorphs (pink); carnivores (blue); ungulates (green), rodents (brown), and bats (red). Bats are monophyletic and species group with expected taxonomic relationships using nucleotide sequence data (A); but they are paraphyletic a when analyzing protein-level differences (B). This shows functional ACE2 diversity may differ from nucleotide data, and a need to better characterize receptor diversity in a wider range of hosts.

While our preliminary results suggest interesting patterns in bats (**Fig. 2**), the limited number of bat ACE2 sequences precludes robust comparison of co-phylogenetic patterns. **In this study**, we propose to sample dozens of species more than 5 bat families in China, and compare sequence with bats we've sampled globally. This will allow us to build a testable, phylogenetically informed models to examine the extent of, and limits to, batCoV host-range; and will allow us to analyze other receptors like DPP4 for MERS-CoV.

Modeling risk of human infection: The use of mathematical, computational models of viral dynamics has become a standard tool to understand risk of pathogen emergence and spread (*56-60*). However, models that characterize the risk of wildlife-to-human infection require data on contact among populations (*61*), evolutionary constraints of pathogens (*29, 62*), and diversity of novel pathogens (*63*). Because these datasets are usually unavailable, mathematical models can often be theoretical, and of reduced value in predicting risk of pathogen spillover and spread. Building on our group's experience in modeling disease emergence (*64-67*), we will develop a mathematical model that explicitly describes the transmission dynamics and evolutionary dynamics of CoVs in wildlife markets and in bat caves. These models will be parameterized with data we have already collected, and new data from this study, to predict whether novel CoV strains we discover are likely to emerge.

Tests of host range *in vitro*: Receptor usage in different animals is a primary determinant of viral host range. However, while the receptor and receptor binding domains (RBDs) of human-infecting CoVs have been studied intensively, bat-CoVs have not (*22, 47*). In this study, we will determine the RBD of bat-CoVs, develop pseudovirus assays (*68*), and work with a humanized mouse model expressing ACE2 receptor. This provides a way to experimentally test hypotheses on the host-range of novel coronaviruses, even from sequence data. However, despite a plethora of novel CoVs in the recent literature (*38, 39, 44, 45*), there has been little work towards this goal. Furthermore, the recent discovery of MERS-CoV, which uses DPP4, and the use of other receptors for other CoVs (*69*) suggest that this work will be highly significant for other CoVs.

B. INNOVATION:

This project is an innovative fusion of virology, ecology, and mathematical modeling. The analysis of CoV genetic diversity in bats and other mammals in southern China, combined with characterization of and co-phylogenetic analysis with CoV functional genes (e. g. ACE2, receptor of SARS-CoV and DPP4 for MERS-CoV) has not yet been attempted, and **will allow us to better understand the patterns of host-switching**. Previous studies using molecular clock analysis have found that the bat SARS-like-CoV to civet/human SARS-CoV divergence

ranged from 7-17 years (mean 4.9) before the 2003 outbreak (*7, 70*). We will use a novel **phylogenetic and mathematical modeling approach to examine how dynamics of contact and pathogen transmission among hosts in markets drives viral evolution and emergence**. We will determine **how many years it takes for a coronavirus to evolve an R_0 >1 and therefore have epidemic potential** using a modeling framework that combines evolutionary changes and multi-host dynamics (Specific Aim 3), expanding on published approaches (*71-73*). We will then **simulate coronavirus emergence under different market conditions to identify most likely scenarios that can inform strategies to prevent future outbreaks.** Finally, we will use pseudovirus binding assays, *in vitro* infections and humanized mice expressing ACE2 to test our analyses on the novel viruses we have, and will, identify.

We will use our multidisciplinary approach to examine fundamental questions on how the wildlife trade, wetmarkets and other target interfaces promote the emergence of coronaviruses, and **assess the risk of future spillover of CoVs from bats or other mammals and humans**. In particular, despite 10 years since the emergence of SARS and the discovery of 60+ novel bat-CoVs, three significant issues remain unanswered: **1) What are the natural limits to CoV host range, and can this be predicted by the host-receptor-virus relationship; 2)** Are the conditions in wildlife markets sufficient to allow enough interspecies transmission that coronaviruses **can evolve the ability to infect new hosts, including humans**, either by accumulation of point mutations, or by recombination events; or **3)** Is the expansion of the wildlife trade **bringing expanded diversity of CoVs into the enhanced human-animal interface present in wet markets**?

C. APPROACH

C1: Specific Aim 1. Assessment of CoV spillover potential at high risk human-wildlife interfaces:

C1a) General strategy and supporting studies: SARS-CoV emerged in live animal markets in Guangdong, with unrelated spillover events in at least five of seven municipalities, suggesting widespread introduction into wildlife markets within this city (*21*). We propose to characterize the species composition of bats and small mammals in wildlife markets where there is a high degree of contact between animals and people. We will identify additional high risk interfaces that may occur in southern China such as guano collection, which we have recently identified as a potential CoV exposure risk in Thailand (*12*). We will interview people at high-risk interfaces and who are enrolled in acute respiratory or influenza-like illness surveillance programs conducted by our colleagues at CDCs in Shanghai, Guangdong, Yunnan, and Guangxi. These data will be used to parameterize the contact process (χ) in our mathematical model of CoV emergence (**see Aim 3**).

We will assess **1)** whether market conditions provide enhanced capacity (increased **evolutionary opportunity**) for bat-CoVs to evolve the ability to infect other hosts, either via repeated inter-species transmission, positive selection or recombination events; and **2)** whether the intake of wildlife from Southeast Asia by China introduces a greater diversity of hosts and a correspondingly diverse group of CoVs (increased **ecological opportunity**). We will conduct CoV pathogen discovery in samples from humans and wildlife at these sites and examine their receptor binding domains to identify their ability to bind to ACE2, DPP4, or CEACAM receptors in humans. We will compare CoV diversity in China with that in wildlife across Southeast Asia (from our current work on other funded programs, and published data) that may potentially enter China's wildlife trade. Data from this aim will be used to assess the likelihood of inter-species bat-CoV transmission (see also **Specific Aim 2**).

Working in high-contact human-wildlife interfaces can be challenging. However, we have already collected significant preliminary data to accomplish Aim 1. We have located and surveyed wildlife markets in Yunnan, Guangdong, GuangXi and Fujian provinces, and have identified populations that hunt and consume bats in Yunnan province. We have begun to characterize the species composition of free-ranging bat populations and have collected samples **from over 1000 bat individuals (28 spp.) from 35 localities in over 15 (two-thirds of all) Chinese Provinces**. We will also utilize archived wild bat, rodent, and civet samples collected by our team in Malaysia, Thailand and Indonesia on another large federally-funded project to provide samples of species regularly imported into China (**section C1b**) (*21, 74*).

Wildlife Markets: Ten years following the SARS-CoV outbreaks, there is little information available on the *current* diversity of bats and other mammals available in the wet markets in southern China. One study found that 91 species of vertebrates, including 40 mammal species, were being traded in Guangxi, China (*75*). Further, little data is available on the origin of wild animals brought into the market system. In some cases, animals may be locally collected, while in other cases animals may be imported from Southeast Asia, including adjacent Vietnam (*74-76*) – factors which will affect the diversity of CoVs. Captive and free-ranging rodents are found in markets and may be an additional host for CoVs (*77*). We have worked with Yunnan Institute of Endemic Diseases Control and Prevention since June 2012 (see **Letters of Support**). We have conducted initial surveillance in Nujiang, Baoshan Denong and Xishuangbanna prefectures and Ruili, which is a major wildlife trade gateway between Myanmar and China **(Fig. 3)**. We have collected 187 small mammals from markets in Yunnan and tested them for coronaviruses using a 1-step PCR assay (*78*), finding 2/21 shrews (*Crocidura attentuata*) are CoV-positive.

Figure 3: Map of wildlife trade routes from Southeast Asia into China. Modified from (*79*).

Other animal samples available for this project: To date, our group has collected more than 90,000 high quality specimens from 15,000 animals representing key wildlife reservoirs for zoonoses such as bats, rodents and primates under our USAID-EPT PREDICT project. Clinical samples include blood, throat swabs, feces and urogenital swabs and represent animals from 10 different countries including Bangladesh, India, Malaysia, Thailand, Indonesia, China, Brazil, Bolivia, Colombia, Peru, and Mexico. 50,000 of these samples originate from Asia, and are currently being screened for novel coronaviruses (**See Section C2a, Fig. 6**). **We have also collected more than 500 bat specimens representing seven species from the Kingdom of Saudi Arabia in collaboration with Saudi Arabia's Ministry of Health and Columbia University.** Nearly 20,000 of our samples come from bats, and will be used to analyze CoV diversity along with novel CoVs we identify.

Identifying novel CoVs in wild bats in China: We have already conducted significant CoV surveillance in China for bats, other wildlife and humans. For this, we use pan-coronavirus PCR protocols based on conserved RNA-dependent RNA polymerase (RdRp) motifs A and C to screen samples at Wuhan Institute of Virology (*80*). Besides a large number of SL-CoVs, we have detected several novel bat-CoVs including strains closely related to CoV HKU4/5, CoV 1A &1B, CoV HKU 2, 6, & 8. For the first time, we have also isolated and characterized a bat-CoV from China that uses ACE2 receptors (see **Section C3a preliminary data**) (*28*). **In all, we have identified sequences from 268 novel bat-CoVs (140 from China alone) from bat species collected in Bangladesh, Thailand, Mexico, Brazil and China (See Section C2a, Fig. 6).** We have an additional 5,000+ clinical samples from free-ranging bats and rodents from Guangdong province, from an ongoing study which are being screened for viral pathogens, including CoVs at Guangdong Entomological Institute.

Survey of people highly exposed to wildlife in Guangdong, China: We have worked with Guangdong CDC since 2008, under a currently active IRB protocol, to interview and sample people working in live animal markets, hunters and restaurant workers with a high level of exposure to animals. We have interviewed volunteer participants about the nature and frequency of animal interactions; collected biological samples (blood, feces, sputum), and trained participants to collect animal blood samples (dried blood spots on filter paper) from animals they butchered or hunted. We enrolled 1300 participants across 12 sites within Guangdong Province (**Fig. 4**).

Figure 4: Sites of current human sample collection by Guangdong CDC for zoonotic pathogen surveillance in Guangdong Province, Southern China. Each star represents a large wildlife market where we have enrolled market and restaurant workers (total = 1,300) for our zoonotic pathogen spillover study. Seventeen people had IgG antibodies to SARS-CoV and a follow-up study is underway.

Samples have been tested for antibodies to animal pathogens, including SARS-CoV. **Of the 1300 serum samples screened using a SARS-CoV ELISA, 17 were positive for IgG antibodies to SARS-CoV.** These patients were not acutely ill at the time of sample collection, and this finding suggests one of three possibilities: 1) that SARS-CoV is still circulating in Guangdong markets; 2) that these people may have been exposed during the time of the 2002-3 outbreak; or 3) that the ELISA used is cross-reacting to another CoV. Review of their history of wildlife exposure is currently underway. In Shanghai, the Shanghai Municipal Center for Disease Control and Prevention (see **Letters of Support**) currently conducts surveillance on people with influenza-like illness in rural communities surrounding Shanghai. We will develop a similar study of people in

these communities who have exposure to wildlife. We will review and re-screen archived blood samples at Guangdong CDC for other bat coronaviruses once we determine candidates that could likely infect humans. to see whether there is exposure to CoVs other than SARS. We will re-screen these samples with specific serological assays based on bat-CoVs that will help differentiate between SARS-CoV IgG and other bat-CoV IgG to see whether there is exposure to CoVs other than SARS (*3, 81*). We will expand our survey to Guangxi, Fujian, Shanghai and Yunnan provinces to survey regions where SARS-CoV was not reported, but where wildlife trade, hunting, and bat guano collection is common.

C1b) Market characterization, wildlife sampling and human surveys: We have conducted surveillance at the wildlife markets of Guangdong where early cases of SARS-CoV were identified. From 2011-2013 we interviewed and sampled animal vendors, hunters and restaurant workers who butcher wildlife (**See Section C1a, Fig. 4**). For this proposed study, we will identify 10 markets in Guangxi, Yunnan, and Fujian Provinces (**Fig. 5**). We will characterize the physical size, number of vendors, diversity and abundance of mammalian species in each market. A questionnaire will be developed based on the one we used in Guangdong, to collect data on the nature and frequency of animal exposure of people who work in markets or hunt wildlife. We will conduct interviews to determine which bat species are sold, typical numbers, and source locations. We will collect information about recent acute respiratory illness and include those who have had undiagnosed acute respiratory symptoms within 3 months of the survey. We will then screen volunteers from this cohort for bat-CoV antibodies using existing and newly developed assays. We will compare exposure rates between people who are highly exposed to wildlife and a control group from the same regions.

Figure 5: Proposed sampling sites in Southern China (Guangdong, Guangxi, and Fujian Provinces) for the current study. Arrows indicate wildlife trade routes. Letters indicate wild animal markets in Guangdong (A-R), Guangxi (S-W), Hunan (X) and Fujian (Y).

In Shanghai, where wildlife markets are less common than southern provinces, we will interview voluntary participants under surveillance by Shanghai CDC for influenza-like illness. We will compare CoV exposure rates in people with acute respiratory illness to a control group from the same region (**see letter of support**).

Wildlife sampling: We will locate wild bat populations used to supply local markets in Yunnan, Guangdong, Guangxi, and Fujian. We will sample a minimum of 30 individuals from 30 different bat species representing but not limited to the following families: *Rhinolophidae, Hipposideridae, Vespertillionidae, Mollossidae,* and *Pteropodidae*, all of which are known to carry *alpha-* or *betacoronaviruses* and are consumed by people (*4, 7, 82*). Bat SL-CoV PCR prevalence is 10%-38% (*4, 24*). Given 10% prevalence in bat populations, sampling 30 individuals would ensure a CoV detection probability of 95%. In all wildlife markets, we will opportunistically sample a variety of insectivorous and frugivorous bats, and other mammals if available, taking fresh feces or rectal swabs, saliva (oropharyngeal swab), and blood. A small number of bats will be sacrificed as vouchers and to collect intestinal tissue for CoV receptor analyses if required. We will use cyt-b to identify host species.

Human exposure to CoVs study: Expanding on our work in Guangdong, we will develop a voluntary study of animal vendors and hunters in Guangxi, Yunnan, and Fujian provinces in cooperation with local Bureaus of Public Health and CDCs. We will develop a survey to identify people with high exposure to wildlife, particularly bats, and will recruit volunteers, collect blood, sputum, and stool sample from each enrolled participant. We will screen sera for antibodies to SARS-CoV, other alpha & beta coronaviruses including MERS-CoV, and novel bat-CoVs. We will screen stool from CoV seropositive participants for CoV nucleic acid. We will also develop specific bat-CoV serological assays and share these with our Chinese collaborators. In each province in southern China we will aim to include 10 markets and survey 20 vendors per market; 20 additional wildlife hunters per province (220 case subjects); 400 control subjects from the general population near the markets in each province (total of 620 people per province). For Shanghai, we will enroll 200 acute respiratory illness cases and 400 non-respiratory controls (600 total), The total number of human subjects will be 2460. The study will be conducted in Guangxi, Yunnan, Fujian and Shanghai provinces (see **Section E, Human Subjects**).

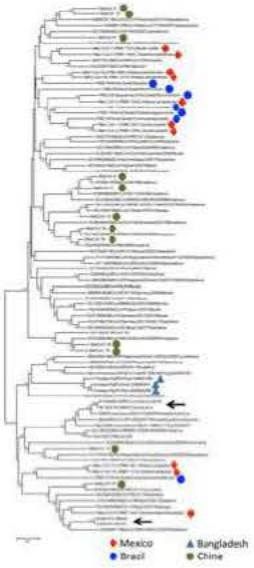
C1c) Data analysis: Human sera and stool samples will be tested at provincial CDC labs (see **letters of support**) and animal samples will be screened at the Wuhan Institute of Virology (Co-I, Shi)**.** Serum or plasma samples will be tested for CoV antibodies using ELISAs specific for SARS-CoV and bat SL-CoVs that we have developed (*4, 68, 83*). Fecal and saliva samples will be tested for CoV viral nucleic acid using a series of pancoronavirus PCR assays that target a region in the RdRp that is highly conserved among coronaviruses and for which we have a positive control, developed by our group under another federally-funded contract (*13, 23,*

84). The RdRp gene will be sequenced from all positive PCR samples and used to build co-phylogenetic trees (see **Specific Aim 2**). We will also test these pathogens for recombination events in markets vs. wild sampled CoVs after viral strains are characterized. Data from **Aim 1** will be used to parameterize mathematical models of viral transmission (**Specific Aim 3**) in markets to estimate relative risk of emergence depending on different diversity of mammals, contact rates, size of markets, and evidence for human exposure to bat-CoVs.

C1d) Potential Pitfalls and Solutions: We may find lower than expected levels of wildlife diversity in markets in Southern China. If this occurs, we have access to tens of thousands of wildlife samples from over 20 countries globally from work on a current NIAID R01 (Daszak PI) to assess diversity of viral pathogens in bats in Asia and Latin America, a large multi-year contract from USAID (Emerging Pandemic Threats: PREDICT program, Daszak PI) to conduct surveillance and pathogen discovery in wildlife in Asia and Latin America and two Nipah virus R01s. We have already discovered >250 novel CoVs from bats in these countries (**Section C2a**) including >100 from China. A second setback would be that access to markets becomes restricted due to political sensitivities. We are working closely with long-term local collaborators at ECNU and the Institute of Virology, Wuhan, both of which institutions are well respected nationally. The Institute of Virology is the National Center of Excellence for viral pathogens, and has Federal authority for viral research. Furthermore, we have shown through our work with Guangdong CDC that we can conduct long-term collaborations in these sites. Finally, we have selected a large number of wildlife market sites, so the closing of one will not affect all sampling activities.

C2: Specific Aim 2. Receptor evolution, host range and predictive modeling of bat-CoV emergence risk:

C2a) General strategy and supporting studies: *Can we use information on CoV sequence, host sequence and behavioral traits and population dynamics at critical human-wildlife interfaces to predict which CoVs are most likely to emerge?* To answer this, we will use data from our characterization of bat-CoVs, host range, receptor genes, serological data, and from field-collected data to build and parameterize three related models. **First**, using phylogenetic reconciliation we will map the co-phylogenetic patterns of bats and their CoVs using neutral and functional markers (RBDs and host receptor genes). We will compare free-ranging and market-sampled species assemblages and test the related hypothesis that wildlife markets disrupt 'natural' bat – CoV host associations and increase recombination and/or accelerated evolution to facilitate emergence. **Second**, we will construct generalized linear models that encompass phylogenetic information to test the two related hypothesis that spillover potential and host-range of bat-CoVs is limited by: 1) opportunity for contact; or 2) phylogenetic relatedness of host species and their receptor genes. **Third**, we will use mathematical matrix modeling to investigate bat-CoV transmission and evolutionary dynamics, and test the potential of novel CoVs to infect humans, bats, and other market animals. This model will be informed by serological data, market surveys, and receptor binding data from bat cell line and humanized mouse inoculation studies.



Phylogenetic studies of known and novel bat-CoVs: Phylogenetic methods can be used to identify recent host shifts and spillover events of CoVs, often these events are due to anthropogenic changes to host ecology, e.g. *Rhinolophus* spp. and human/civet SARS-CoV in the wildlife trade (*4, 7*). It has been proposed that repeated passage between civets and humans in wet markets facilitated SARS-CoV evolution towards greater human and civet ACE2 receptor affinity (*85*), and accelerated evolution and positive selection in CoVs was detected after host spillover (*86*). It is not known if bat-CoVs follow predictable patterns of co-phylogeny between host and virus; many studies found unique CoV strains circulating in different bat lineages, but also multiple CoV strains have been identified in the same bat species and individuals (*7, 40, 44, 45*).

Figure 6 (above): Phylogenetic tree (RdRp) of selected bat-CoVs from Genbank, including as subset of the 268 novel bat-CoVs discovered by our group through our USAID-EPT PREDICT pathogen discovery work in China, Brazil, Bangladesh and Mexico.

Wildlife trade and market dynamics may promote the cross-species transmission of distinct bat-CoV strains and facilitate viral recombination within these hosts (*46*); the extent of this will depend on the role of host phylogeny vs. contact in limiting bat-CoV spillover. Using our extensive database of bat and other wild animal CoVs that we have characterized, isolated, or are available on Genbank, we will examine these constraints for known and novel CoVs we identify. Over the past four years, we have conducted large surveys of bat pathogens globally, including the discovery of sequences from 268 novel bat-CoVs (including 140 from China) (**Fig. 6**).

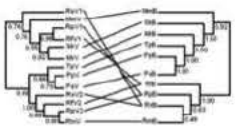


Figure 7 (left): Host-pathogen co-phylogeny of bat-CoVs from China (*43*). Bat genera: R, *Rhinolophus*; Mm *Miniopterus*; Mr, *Myotis*; P, *Pipistrellus*; V= viral sequence, B= bat sequence. This figure suggests rhinolophid CoVs may have a greater ability to jump hosts. Warrants further investigation using functional genetic markers and data from more species.

C2b) Co-phylogenetic analysis of bat-CoVs: We will use coronavirus and host sequence data generated in this project, from archived samples that we collected from bats just after the SARS outbreak, and previously published CoV strains from a diverse range of host species to quantify co-evolutionary patterns and host range in bat-CoVs. Combined analyses of host and viral phylogenies **will allow us better understand if host phylogeny (and receptor gene similiarity) can predict CoV host switching and whether or not market systems have disrupted the "natural" patterns of CoV association** (e.g. **Fig. 7,** from (*43*)). We will reconstruct phylogenetic relationships of CoVs using a combination of the HEL, N, RdRp, and S genes, as each has a different evolutionary rate and will allow us to test patterns of cophylogeny at different taxonomic scales. We will reconstruct host species relationships from tissue collected in our study using both neutral (mitochondrial and nuclear, e.g. cytB and RAG2) and functional (e.g. ACE2 CoV receptor) host genetic markers. Multiple alignments will be performed MAFFT (*87*), and phylogenies estimated using maximum likelihood (*88*) and Bayesian inference (*89*) for each viral and host gene, and concatenated virus datasets when no viral recombination is detected. We will test for statistical significance using ParaFit implemented in CopyCat (*90*) and AxPcoords (*91*), and visualize these using TreeMap v2.02β (*92*). These methods will allow us to identify which particular host-virus associations contribute most to the observed patterns. We will partition our dataset by collection localities and higher-level taxonomic groups to test co-phylogenetic significance at multiple spatial and taxonomic scales. To test the null hypothesis that there is no pattern of co-evolution we will perform permutations to randomized hosts–virus associations and then measure congruence relative to the host tree. By comparing the patterns of host-CoV co-phylogeny in natural bat communities (cave sites) vs. wet markets, we will be able to identify anomalies that may likely signal recent spillover events. To test for genetic recombination in market vs. wild-collected bat-CoVs, we will use sliding window analysis (*7*) and RDP3 v3.44 software (*93*). We will use previous methods to test for positive selection and identify specific virus residues under selective pressure (*94*).

Quantifying CoV strain sharing between host species: We will use viral sequence data from RdRp and S genes to delimit unique CoV "species" or "genotypes" at different taxonomic and sampling levels. We will test for non-random patterns of association of viral community assemblages between species (*95-97*) (*98*). This will involve calculating Jaccard's index of similarity (J) for the viral assemblages between pairs of species and testing for deviations from that expected by random chance using Monte Carlo randomizations (*99*). Deviation from the null model will be calculated as the difference between the mean J observed (J_{obs}) in the data and the mean J expected, such that $J_{dev} = J_{obs} - J_{null}$. Positive values of J_{dev} will thus indicate that CoV 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.

C2c) Predictive model of CoV host-range and diversity: We will develop a predictive model of host-range for bat-CoVs using data of bat distribution in natural caves and the markets, geographic ranges, ecological and behavioral characteristics of host species from our field studies and the literature, host and viral phylogenies, and associations of host species to particular CoV strains/clades. We will include phylogenetic distance between bat species and other mammal hosts from various neutral and receptor genes generated in this study. We will use CoV similarity indices (Jaccard, above) as our response variables in multiple regression models, i.e. generalized linear models (GLMs) and phylogenetic generalized linear mixed models (PGLMMs) with relevant bat ecological, phylogenetic, morphological, behavioral, and life history traits as our predictor variables, to assess the relative contribution of host phylogeny, viral traits, or species-specific ecological traits in explaining CoV diversification and sharing. We will calculate indices of host specificity that account for host phylogeny (*100, 101*), to further test hypotheses of whether bat-CoVs are more likely shared between host ecological groups or among species with similar life-history traits vs. relatedness. All statistical analyses will be conducted in R with relevant packages for community ecology and species diversity (vegan, fossil), and phylogenetic modeling (ade4, ape).

Extension of this model beyond China will allow us to map **a global spatial and phylogenetic risk gradient for CoV emergence based on host species traits, mammalian phylogeny (including functional CoV receptor genes), and relatedness of CoVs**. Further, we can use the results from our logistic regressions to identify gaps in surveillance, where bat species are found to share a lower than expected number of CoV strains given a threshold level of contact and relatedness with other host species. We will test our predictions of host range from the analytical model for bat-CoVs using synthetic reconstruction of bat-CoVs and *in vitro* studies of ortholog

receptor binding with different mammalian cell lines (**Aim 3**). Specifically, we will evaluate the ability of novel bat-CoVs to recognize and bind to selected receptors (ACE2, CEACAM, APN, receptor for alpha-CoV, or DPP4/CD26, receptor of MERS-CoV) reconstructed from divergent bat taxa. We envision an iterative process over the first few years of this grant whereby initial data are generated from known host-CoV associations, results from the model will be tested experimentally, and then data from experimental studies will be used to refine the models and better inform field sampling in China and globally.

Analyses of literature database: We have built a database of virus-host associations for 131 bat species and all 50 unique ICTV recognized bat viruses. We used a logistic GLM regression approach with host and virus variables, and found that host phylogeny (i.e. phylogenetic distance to other bat host species) was a strong predictor of observed virus sharing across bat species (trend with phylogeny only shown in **Fig 8**). We will adapt this approach by using host genetic distance of functional receptor genes instead of neutral markers, and CoV data collected from our standardized survey efforts.

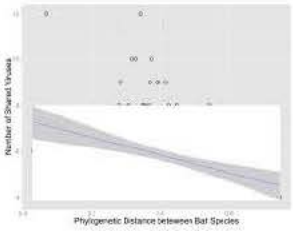


Fig 8. Scatterplot showing a decrease in the number of shared viruses with decreasing phylogenetic relatedness among bat species. Dataset includes all bat species pairs with >3 shared viruses for ~200 bat-virus associations form the literature (Olival, unpublished). Pairwise phylogenetic distance from maximum likelihood tree using cytochrome B mtDNA data.

C2d) Modeling the dynamics of CoV spillover risk: A key question in EID research is the role of viral evolution in enabling pathogen emergence. While some EID pathogens cause epidemic or pandemic disease because they readily transmit among humans ($R_{0,Human} > 1$, e.g., HIV, A/H1N1pdm), or only spillover directly from animals ($R_{0,Human} = 0$, e.g. West Nile Virus). Others, including MERS-CoV, may spillover regularly to humans, and even cause small clusters of human-to-human transmission, but have not yet caused a major epidemic or pandemic ($1 > R_{0,Human} > 0$, e.g., Nipah virus, monkeypox, Influenza H5N1). A looming issue is the likelihood of such a pathogen evolving to become a major epidemic or pandemic (i.e., $R_{0,Human} > 1$). Divergence times between ancestral bat-CoVs and hCoVs can vary widely and provide a timeline of past spillover events, e.g. 560+ years between hCoV-NL63 and its progenitor alpha-CoV (*6*) and ~20 years between bat SARS-like CoVs and human or civet SARS-CoV (*7, 70*).

The limits on SARS emergence are still unclear: Were the bat SL-CoVs unable initially to bind to human receptors, or was it necessary for a precursor CoV to evolve and adapt to humans for SARS-CoV to emerge? Were civets or similar non-bat, non-human hosts a critical intermediate evolutionary step in the transition from bats to humans, or were they incidentally infected along with humans simply by virtue of similar receptors? To examine the timeline for different emergence pathways, we have built a model framework (below) to represent the wildlife market environment and include viral ecology and evolution. We will use a matrix framework (*72, 102*) to determine how the pathogen is transmitted among different host species and between locations. We have already built the framework of this model (below), and have listed the data that we will collect in the current study to parameterize it (**Table 1, below**). To incorporate strain variation and evolution, we will adapt the approach of Antia *et al.* (*71*) by integrating a branching process approach to our matrix framework. **We will use these techniques to develop "What If" scenarios that predict how different strains of CoV would emerge, and potentially evolve, in different market systems within Asia and elsewhere** (e.g. scenarios with different host diversity and different levels of host-host and host-human contact within markets).

To examine strain evolution, we will model ***n*** possible strains, where strain 1 is the initial variant, and strain ***n*** is the variant that has a human $\mathbf{R_0}$ >1, with ***n-2*** variants in between, which may each have their own $\mathbf{R_0}$ that depends on the host community using 'Who-Acquires-Infection-From-Whom' (WAIFW) matrix framework (below). Following Antia *et al.* (*71*), we assume the mutation rate, μ, is the same for all variants, that only single mutations can occur, and we ignore back-mutation. However, we will reconsider these assumptions if changes in these can alter the expected outcomes of the mathematical results. We also assume that the total number of secondary infections generated by an individual with variant ***i*** is Poisson distributed with mean $\mathbf{R_0^{(i)}}$. A proportion μ of the variants will have mutated or recombined into type ***i+1***, while the proportion (1- μ) remains the same, as type ***i***. We will separate out the cases of mutation and recombination by placing different restrictions on the changes that could occur in the strains as they move from type ***i*** to type ***i+1***. These assumptions result in the probability generating functions:

$$f_i(s_1, s_2, \ldots, s_m) = \exp\left(-(1-\mu)R_0^{(i)}(1-s_i)\right)\exp\left(-\mu R_0^{(i)}(1-s_{i+1})\right) \quad \text{for} \quad i < m$$

(5) $$\text{otherwise} \quad f_m(s_1, s_2, \ldots, s_m) = \exp\left(-R_0^{(m)}(1-s_m)\right)$$

Through this branching process approach we can gain insight into the limitations and possibilities that stochastic processes may impose on the evolution of strain diversity in both limited diversity settings (e.g., only bats and humans), and highly diverse environments (e.g., markets with other hosts such as civets and bamboo rats). We can also adapt this methodology to compare mutation, which we expect to take small incremental movements in a fitness landscape that may have low fitness valleys between a wild-host adapted strain and a human or other host adapted strain, and recombination which may be able to take larger leaps across a given fitness landscape. Using this framework we can vary the $\mathbf{R_0}^{(i)}$ depending on the fitness of the mutants in various hosts, and the host diversity and abundance, simulating the complex fitness landscapes of real CoV systems. We can calculate the number of secondary hosts infected as $\mathbf{R_0} = \chi\phi\tau$, where τ is the duration of infectiousness, and χ is the rate of contact. Our receptor binding studies and predictive GLM models of host range will be used to inform ϕ, the joint probability that a susceptible host becomes infected when exposed. We will model our system both mathematically from a simple $\mathbf{R_0}$ perspective for insight, as well as using a spatial stochastic-birth-death simulation implementation to understand the implications of multiple scales of variation, including mutation and recombination and the implications for stochasticity for CoV emergence. To do this we will expand our basic equation, $\mathbf{R_0} = \chi\phi\tau$, into a matrix formulation to incorporate the multiple hosts within this system. Each strain and spatial location (e.g., market), can be represented by a different matrix. Thus we have:

$$X_k = \begin{bmatrix} \chi_{1,1,k} & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \chi_{i,j,k} \end{bmatrix} \quad \Phi_k = \begin{bmatrix} \phi_{1,1,k} & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \phi_{i,j,k} \end{bmatrix} \quad T_k = \begin{bmatrix} \tau_{1,1,k} & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \tau_{i,j,k} \end{bmatrix} \tag{1}$$

which we can use to define a 'WAIFW' $\mathbf{\Omega_k}$ matrix (*73, 103*) of which the eigenvalue gives us an estimate of $\mathbf{R_0}$ for the whole system, for a given strain and location. The 'WAIFW' matrix is:

$$\Omega_k = \begin{bmatrix} \chi_{1,1,k}\phi_{1,1,k}\tau_{1,1,k} & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \chi_{i,j,k}\phi_{i,j,k}\tau_{i,j,k} \end{bmatrix} \tag{2}$$

Critically, this enables us to analyze certain 'what-if' scenarios. For example, we can examine the role of civets in emergence by assuming that the strain which initially infected civets had to evolve in order to then infect humans. This would give us two strains in a single location, each with its own $\mathbf{R_0}$. Alternatively, we can assume that all three SARS-CoV host species (bats, civets, humans) were in the same market place, and a single CoV strain. In this case we would have a single matrix, with all three species, and values in every cell of the matrix. By keeping the separate pieces of the $\mathbf{R_0} = \chi\phi\tau$ equation, in the matrix form, we can examine potential public health control measures (e.g. quarantine, culling or separating species into different market locations) (*104*), which might also vary depending on the nature of receptor binding and strain evolution. To account for assumptions, we will investigate the implications of mixing in a stochastic environment. We have already built a stochastic-birth death, discrete event simulation of the spread of EIDs for Avian Influenza in multi-species markets and farms. We will adapt these simulations for strain and receptor diversity interactions with multiple species of CoV hosts. **This suite of modeling approaches will allow us to integrate our ecological and molecular approaches to understanding the potential pandemic emergence threat posed by the whole suite of bat-CoVs.**

Table 1: Data Needs for Model:

Parameter	Description	Sources
τ_{Human}, τ_{Bat}, $\tau_{Other?}$	Duration of infectiousness, Humans, Bats, other spp.	Humans (*57, 105-108*), Bats (*7*), other species (*108, 109*)
$\phi_{Human->Human}$	Joint probability an infected Human can transmit to susceptible Human	(*57, 105-107*)
$\phi_{Bat->Human}$, $\phi_{Other->Human}$ $\phi_{Bat->Other}$, $\phi_{Bat->Human}$ $\phi_{Other->Other}$	Joint probability an infected host can transmit to susceptible; can use receptor binding in host species for parameterization	*(*109*)
$\phi_{?}$	As above	Generally assume 0 or $\phi_{i,j} = \phi_{j,i}$ *
$\chi_{i,j}$	Contact rates	Market Surveys, using map overlap for non-market areas.
μ	Mutation rate	Literature
ξ	Recombination rate	Literature
N_{Human}, N_{bat}, N_{other}	Population density of bats, humans, other	Market surveys, census & transect data.

* Use knowledge of receptor bindings to appropriately upscale or downscale relative to human-to-human case of SARS and laboratory studies on other animals. We will run sensitivity analyses for these parameters.

We will assume that twice the estimate for SARS $R_{0,Human}$ rounded up to the max of the 95% CI to give 5 or 10 represents a near maximum, and 0 forms a lower boundary. We will assume τ is constant regardless of species and again do sensitivity analysis using SARS-CoV values. We test the following hypotheses: 1) That recombination can either substantially boost (H_A: $\Delta Pr > 0$) or mutation have the same effect (H_A: $\Delta Pr > 0$) on the probability of CoV spillover into humans, or that only recombination and mutation together provide a substantial boost to spillover probability (H_A: $\Delta Pr > 0$); 2) That known (e.g. civets) or unknown intermediate animal hosts or no intermediate hosts are necessary for CoV spillover to humans; 3) That high diversity of intermediate hosts either increases or decreases the probability of CoV spillover into humans. We will use our modeling framework to examine the potential CoV spillover in different markets, using the market data from Specific Aim 1, evolutionary characteristics of the CoVs from Aim 2, and specifically-acquired data to parameterize the model. **Table 1 (above)** lists parameters in the model, and gives available sources for data.

Previous experience of modeling disease emergence: Our group has used mathematical models to test hypotheses on zoonotic disease emergence for over 15 years. We use computational models that are tailored for the specific pathogen type or combination of hosts involved, and parameterize these with extremely detailed datasets specific for the emergence event. We then run simulations to test hypotheses on the spillover of viruses and the emergence of zoonoses. For Nipah virus (NiV), another bat-borne zoonosis, we obtained data from pig production facilities in Malaysia (*110, 111*), from experimental infection of bats and *in vitro* under BSL-4 conditions for viral transmission parameters (*112, 113*). We used this approach to demonstrate the cause of NiV emergence (*111*). We have successfully used similar approaches to demonstrate viable causal mechanisms for the emergence of Hendra virus (*114*), Avian influenza (*115-117*) and West Nile virus (*118-120*).

C2e) Potential pitfalls and solutions: The diversity of coronaviruses that we identify may be inadequate for robust co-phylogenetic analysis. We have already shown proof of concept in preliminary data through USAID and NIAID funded projects that we have detected new coronaviruses in most bat species examined; there has been a large amount of research from several groups showing a broad diversity of coronaviruses; previous studies from us and other groups have provided evidence of a diversity of coronaviruses associated with bats and there is high likelihood that we will identify more. In China specifically, 23% of bat samples we have screened were positive for CoVs, thus we do not anticipate a lack of diverse CoVs (*28*). For modeling studies, not all necessary parameters may be easily obtained. We will use information from the SARS-CoV outbreak, where we have detailed data from the WHO investigations on serology and viral isolation from market wildlife; and from our recent and current work in Guangdong province; and an ongoing study on avian influenza in Shanghai and Guangdong markets (Co-I Zhang). Lastly, for parameters that we cannot actually estimate, we may be able to posit reasonable limits. For example we can constrain the probability of spillover: it must be greater than 0, since SARS did in fact spillover (*106*), but it is very unlikely that this probability is higher than the within species transmission probability. If the rate of transmission within a host species is unestimatable, we can use data from other diseases in similar species, such as bat rabies. Thus we can readily perform a sensitivity analysis for unknown parameters within a range that is biologically plausible, using sensible constraints.

C3: Specific Aim 3. Testing predictions on CoV inter-species transmission:

How can we test predictive strategies to understand which viruses have the capacity to 'jump hosts'? To answer this, we will analyze the interspecies infection or transmission of CoVs we have identified, particularly the SARS-like CoVs and CoV HUK4/5 that is closely related to MERS-CoV (hCoV-EMC) from Saudi Arabia. Our main approach will be: 1) *in vitro* infection experiments using pseudoviruses carrying the spike proteins (wild type or mutants) or live viruses in cell lines of different origins; 2) binding affinity assays between the spike proteins (wild type or mutants) and different cellular receptor molecules; and 3) humanized mouse experiments if viruses are identified of significant human infection potential (see **Ralph Baric, Letter of Support**).

C3a) General strategy and supporting studies: We will sequence the spike (or other receptor binding/fusion) protein genes from all bat-CoVs we identify, creating mutants of these to identify how significantly each would need to evolve to use ACE2 or CD26/DPP4 (receptor for MERS). We will then use receptor-mutant pseudovirus binding assays, *in vitro* studies with a wide range of cell lines from bats, other mammals including primates and human cell lines, and with humanized mice where particularly interesting viruses are identified phylogenetically, or isolated (see **Ralph Baric, Letter of Support**). These tests will provide direct public health-relevant data, and also iteratively improve our predictive model to better target bat species and CoVs during our field studies to obtain bat-CoV strains of the greatest interest for understanding the mechanisms of cross-species transmission.

Experience working with receptor mutants & pseudovirus binding assays: We have established a stable pseudovirus assay for SARS-CoV and SARS-like CoV and tested the infectivity of these spike proteins in cells expressing ACE2 from human, civet and bats (*52, 68*). We have demonstrated that several bat species are susceptible to the SARS-CoV and that some SARS-like CoV strains can use human ACE2 for cellular entry (*52*). Minor mutations in S proteins or ACE2s greatly affected the receptor binding and finally abolish the pseudovirus entry (*68, 121*) (**Fig. 9**). Recently, we have discovered a number of alpha and beta CoVs including HKU4/5 (*122*)(Ge *et al.*, Co-infection of alphacoronaviruses in one bat community, in China. unpublished results). The established approaches will enable us to analyze the receptor usage of these CoV S proteins and understand the host range and potential interspecies transmission ability of these novel CoVs and finally predict the potential spillover probability of these viruses to humans or other hosts at a molecular level.

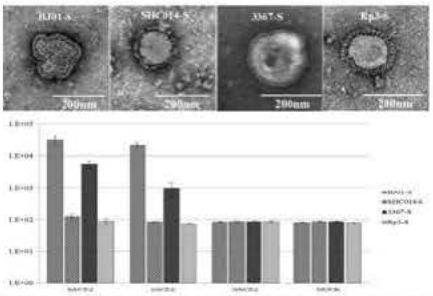


Figure 9: Top panel: HIV pseudovirus carrying spike proteins from human SARS-CoV (BJ01-S) and bat SARS-like CoV (SHC014-S, 3367-S and Rp-S). Bottom panel: Infectivity assay with the above pseudoviruses in HeLa cell lines expressing ACE2 from human, civet and bat.

In vitro cell lines & Humanized mouse model: We have developed primary cell lines and transformed cell lines from 9 bat species using kidney, spleen, heart, brain and intestine. We have used these for virus isolation, infection assays and receptor molecule gene cloning. We also have a large number of cell lines from humans and animals that we will use for virus infectivity assays. We have obtained a letter of support from Dr Ralph Baric, who is keen to collaborate with us initially to infect his humanized mouse model with our bat SL-CoV that uses ACE2, and subsequently to use other CoVs we identify (see **Dr Ralph Baric, Letter of Support**).

C3b) Receptor-mutant pseudovirus binding assays: We will amplify ACE2, DPP4 or other receptor genes of human and bats and clone them into eukaryotic expression vector pcDNA3.1 to construct cells expressing these molecules. We will amplify full length spike genes (S) of bat-CoVs detected from different bat species. The full length S gene, particularly RBDs, will be codon optimized, then cloned into eukaryotic expression vector pcDNA3.1 (*68, 123*). For packaging pseudovirus, S-expressing plasmids (or empty vector control) and pHIV-Luc (pNL4.3.Luc.R$^-$ E$^-$ -Luc) bone plasmid will be co-transfected into 4 x 10^6 293T cells using calcium phosphate transfection system (Promega), after 8 hours, replacing the medium with fresh medium, and supernatants will be harvested at 48 hours post transfection and separated from cell debris by centrifugation at 3,000g, then by passing through a 0.45μm filter (Millipore). The filtered supernatants will be stored at -80°C in aliquots until the use. We will use prepared pseudoviruses bearing different S proteins to infect human and bat ACE2 or DPP4 receptor expressing cells (in Hela cell model), 24 hours post infection, receptor usage by different S proteins will be determined by measuring luciferase activities. We will also induce site mutations in S proteins using site-directed mutation method, then do receptor-mutant pseudovirus binding assays. Pseudovirus infectivity on different human cell lines (A549, 293T, Caco, Huh7, and etc), primary and immortalized bat cell lines (listed below) and other mammalian cell line (mouse, pig, hamster, monkey, and ect) will be also determined by luciferase assay. The results will provide information whether bat-CoVs could use known bat and human ACE2, DPP4 or other known CoV receptors to enter cells, and allow us to determine critical receptor binding sites, viral host range, and to better predict the capacity of our CoVs to infect people.

C3c) *In vitro* studies: We will isolate bat-CoVs using Vero E6 cell (susceptible SARS-CoV and MERS-CoV) and primary or transformed bat cell lines that we have developed from *Myotis davidii, Rhinolophus sinicus, Myotis chinensis, Rousettus leschenaultia* and other bats of China (*124, 125*). CoV PCR-positive bat samples (in 200 μl buffer) will be 3,000-12,000 rpm gradient centrifuged, and supernatant will be diluted at 1:10 in DMEM medium, then added to cells, incubated at 37°C for 1 h, the inoculum removed and replaced by fresh DMEM medium with 2% fetal calf serum, and cells checked daily for cytopathic effect (CPE). Double dose triple antibiotics (penicillin 200 IU/ml, streptomycin 0.2 mg/ml, amphotericin 0.5 μg/ml-Gibco) will be included in all culture media. Three blind passages will be carried out for each sample and the culture supernatant and cell pellet examined for presence of virus by RT-PCR using primers targeting the RdRp or S gene after each passage (*28, 126*). Live bat-CoVs will be sequenced to confirm viral receptor and by comparing viral infection in ACE2 or DPP4 expression cells and virus infectivity and replication on different human cell lines (A549, 293T, Caco, Huh7, and etc), bat cells and others (mouse, pig, hamster, monkey) using plaque assay, real time-PCR, and Immunological Fluorescence Assay (IFA). These *in vitro* assays will be used to test viral host species range and transmission possibility of bat-CoVs to human and other mammal, as predicted by our GLM and matrix models.