

Coronavirus sampling and surveillance in bats from 1996–2019: a systematic review and meta-analysis

Received: 15 June 2022

Accepted: 24 March 2023

Published online: 25 May 2023

 Check for updates

Lily E. Cohen¹✉, Anna C. Fagre², Binqi Chen³, Colin J. Carlson³
& Daniel J. Becker⁴

The emergence of SARS-CoV-2 highlights a need for evidence-based strategies to monitor bat viruses. We performed a systematic review of coronavirus sampling (testing for RNA positivity) in bats globally. We identified 110 studies published between 2005 and 2020 that collectively reported positivity from 89,752 bat samples. We compiled 2,274 records of infection prevalence at the finest methodological, spatiotemporal and phylogenetic level of detail possible from public records into an open, static database named *datacov*, together with metadata on sampling and diagnostic methods. We found substantial heterogeneity in viral prevalence across studies, reflecting spatiotemporal variation in viral dynamics and methodological differences. Meta-analysis identified sample type and sampling design as the best predictors of prevalence, with virus detection maximized in rectal and faecal samples and by repeat sampling of the same site. Fewer than one in five studies collected and reported longitudinal data, and euthanasia did not improve virus detection. We show that bat sampling before the SARS-CoV-2 pandemic was concentrated in China, with research gaps in South Asia, the Americas and sub-Saharan Africa, and in subfamilies of phyllostomid bats. We propose that surveillance strategies should address these gaps to improve global health security and enable the origins of zoonotic coronaviruses to be identified.

Since the emergence of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) in 2002, coronaviruses (Coronaviridae: Orthocoronavirinae) have been recognized as potential pandemic threats. The group comprises four genera containing an estimated hundreds, or thousands, of viruses¹. The delta- and gammacoronaviruses are primarily bird pathogens, although they also infect some mammals; notably, porcine deltacoronavirus was reported to infect humans in 2021 (ref. 2). The alpha- and betacoronaviruses contain all other known human-infective coronaviruses. Betacoronaviruses include SARS-CoV, Middle East respiratory syndrome-related coronavirus (MERS-CoV)

and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), all of which have caused morbidity and mortality in humans³. While alpha- and betacoronaviruses can infect many different hosts, substantial diversity of coronaviruses occurs in bats, which are probably the ancestral hosts of these coronavirus genera^{4,5}. Owing to this, coronaviruses, along with other clades of zoonotic viruses including filoviruses, lyssaviruses and henipaviruses, continue to be extensively monitored in wild bats⁶.

Research into the natural origins of SARS-CoV-2 and continuing interest in coronavirus ecology and evolution have highlighted

¹Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²Department of Microbiology, Immunology, and Pathology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA. ³Center for Global Health Science and Security, Georgetown University Medical Center, Washington, DC, USA. ⁴Department of Biology, University of Oklahoma, Norman, OK, USA. ✉e-mail: lily.cohen@icahn.mssm.edu

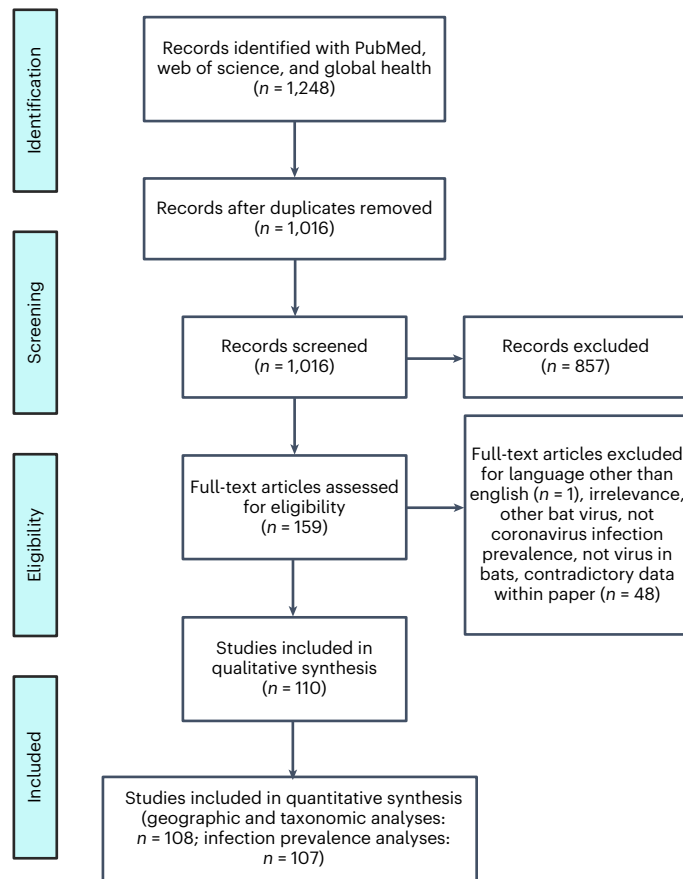


Fig. 1 | PRISMA reporting for systematic review and meta-analysis.

the value of wild bat surveillance. However, field sampling is often carried out opportunistically in response to concerns about spillover, and capacity for systematic sampling is financially or logistically constrained⁷. For example, comparative analyses of bat filovirus and henipavirus positivity have shown that only a small fraction of studies report longitudinal data, limiting inference into temporal dynamics of infection in bats⁶. Single sampling events can bias prevalence estimates in biologically meaningful ways, for example if sampling is more convenient in one season over another, and may lead to non-randomly missing data. Unlike single sampling studies, spatiotemporal designs can identify seasonal and environmental drivers of viral prevalence and shedding intensity, but they are logistically challenging and often have either spatial or temporal replication but not both⁶.

If the ultimate goal is to explain and predict pathogen spillover—a dynamic process that is driven by geographical and temporal variation in infection prevalence and shedding from reservoir hosts^{6,8}, there is a critical need to resolve the relative importance of spatiotemporal, taxonomic and methodological factors (for example, tissues sampled, use of euthanasia, diagnostic method) that may impact virus positivity. Unfortunately, a lack of standardized and aggregated data from disparate studies limits our ability to quantify whether and how these many different factors shape global assessments of coronavirus infection in bats and downstream spillover risk.

To provide baseline data to inform future surveillance efforts, we compiled a standardized global database of infection prevalence estimates using published pre-pandemic coronavirus testing data from wild bat samples and included metadata on bat and viral taxonomy, study methodology, bat demography, bat seasonality and ecological context. We used our database to test several standing hypotheses, including that (1) longitudinal sampling results in higher virus detection rates^{6,9}, (2) seasonality affects virus shedding and detection rates^{1,10}

and (3) viral detection varies in different sample types¹¹. More broadly, we evaluated the global state of coronavirus surveillance in bat hosts before SARS-CoV-2-motivated research efforts.

Results

Dataset description

We first identified global biases in the distribution and intensity of pre-pandemic bat coronavirus surveillance. From publicly available literature published between 2005 and 2020, we recovered 89,752 tests for coronaviruses in bats from 110 studies^{12–121} (Fig. 1 and Supplementary Table 1). Within the pooled-coronavirus genera (alpha- and betacoronavirus) infection prevalence dataset, which comprised data from 107 studies, approximately 95% of studies used PCR targeting the RNA-dependent RNA polymerase (RdRp) gene to detect viruses; other gene targets included subunits of the coronavirus spike protein, the nucleocapsid gene or the envelope protein. Of the 106/107 studies detecting coronaviruses by PCR, approximately 56% used single-round PCR, as opposed to nested PCR or multiple PCR assays in parallel to target different genes in the same RNA sample. More than half of these studies (53.8%) designed their primers using protocols from four studies^{11–124}. Of the pooled-coronavirus genera infection prevalence records, 35% was derived from studies that had euthanized bats. Supplementary Table 2 lists the sample types analysed and the associated percentages of positive and zero-infection prevalence. Faecal samples and rectal swabs were the most common samples used to detect coronavirus RNA. Sex and/or reproductive status of bats was only described in 13 of 110 studies in our full database, limiting downstream analyses of sex biases in coronaviruses infection or possible impacts of reproductive stress on viral susceptibility and shedding⁸.

Spatial bias in coronavirus surveillance

Before the COVID-19 pandemic, we identified studies reporting sampling of wild bats for coronavirus infection in 52 countries on 6 continents. However, the distribution and frequency of viral surveillance was uneven (Fig. 2). Individual countries had 1 to 32 bat coronavirus studies (Fig. 2a), with the number of total samples tested ranging from 4 to 26,051 (Fig. 2b). Whereas sampling occurred in all North American countries, Central and South America had sparse surveillance. Sampling in sub-Saharan Africa and in Central and South Asia has been inconsistent, with most surveillance carried out in China and in some other regions of Southeast Asia. A generalized linear model (GLM) of binary sampling effort ($\chi^2 = 13.02$, $P = 0.01$, $R^2 = 0.04$) confirmed that countries in Asia and Europe were marginally more likely to have data on bat coronaviruses than those in the Americas and in Oceania (Supplementary Table 3). We found substantial geographic biases for the relative intensity of sampling, specifically the number of studies ($\chi^2 = 17.92$, $P = 0.001$, $R^2 = 0.06$) and the number of tested samples ($\chi^2 = 20671$, $P < 0.001$, $R^2 = 0.12$). Post-hoc comparisons using GLMs revealed that there were more bat coronavirus studies per country in Asia than in Africa or Europe (Supplementary Table 4). Similarly, the greatest contrast in total number of tested bat samples was between Asia and Europe (risk ratio = 4.64), and between the Americas and Europe (risk ratio = 2.11; Supplementary Table 5).

Taxonomic biases in surveillance

More than 1 in 4 bat species (343 species of the 1,287 included in the most recent bat phylogeny¹²⁵) were sampled in pre-COVID-19 pandemic coronavirus surveillance. Bats have been sampled evenly across the phylogeny (Fig. 3a). Of the 19 bat families included in this phylogeny, 15 had at least 1 member species sampled in our dataset. Unsampled bat families included the Furiferidae, Natalidae, Myzopodidae and Thyropteridae. Indeed, we only identified intermediate phylogenetic signal in binary sampling effort ($D = 0.86$) that departed from both phylogenetic randomness ($P < 0.001$) and Brownian motion models of evolution ($P < 0.001$). Similarly, phylogenetic factorization¹²⁶,

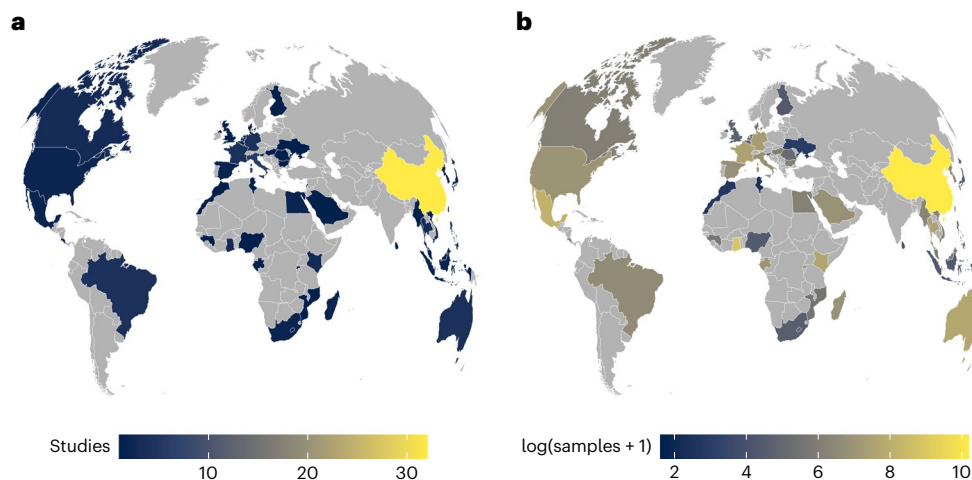


Fig. 2 | Geographic distribution of bat coronavirus sampling effort. Geographic distribution is defined by the number of studies per country (a) and the number of samples tested per country (b). Sampled countries varied in having 1 to 32 bat coronavirus studies (a), with the number of total samples tested ranging from 4 to 26,051 (b). A disproportionate number of bat coronavirus studies and testable

samples were conducted and assayed in China, probably reflecting interest in the subgenus *Sarbecovirus* and the risk of future SARS-like virus emergence. Many areas were severely understudied, particularly relative to ecological and evolutionary risk factors for emergence¹³¹. In particular, sampling in Central and South America, sub-Saharan Africa and Central and South Asia was notably limited.

a graph-partitioning algorithm based on the bat phylogeny, did not identify any bat clades that differed considerably in their fraction of sampled species. In contrast, we observed stronger taxonomic biases in sampling intensity. The number of studies per sampled species ranged from 1 to 23 (*Miniopterus schreibersii* and *Rhinolophus ferrumequinum*), whereas the number of total samples tested ranged from 1 to 16,499 (*Rhinolophus sinicus*). The number of studies per sampled species showed low phylogenetic signal ($\lambda = 0.02$) that departed from Brownian motion models of evolution ($P < 0.001$) but not phylogenetic randomness ($P = 0.56$). Phylogenetic factorization did, however, more flexibly identify 3 bat clades with greater mean numbers of studies than the paraphyletic remainder (Fig. 3b): a subclade of the genus *Myotis* (including both European and Asian species), a subclade of the tribe Pipistrellini (including the genera *Pipistrellus* and *Nyctalus*) and a subclade of the family Rhinolophidae (Supplementary Table 8); notably, all highly sampled clades consisted exclusively of Old World bat species.

For the total number of tested samples per species, we instead observed more intermediate phylogenetic signal ($\lambda = 0.27$) that departed from both Brownian motion models of evolution ($P < 0.001$) and phylogenetic randomness ($P < 0.001$). Accordingly, phylogenetic factorization identified a total of 39 clades with differential intensities of sampling effort, 15 of which had relatively more tested samples and 24 had relatively fewer tested samples (Fig. 3c). The top clades with comparatively fewer total samples included a large portion of the suborder Yangochiroptera; the above-mentioned subclade of the tribe Pipistrellini; members of the phyllostomid subfamilies Stenodermatinae, Glossophaginae and Phyllostominae; and the sister families Rhinolophidae and Hipposideridae; these results suggest a greater number of publications on some of these bat taxa but fewer tested samples. However, smaller subclades of the Hipposideridae and Rhinolophidae families were some of the most heavily sampled, suggesting key biases in sampling effort within these taxa that have been the subject of much coronavirus research (Supplementary Table 9). Finally, members of several genera within the Pteropodinae subfamily were undersampled (that is, *Pteropus*, *Eidolon* and *Acerodon*), while others displayed greater sampling effort (that is, the subfamily Rousettinae).

Heterogeneity in coronavirus infection prevalence

Using a phylogenetic meta-analysis model that accounted for sampling variance, bat phylogeny, additional species effects, and within- and

between-study variation^{127,128}, we observed high heterogeneity among coronavirus infection prevalence estimates ($I^2 = 84.2\%$, $Q_{1,854} = 8,620.69$, $P < 0.0001$). This heterogeneity was mainly due to within-study (43.65%) and between-study effects (31.53%), with smaller contributions from bat phylogeny (9.02%) and additional species effects (0.001%). When repeating this intercept-only model for alphacoronavirus- and betacoronavirus-specific datasets, prevalence showed similar patterns of heterogeneity (alphacoronavirus: $I^2 = 79.10\%$, $Q_{1,553} = 4,973.72$, $P < 0.0001$; betacoronavirus: $I^2 = 74.10\%$, $Q_{1,428} = 3,871.49$, $P < 0.0001$), mainly due to within-study (alphacoronavirus: 35.50%; betacoronavirus: 30.21%) and between-study effects (alphacoronavirus: 36.94%; betacoronavirus: 29.88%) and secondarily by phylogeny (alphacoronavirus: 6.66%; betacoronavirus: 14.02%) or other species-level effects (alphacoronavirus: 0.001%; betacoronavirus: 0%).

Methodological and biological predictors of prevalence

When considering the suite of methodological and biological predictors in our phylogenetic meta-analysis models, fixed effects explained approximately 20% of the variance in infection prevalence (pooled-coronavirus genera $R^2 = 0.19$; alphacoronavirus-only $R^2 = 0.21$; betacoronavirus-only $R^2 = 0.19$). Sample type, sampling method and study format were the strongest predictors of coronavirus prevalence (Table 1). Within our pooled-coronavirus dataset, lung or respiratory samples (untransformed $\beta = -0.09$; 95% confidence interval (CI): -0.14 to -0.04 , $P = 0.001$), oropharyngeal samples (untransformed $\beta = -0.08$; 95% CI: -0.14 to -0.03 , $P = 0.004$), pooled swabs/samples (untransformed $\beta = -0.07$; 95% CI: -0.12 to -0.03 , $P = 0.003$) and pooled tissue (untransformed $\beta = -0.13$; 95% CI: -0.22 to -0.04 , $P = 0.006$) all had lower prevalence than faecal/rectal or intestinal samples, with weaker associations observed for only alphacoronaviruses and only betacoronaviruses (Fig. 4). Across all three datasets, repeat sampling was associated with a 0.70–1.6% increase in coronavirus prevalence (pooled coronavirus: untransformed $\beta = 0.15$; 95% CI: 0.05–0.25, $P = 0.003$; alphacoronavirus: untransformed $\beta = 0.14$; 95% CI: 0.03–0.26, $P = 0.03$; betacoronavirus: untransformed $\beta = 0.13$; 95% CI: 0.03–0.23, $P = 0.009$) as compared to one-time (single) sampling (Fig. 4). Similarly, longitudinal study design predicted a small increase (-0.23 – 0.33%) in positive viral detection in the pooled coronavirus (untransformed $\beta = 0.06$; 95% CI: 0.01–0.11, $P = 0.01$) and alphacoronavirus-only (untransformed $\beta = 0.07$; 95% CI: 0.02–0.12, $P = 0.008$) datasets, as

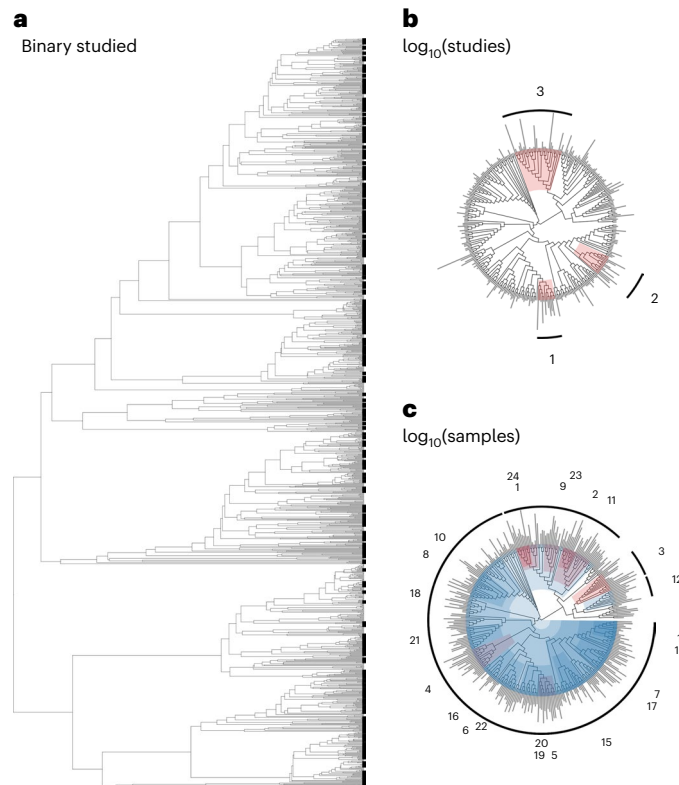


Fig. 3 | Evolutionary distribution of bat coronavirus sampling effort.

Sampling effort is defined as whether a bat species has been sampled (a), the number of studies (b) and the number of samples tested (c). Clades identified by phylogenetic factorization with greater or lesser sampling effort compared with a paraphyletic remainder are shown in red and blue, respectively, alongside clade numbers per analysis. Phylogenetic factorization did not identify any taxonomic patterns in binary sampling effort across the bat phylogeny (a), but did identify a number of bat clades within sampled bat species that have been particularly well-sampled for coronaviruses, both in terms of number of studies (b; Supplementary Table 8) and number of samples (c; Supplementary Table 9, only the first 24 phylogenetic factors are displayed). For analyses of total studies and tested samples, segment length corresponds to the relative degree of sampling effort.

opposed to cross-sectional sampling. Other model variables including sampling season, bat family, PCR type and gene target showed weak or no association with coronavirus positivity across all datasets. Notably, use of euthanasia was not associated with greater ability to detect coronavirus RNA (pooled coronavirus: untransformed $\beta = -0.01$; 95% CI: -0.07 to 0.05 , $P = 0.86$; alphacoronavirus: untransformed $\beta = -0.01$; 95% CI: -0.08 to 0.05 , $P = 0.73$; betacoronavirus: untransformed $\beta = 0.004$; 95% CI: -0.05 to 0.06 , $P = 0.89$).

Discussion

Since the onset of the COVID-19 pandemic, increased attention has been paid to bats as potential reservoir hosts of coronaviruses, presumably including viruses with zoonotic potential^{129–131}. While other studies have reported data on the geographical and taxonomic distribution of reported bat hosts^{131,132}, we generated a standardized, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-compliant open and static database of coronavirus surveillance in bats, which provides disaggregated data (including negative results). In doing so, our study takes an important step towards building an open database of wildlife disease surveillance with relevance to pandemic prediction and preparedness¹³³.

Our database is a snapshot of bat coronavirus research before the COVID-19 pandemic and includes 110 studies, 2,274 records of infection

prevalence and a total of 89,752 bat samples. Our geographic and taxonomic analyses reveal that most bat sampling has taken place in China, with gaps in surveillance in South Asia, the Americas, sub-Saharan and East Africa. Additionally, very few such studies were carried out in the United States and Canada.

Progress towards addressing gaps in surveillance has been made since the onset of the pandemic; for example, recent bat surveillance in Latin America and Madagascar has been reported^{131,134–138}. Although phylogenetic coverage of bat species is a strength of the dataset, we identified taxonomic patterns in the intensity of sampling efforts. Our analyses confirm previous findings, such as a greater number of surveillance studies in the Rhinolophidae and a disproportionate number of studies in China¹³⁹. However, we also characterized finer-scale variation in sampling effort relevant to prioritizing future surveillance. For example, although many studies have been conducted on rhinolophid bats, the Rhinolophidae and Hipposideridae families also had low sample sizes for coronavirus diagnostics, suggesting low power to detect viruses on a per-species basis. Further, subclades of the Hipposideridae and Rhinolophidae as well as the Rousettinae subfamily of pteropid bats were some of the most heavily sampled taxa versus considerable undersampling within subfamilies of phyllostomid bats in particular. Strengthening surveillance efforts in undersampled regions and specific bat taxa is important; for example, greater sampling of rhinolophid and hipposiderid species that fall outside identified well-sampled subclades is likely to uncover novel coronaviruses (Supplementary Table 9). Sampling the understudied Neotropical subfamilies Stenodermatinae and Glossophaginae might also have potential to uncover novel betacoronaviruses, as predicted by recent models¹³¹.

After controlling for bat phylogeny, sampling variance, and both study- and observation-level heterogeneity, we found that sample type, repeat sampling and longitudinal study design were the most important predictors of coronavirus prevalence. We did not find consistent support for seasonality in coronavirus prevalence¹⁰, whereas we did find support for longitudinal sampling enabling coronavirus detection^{6,9} and for successful coronavirus detection varying by sample type¹¹. Specifically, lung or respiratory samples, urinary samples, oropharyngeal samples, pooled swabs and pooled tissue were associated with lower prevalence across all studies, with weaker effects generally observed in alphacoronavirus- and betacoronavirus-only datasets. In contrast, repeat sampling and longitudinal study designs, as well as intestinal and faecal and rectal samples, were consistently associated with viral detection. This might reflect gastrointestinal tropism of coronaviruses in bats¹¹.

To optimize coronavirus detection, combining the above set of sampling approaches¹⁴⁰, particularly using faecal samples or rectal swabs, should enhance detection of coronaviruses from wild bats. Moreover, longitudinal study designs will be crucial to pinpoint how coronaviruses are transmitted among wild bat hosts^{140,141} and identify the intrinsic and extrinsic drivers of virus shedding^{142,142}. Euthanasia did not affect the likelihood of virus detection, which means that coronavirus surveillance can be accomplished with minimally invasive (for example, rectal swab) and readily accessible samples (for example, museum-derived, such as whole specimens or individual organs) rather than requiring terminal sampling¹⁴³. Avoiding euthanasia reduces negative impacts of virus surveillance studies on bat population dynamics and enables longitudinal, mark-recapture designs. However, we note that selective terminal sampling can still provide other important benefits for virus surveillance, including the ability to post hoc confirm the species identity of voucher specimens, study tissue tropism and receptor usage of coronaviruses and provide lasting evidence of specific bat–virus associations in scientific collections^{143,144}.

Our systematic review identified multiple challenges in synthesizing viral surveillance data from wildlife studies. Although study-level effects can be accounted for in part with random effects

Table 1 | Meta-analysis of coronavirus prevalence across studies

	Alphacoronavirus or betacoronavirus			Alphacoronavirus only			Betacoronavirus only		
	Q	d.f.	P	Q	d.f.	P	Q	d.f.	P
Sampling method	16.066	2	0.0003	9.347	2	0.0093	17.818	2	0.0001
Study format	6.302	1	0.0121	7.058	1	0.0079	2.252	1	0.1334
PCR type	1.368	1	0.2422	0.4157	1	0.5191	2.993	1	0.0837
Sample type	38.005	8	<0.0001	17.612	8	0.0243	30.033	8	0.0002
Euthanasia use	0.0332	1	0.8555	0.1166	1	0.7328	0.0186	1	0.8915
Bat family	11.5996	12	0.4783	10.8095	12	0.5453	14.9070	12	0.2466
Sampling season	8.3251	4	0.0804	9.9849	4	0.0407	6.9559	4	0.1382
Gene target	2.2751	2	0.3206	0.5962	2	0.7422	2.9593	2	0.2277

Analysis of variance (ANOVA) table from the phylogenetic meta-analysis model fit using REML to all data and each data subset (alphacoronavirus only or betacoronavirus only). For each variable, we provide Cochran's Q, the associated degrees of freedom and the two-sided P value.

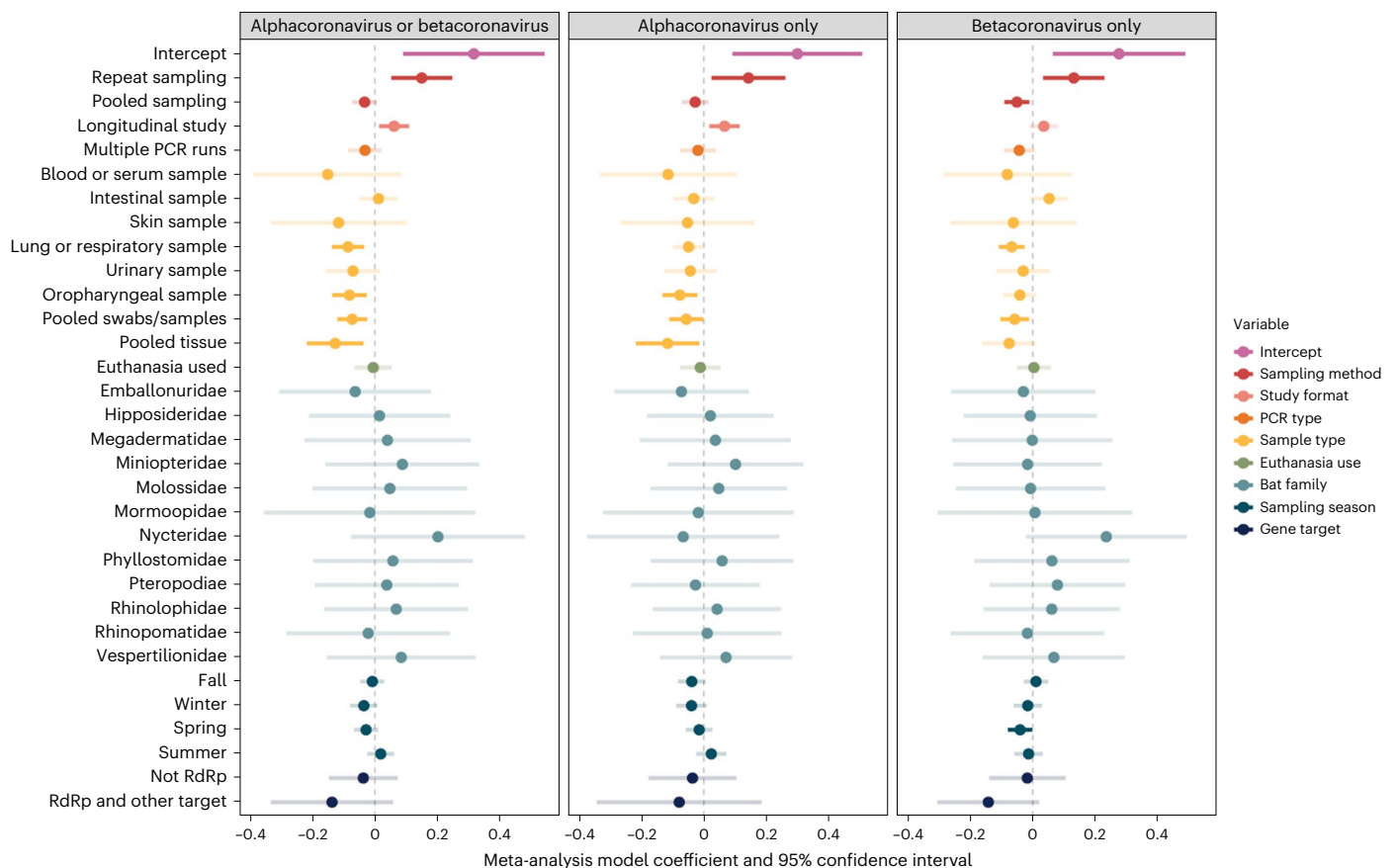


Fig. 4 | Methodological and biological predictors of coronavirus prevalence in wild bats. Phylogenetic meta-analysis model coefficients and 95% confidence intervals, estimated using REML for each of our three datasets. Colours indicate the nine variables included in each model (binary covariates for sampling season). Estimate confidence intervals are shaded by whether they cross zero (the vertical dashed line), with increased transparency denoting non-significant effects. The intercept contains the following reference levels: single sampling

(sampling method); cross-sectional study (study format); single PCR (PCR type); faecal, rectal or anal sample (sample type); euthanasia not used (euthanasia use); Craseonycteridae (bat family); not fall, not winter, not spring and not summer (sampling season); and RNA-dependent RNA polymerase (RdRp) only (gene target). Sample sizes are 1,854 prevalence estimates for all coronaviruses, 1,553 prevalence estimates for only alphacoronaviruses and 1,428 prevalence estimates for only betacoronaviruses.

in meta-analysis, we note that at least some of our non-significant results could be due to variability in study format, sampling design and reporting. To reduce this limitation in the future, we encourage researchers to report data at the finest resolution possible (for example, fully stratified by location, timepoint, bat species, virus species or strain, and sample type). Developing and adopting data standards for reporting these types of data—and real-time channels to aggregate them with standardized metadata—could substantially improve our

ability to address research questions regarding transmission dynamics, bat immunology, viral evolution and spillover risk.

Methods

Systematic review

To identify studies quantifying the proportion of wild bats positive for alpha- or betacoronaviruses using PCR or serological methods, we followed the PRISMA protocol (Fig. 1)⁴⁵. We systematically searched

Web of Science, PubMed and Global Health (a database comprising publications from the Public Health and Tropical Medicine database and CAB Abstracts). PubMed searches used the following string: (bat* OR Chiroptera*) AND (coronavirus* OR CoV*). Web of Science and Global Health (comprising CAB Abstracts and Public Health and Tropical Medicine database) searches used the following string: (bat* OR Chiroptera*) AND (coronavirus* OR CoV*) AND (wild*). Searches were performed on 24 September 2020 and included studies published in or after 1984.

We screened a total of 1,016 abstracts for studies that included sampling of wild bats for coronaviruses. Publications were excluded if they did not assess coronavirus prevalence in bats or were published in languages other than English (this led to the exclusion of only a single dissertation, written in Portuguese). In total, we identified a total of 159 candidate articles that we screened for these data. Of these, 110 studies tested bats for coronaviruses, reported reusable data and were included in our final, publicly available dataset. Geographic and taxonomic analyses, which did not rely on population-level prevalence estimates, were performed on a 108-study subset of the public dataset which excludes records with genus- or family-level versus species-level bat data and includes data that could not be used to calculate prevalence (for example, number of samples corresponds to geographic region rather than bat species). Infection prevalence analyses were performed on a 107-study subset of the public dataset. Each of these two datasets were then divided into three more: pooled-coronavirus genera (alphacoronaviruses and betacoronaviruses), alphacoronavirus genus-only and betacoronavirus genus-only (Supplementary Table 1). The datasets used for geographic and taxonomic analyses, which included data that could not be used to calculate prevalence (for example, number of samples corresponds to geographic region rather than bat species) had 37 (pooled-coronavirus genera), 21 (alphacoronavirus genus-only) and 9 (betacoronavirus genus-only) more rows than the corresponding infection prevalence datasets.

Our aim was to provide a comprehensive record of bat coronavirus surveillance up to the beginning of the COVID-19 pandemic, and our sample necessarily omits more recent publications that have reanalysed samples, motivated by investigations into the evolutionary origins of SARS-CoV-2 and other L2 lineage sarbecoviruses. It also omits the final dataset compiled by the USAID PREDICT dataset and released at the end of 2020. Standardized PREDICT format is a substantively different kind of data compared with all other studies we analysed; these data have been extensively analysed elsewhere¹. Additionally, only 16 of the 110 studies in our database reported financial support from the PREDICT programme, suggesting that a substantial breadth of data collection exists in the literature beyond any one collaborative project.

Data collection

Our initial dataset consists of a total of 110 studies and 2,274 records. Each record provides an infection prevalence estimate at the finest spatiotemporal, methodological and phylogenetic scale reported. More precisely, each unique record includes a distinct combination of coronavirus genus; bat genus, family and/or species; sample type; detection method (that is, PCR or serology); gene/protein target; date/sampling season and geographic location (sampling country, state, and specific site and/or geographic coordinates, if available). Sampling season was determined by month of sampling according to National Oceanic and Atmospheric Administration meteorological definitions; in the Northern Hemisphere, sample seasons equated to fall (September–November), winter (December–February), spring (March–May) and summer (June–August), while in the Southern Hemisphere these groupings were inverted (for example, December–February was classified as summer)¹⁴⁶. Detection estimates derived at finer phylogenetic scales (for example, virus strain) were aggregated

to genus. Prevalence estimates that combined two or more sample subtypes (for example, lung and small intestine) and that could not be further separated were recorded as pooled. As observed previously for bat filoviruses and henipaviruses, some studies pooled coronavirus detection estimates for more than one bat species⁶. Rows with these pooled prevalence estimates were excluded from subsequent statistical analyses. Study formats were classified as longitudinal and cross-sectional: prevalence estimates derived from repeated sampling at one location were marked as longitudinal, while those derived from one location on a specific date were listed as cross-sectional. Thus, most studies (92.7%) yielded more than one detection estimate record: for example, a longitudinal study that provides individual coronavirus detection estimates from two types of samples in a given bat species on six separate dates spanning several years would result in at least 12 records in the dataset.

In addition to these spatial and temporal components, we recorded data on detection methodology (for example, single or nested/multiple PCR for RNA detection or lateral flow immunoassay for antigen detection), additional virus taxonomy (for example, subgenus, strain), PCR primers (and their gene targets) and whether the authors included information on the sex of the sampled bats or the use of euthanasia. We note that infection prevalence estimates are based on the number of samples tested for coronaviruses rather than the number of individual bats, as studies often tested multiple samples per individual specimen (for example, saliva, faeces, blood, tissue).

Geographic and taxonomic analyses of sampling effort

With these data, we assessed geographic and taxonomic patterns in bat sampling effort. For the former, we fitted a GLM, with whether a country had been sampled for bat coronaviruses as a binomial response and region as the predictor in R. For sampled countries ($n = 52$), we fitted equivalent GLMs that modelled the number of unique studies and the total samples per country as a Poisson-distributed response. For each GLM, we assessed fit using McFadden's R^2 and the 'performance' package¹⁴⁷. We also adjusted for the inflated false-discovery rate in post-hoc comparisons using 'emmeans'¹⁴⁸. Here and below, all statistical tests are two-tailed.

For taxonomic patterns, we derived equivalent response variables across bat species, using a recent phylogeny as a taxonomic backbone¹⁵. We note that despite being a recent synthesis, the number of bat species included this phylogeny ($n = 1,287$) remains an underestimate of known bat diversity (over 1,460 species); as such, corresponding taxonomic analyses necessarily exclude approximately 12% of extant bat species. Additionally, only four species in our dataset were absent from this phylogeny (*Pipistrellus taiwanesis*, *Pipistrellus montanus*, *Myotis rufoniger*, *Rhinolophus cornutus*) and were excluded from phylogenetic analyses. We also reclassified species in the genus *Miniopterus* from the Vespertilionidae to be the sole members of the family Miniopteridae¹⁴⁹. For all bat species in our phylogeny, we derived a binary response for whether a species had been sampled for coronaviruses. For those sampled species ($n = 343$), we derived the number of unique studies and the total samples. Using the 'caper' package¹⁵⁰, we first estimated phylogenetic signal in sampling effort (that is, the propensity for related bat species to be sampled in a similar intensity). For binary sampling effort, we calculated D , where a value of 1 indicates a phylogenetically random trait distribution and 0 indicates phylogenetic clustering under a Brownian motion model of evolution¹⁵¹. For sampled species, we estimated Pagel's λ for the \log_{10} -transformed number of studies and samples¹⁵². Next, we applied a graph-partitioning algorithm, phylogenetic factorization, to more flexibly identify any bat clades across taxonomic levels that differ in sampling effort. With a standardized taxonomy from our bat phylogeny¹⁵, we used the 'phylofactor' package to partition binary sampling effort, number of studies and number of samples in a series of iterative GLMs for each edge in the tree^{16,153}. As in our geographic analyses, we

modelled these variables with binomial and Poisson distributions. We then determined the number of significant clades using Holm's sequentially rejective test with a 5% family-wise error rate¹⁵⁴.

Phylogenetic meta-analysis of infection prevalence

We first used the 'metafor' package to calculate Freeman–Tukey double arcsine-transformed proportions of coronavirus infection-positive bats and their corresponding sampling variances^{10,18,20}. We then built two hierarchical meta-analysis models for three infection prevalence datasets: the global dataset, an alphacoronavirus-specific dataset and a betacoronavirus-specific dataset (see Supplementary Table 1 for the sample size per model). Each model was fitted using restricted maximum likelihood (REML) and included bat species and phylogeny (using the previous bat tree) as random effects alongside an observation-level random effect nested within a study-level effect¹⁷. The first model (that is, model 1) for each dataset only included an intercept and was used to estimate I^2 , which quantifies the contribution of true heterogeneity (rather than noise) to variance in infection prevalence¹⁵⁵. We report both the overall I^2 per dataset as well as the proportional I^2 for each random effect, and we used Cochran's Q to test whether such heterogeneity was greater than that expected by sampling error alone. The second model (that is, model 2) for each dataset included the following moderators: sampling method (repeat vs single), study format (longitudinal vs cross-sectional sampling), PCR type (nested/multiple vs single), sample analysed, whether terminal sampling was performed, bat family, sampling season and gene target. We calculated variance inflation factors for all moderators in the linear model; the moderators displayed no substantial collinearity¹⁵⁶. To facilitate estimating model coefficients, we removed levels for any moderators with $n < 3$. For each iteration of model 2, we assessed moderator significance using the Q test (that is, a Wald-like test of all coefficients per moderator) and estimated a pseudo- R^2 as the proportional reduction in the summed variance components compared against those from an intercept-only model¹⁵⁷.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The primary dataset is available on GitHub (www.github.com/viralemergence/datacov; <https://doi.org/10.5281/zenodo.6644163>) and comprises data extracted from papers obtained during a systematic search of PubMed (<https://pubmed.ncbi.nlm.nih.gov>), Web of Science (<https://www.webofscience.com>) and Global Health (<https://www.cabdirect.org/globalhealth>). Source data are provided with this paper.

Code availability

Data were analysed in R Studio (v2021.9.2 'Ghost Orchid'). The unprocessed data and scripts to generate the primary dataset (and all other derived datasets) and to replicate all analyses and visualizations are available at www.github.com/viralemergence/batgap; <https://doi.org/10.5281/zenodo.6644081>.

References

- Anthony, S. J. et al. Global patterns in coronavirus diversity. *Virus Evol.* **3**, vex012 (2017).
- Lednický, J. A. et al. Independent infections of porcine deltacoronavirus among Haitian children. *Nature* **600**, 133–137 (2021).
- Zhu, Z. et al. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **21**, 1–14 (2020).
- Woo, P. C. Y. et al. Molecular diversity of coronaviruses in bats. *Virology* **351**, 180–187 (2006).
- Woo, P. C. Y. et al. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *J. Virol.* **86**, 3995–4008 (2012).
- Becker, D. J., Crowley, D. E., Washburne, A. D. & Plowright, R. K. Temporal and spatial limitations in global surveillance for bat filoviruses and henipaviruses. *Biol. Lett.* **15**, 20190423 (2019).
- Nusser, S. M., Clark, W. R., Otis, D. L. & Huang, L. Sampling considerations for disease surveillance in wildlife populations. *Wildfire* **72**, 52–60 (2008).
- Plowright, R. K. et al. Ecological dynamics of emerging bat virus spillover. *Proc. Biol. Sci.* **282**, 20142124 (2015).
- Giles, J. R. et al. Optimizing noninvasive sampling of a zoonotic bat virus. *Ecol. Evol.* **11**, 12307–12321 (2021).
- Seltmann, A. et al. Seasonal fluctuations of astrovirus, but not coronavirus shedding in bats inhabiting human-modified tropical forests. *Ecohealth* **14**, 272–284 (2017).
- Watanabe, S. et al. Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg. Infect. Dis.* **16**, 1217–1223 (2010).
- Afelt, A. et al. Distribution of bat-borne viruses and environment patterns. *Infect. Genet. Evol.* **58**, 181–191 (2018).
- Ali, M. et al. Cross-sectional surveillance of Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels and other mammals in Egypt, August 2015 to January 2016. *Euro Surveill.* **22**, 30487 (2017).
- Anindita, P. D. et al. Detection of coronavirus genomes in Moluccan naked-backed fruit bats in Indonesia. *Arch. Virol.* **160**, 1113–1118 (2015).
- Annan, A. et al. Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. *Emerg. Infect. Dis.* **19**, 456–459 (2013).
- Anthony, S. J. et al. Coronaviruses in bats from Mexico. *J. Gen. Virol.* **94**, 1028–1038 (2013).
- Ar Gouilh, M. et al. SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world. *Virology* **517**, 88–97 (2018).
- Asano, K. M. et al. Alphacoronavirus in urban Molossidae and Phyllostomidae bats. *Braz. Virol. J.* **13**, 110 (2016).
- August, T. A., Mathews, F. & Nunn, M. A. Alphacoronavirus detected in bats in the United Kingdom. *Vector Borne Zoonotic Dis.* **12**, 530–533 (2012).
- Balboni, A., Palladini, A., Bogliani, G. & Battilani, M. Detection of a virus related to betacoronaviruses in Italian greater horseshoe bats. *Epidemiol. Infect.* **139**, 216–219 (2011).
- Balboni, A., Gallina, L., Palladini, A., Prosperi, S. & Battilani, M. A real-time PCR assay for bat SARS-like coronavirus detection and its application to Italian greater horseshoe bat faecal sample surveys. *ScientificWorldJournal* **2012**, 989514 (2012).
- Berto, A. et al. Detection of potentially novel paramyxovirus and coronavirus viral RNA in bats and rats in the Mekong Delta region of southern Viet Nam. *Zoonoses Public Health* **65**, 30–42 (2018).
- Bittar, C. et al. Alphacoronavirus detection in lungs, liver, and intestines of bats from Brazil. *Microb. Ecol.* **79**, 203–212 (2020).
- Brandão, P. E. et al. A coronavirus detected in the vampire bat *Desmodus rotundus*. *Braz. J. Infect. Dis.* **12**, 466–468 (2008).
- Carrington, C. V. F. et al. Detection and phylogenetic analysis of group 1 coronaviruses in South American bats. *Emerg. Infect. Dis.* **14**, 1890–1893 (2008).
- Chen, Y.-N., Su, B.-G., Chen, H.-C., Chou, C.-H. & Cheng, H.-C. Detection of specific antibodies to the nucleocapsid protein fragments of severe acute respiratory syndrome-coronavirus and Scotophilus bat coronavirus-512 in three insectivorous bat species. *Taiwan. Vet. J.* **44**, 179–188 (2018).

27. Chen, Y.-N. et al. Detection of the severe acute respiratory syndrome-related coronavirus and Alphacoronavirus in the bat population of Taiwan. *Zoonoses Public Health* **63**, 608–615 (2016).
28. Chu, D. K. W. et al. Coronaviruses in bent-winged bats (*Miniopterus* spp.). *J. Gen. Virol.* **87**, 2461–2466 (2006).
29. Corman, V. M. et al. Evidence for an ancestral association of human coronavirus 229E with bats. *J. Virol.* **89**, 11858–11870 (2015).
30. Davy, C. M. et al. White-nose syndrome is associated with increased replication of a naturally persisting coronavirus in bats. *Sci. Rep.* **8**, 15508 (2018).
31. Dominguez, S. R., O’Shea, T. J., Oko, L. M. & Holmes, K. V. Detection of group 1 coronaviruses in bats in North America. *Emerg. Infect. Dis.* **13**, 1295–1300 (2007).
32. Drexler, J. F. et al. Amplification of emerging viruses in a bat colony. *Emerg. Infect. Dis.* **17**, 449–456 (2011).
33. Drexler, J. F. et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J. Virol.* **84**, 11336–11349 (2010).
34. Du, J. et al. Genetic diversity of coronaviruses in *Miniopterus fuliginosus* bats. *Sci. China Life Sci.* **59**, 604–614 (2016).
35. Falcón, A. et al. Detection of alpha and betacoronaviruses in multiple Iberian bat species. *Arch. Virol.* **156**, 1883–1890 (2011).
36. Fischer, K. et al. Insectivorous bats carry host specific astroviruses and coronaviruses across different regions in Germany. *Infect. Genet. Evol.* **37**, 108–116 (2016).
37. Ge, X.-Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol. Sin.* **31**, 31–40 (2016).
38. Ge, X.-Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
39. Geldenhuys, M., Weyer, J., Nel, L. H. & Markotter, W. Coronaviruses in South African bats. *Vector Borne Zoonotic Dis.* **13**, 516–519 (2013).
40. Gloza-Rausch, F. et al. Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. *Emerg. Infect. Dis.* **14**, 626–631 (2008).
41. Góes, L. G. B. et al. Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil. *Infect. Genet. Evol.* **44**, 510–513 (2016).
42. Goffard, A. et al. Alphacoronaviruses detected in French bats are phylogeographically linked to coronaviruses of European bats. *Viruses* **7**, 6279–6290 (2015).
43. Gouilh, M. A. et al. SARS-coronavirus ancestor’s foot-prints in South-East Asian bat colonies and the refuge theory. *Infect. Genet. Evol.* **11**, 1690–1702 (2011).
44. van Gucht, S. et al. No evidence of coronavirus infection by reverse transcriptase-PCR in bats in Belgium. *J. Wildl. Dis.* **50**, 969–971 (2014).
45. Hall, R. J. et al. New alphacoronavirus in *Mystacina tuberculata* bats, New Zealand. *Emerg. Infect. Dis.* **20**, 697–700 (2014).
46. Han, H.-J. et al. Novel coronaviruses, astroviruses, adenoviruses and circoviruses in insectivorous bats from northern China. *Zoonoses Public Health* **64**, 636–646 (2017).
47. He, B. et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J. Virol.* **88**, 7070–7082 (2014).
48. Holz, P. H. et al. Virus survey in populations of two subspecies of bent-winged bats (*Miniopterus orianae bassanii* and *oceanensis*) in south-eastern Australia reveals a high prevalence of diverse herpesviruses. *PLoS ONE* **13**, e0197625 (2018).
49. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, e1006698 (2017).
50. Hu, D. et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.* **7**, 154 (2018).
51. Hu, D. et al. Virome analysis for identification of novel mammalian viruses in bats from Southeast China. *Sci. Rep.* **7**, 10917 (2017).
52. Huong, N. Q. et al. Coronavirus testing indicates transmission risk increases along wildlife supply chains for human consumption in Viet Nam, 2013–2014. *PLoS ONE* **15**, e0237129 (2020).
53. Ithete, N. L. et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis.* **19**, 1697–1699 (2013).
54. Jeong, J. et al. Persistent infections support maintenance of a coronavirus in a population of Australian bats (*Myotis macropus*). *Epidemiol. Infect.* **145**, 2053–2061 (2017).
55. Joffrin, L. et al. Bat coronavirus phylogeography in the Western Indian Ocean. *Sci. Rep.* **10**, 6873 (2020).
56. Kemenesi, G. et al. Molecular survey of RNA viruses in Hungarian bats: discovering novel astroviruses, coronaviruses, and caliciviruses. *Vector Borne Zoonotic Dis.* **14**, 846–855 (2014).
57. Kim, H. K. et al. Detection of severe acute respiratory syndrome-like, Middle East respiratory syndrome-like bat coronaviruses and Group H rotavirus in faeces of Korean bats. *Transbound. Emerg. Dis.* **63**, 365–372 (2016).
58. Kivistö, I. et al. First report of coronaviruses in Northern European bats. *Vector Borne Zoonotic Dis.* **20**, 155–158 (2020).
59. Kudagammana, H. D. W. S. et al. Coronaviruses in guano from *Pteropus medius* bats in Peradeniya, Sri Lanka. *Transbound. Emerg. Dis.* **65**, 1122–1124 (2018).
60. Lacroix, A. et al. Wide diversity of coronaviruses in frugivorous and insectivorous bat species: a pilot study in Guinea, West Africa. *Viruses* **12**, 855 (2020).
61. Lau, S. K. P. et al. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* **367**, 428–439 (2007).
62. Lau, S. K. P. et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* **84**, 2808–2819 (2010).
63. Lau, S. K. P. et al. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of *Pipistrellus* bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J. Virol.* **87**, 8638–8650 (2013).
64. Lau, S. K. P. et al. Novel bat alphacoronaviruses in Southern China support Chinese horseshoe bats as an important reservoir for potential novel coronaviruses. *Viruses* **11**, 423 (2019).
65. Lau, S. K. P. et al. Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault’s rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. *J. Virol.* **86**, 11906–11918 (2012).
66. Lau, S. K. P. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl Acad. Sci. USA* **102**, 14040–14045 (2005).
67. Lazov, C. M. et al. Detection and characterization of distinct alphacoronaviruses in five different bat species in Denmark. *Viruses* **10**, 486 (2018).
68. Lecis, R., Mucedda, M., Pidinchedda, E., Pittau, M. & Alberti, A. Molecular identification of Betacoronavirus in bats from Sardinia (Italy): first detection and phylogeny. *Virus Genes* **55**, 60–67 (2019).
69. Lee, S. et al. Genetic characteristics of coronaviruses from Korean bats in 2016. *Microb. Ecol.* **75**, 174–182 (2018).

70. Lelli, D. et al. Detection of coronaviruses in bats of various species in Italy. *Viruses* **5**, 2679–2689 (2013).
71. Leopardi, S. et al. The close genetic relationship of lineage D Betacoronavirus from Nigerian and Kenyan straw-coloured fruit bats (*Eidolon helvum*) is consistent with the existence of a single epidemiological unit across sub-Saharan Africa. *Virus Genes* **52**, 573–577 (2016).
72. Li, W. et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
73. Liang, J. et al. Detection of diverse viruses in alimentary specimens of bats in Macau. *Viol. Sin.* **32**, 226–234 (2017).
74. Lin, X.-D. et al. Extensive diversity of coronaviruses in bats from China. *Virology* **507**, 1–10 (2017).
75. Luo, C.-M. et al. Discovery of novel bat coronaviruses in South China that use the same receptor as Middle East respiratory syndrome coronavirus. *J. Virol.* **92**, e00116-18 (2018).
76. Luo, Y. et al. Longitudinal surveillance of betacoronaviruses in fruit bats in Yunnan Province, China during 2009–2016. *Viol. Sin.* **33**, 87–95 (2018).
77. Maganga, G. D. et al. Genetic diversity and ecology of coronaviruses hosted by cave-dwelling bats in Gabon. *Sci. Rep.* **10**, 7314 (2020).
78. Memish, Z. A. et al. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19**, 1819–1823 (2013).
79. Mendenhall, I. H. et al. Identification of a lineage D Betacoronavirus in cave nectar bats (*Eonycteris spelaea*) in Singapore and an overview of lineage D reservoir ecology in SE Asian bats. *Transbound. Emerg. Dis.* **64**, 1790–1800 (2017).
80. Misra, V. et al. Detection of polyoma and corona viruses in bats of Canada. *J. Gen. Virol.* **90**, 2015–2022 (2009).
81. Monchatre-Leroy, E. et al. Identification of alpha and beta coronavirus in wildlife species in France: bats, rodents, rabbits, and hedgehogs. *Viruses* **9**, 364 (2017).
82. Moreira-Soto, A. et al. Neotropical bats from Costa Rica harbour diverse coronaviruses. *Zoonoses Public Health* **62**, 501–505 (2015).
83. Moreno, A. et al. Detection and full genome characterization of two beta CoV viruses related to Middle East respiratory syndrome from bats in Italy. *Viol. J.* **14**, 239 (2017).
84. Nziza, J. et al. Coronaviruses detected in bats in close contact with humans in Rwanda. *Ecohealth* **17**, 152–159 (2020).
85. Obameso, J. O. et al. The persistent prevalence and evolution of cross-family recombinant coronavirus GCCDC1 among a bat population: a two-year follow-up. *Sci. China Life Sci.* **60**, 1357–1363 (2017).
86. Osborne, C. et al. Alphacoronaviruses in New World bats: prevalence, persistence, phylogeny, and potential for interaction with humans. *PLoS ONE* **6**, e19156 (2011).
87. Pauly, M. et al. Novel alphacoronaviruses and paramyxoviruses cocirculate with type 1 and severe acute respiratory system (SARS)-related betacoronaviruses in synanthropic bats of Luxembourg. *Appl. Environ. Microbiol.* **83**, e01326-17 (2017).
88. Pfefferle, S. et al. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerg. Infect. Dis.* **15**, 1377–1384 (2009).
89. Poon, L. L. M. et al. Identification of a novel coronavirus in bats. *J. Virol.* **79**, 2001–2009 (2005).
90. Prada, D., Boyd, V., Baker, M. L., O’Dea, M. & Jackson, B. Viral diversity of microbats within the South West Botanical Province of Western Australia. *Viruses* **11**, 1157 (2019).
91. Razanajatovo, N. H. et al. Detection of new genetic variants of Betacoronaviruses in endemic frugivorous bats of Madagascar. *Viol. J.* **12**, 42 (2015).
92. Reusken, C. B. E. M. et al. Circulation of group 2 coronaviruses in a bat species common to urban areas in Western Europe. *Vector Borne Zoonotic Dis.* **10**, 785–791 (2010).
93. Rico Chavez, O. et al. Viral diversity of bat communities in human-dominated landscapes in Mexico. *Vet. Méx. OA.* <https://doi.org/10.21753/vmoa.2.1.344> (2015).
94. Rihtaric, D., Hostnik, P., Steyer, A., Grom, J. & Toplak, I. Identification of SARS-like coronaviruses in horseshoe bats (*Rhinolophus hipposideros*) in Slovenia. *Arch. Virol.* **155**, 507–514 (2010).
95. Rizzo, F. et al. Coronavirus and paramyxovirus in bats from Northwest Italy. *BMC Vet. Res.* **13**, 396 (2017).
96. Seltmann, A. et al. Seasonal fluctuations of astrovirus, but not coronavirus shedding in bats inhabiting human-modified tropical forests. *Ecohealth* **14**, 272–284 (2017).
97. Shehata, M. M. et al. Surveillance for coronaviruses in bats, Lebanon and Egypt, 2013–2015. *Emerg. Infect. Dis.* **22**, 148–150 (2016).
98. Shirato, K. et al. Detection of bat coronaviruses from *Miniopterus fuliginosus* in Japan. *Virus Genes* **44**, 40–44 (2012).
99. Smith, C. S. et al. Coronavirus infection and diversity in bats in the Australasian Region. *Ecohealth* **13**, 72–82 (2016).
100. Su, B.-G., Chen, H. C., Cheng, H.-C. & Chen, Y.-N. Detection of bat coronavirus and specific antibodies in chestnut bat (*Scotophilus kuhlii*) population in Central Taiwan. *Taiwan Vet. J.* **42**, 19–26 (2016).
101. Subudhi, S. et al. A persistently infecting coronavirus in hibernating *Myotis lucifugus*, the North American little brown bat. *J. Gen. Virol.* **98**, 2297–2309 (2017).
102. Suzuki, J., Sato, R., Kobayashi, T., Aoi, T. & Harasawa, R. Group B betacoronavirus in rhinolophid bats, Japan. *J. Vet. Med. Sci.* **76**, 1267–1269 (2014).
103. Tang, X. C. et al. Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* **80**, 7481–7490 (2006).
104. Tao, Y. et al. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. *J. Virol.* **91**, e01953-16 (2017).
105. Tong, S. et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* **15**, 482–485 (2009).
106. Tsuda, S. et al. Genomic and serological detection of bat coronavirus from bats in the Philippines. *Arch. Virol.* **157**, 2349–2355 (2012).
107. Valitutto, M. T. et al. Detection of novel coronaviruses in bats in Myanmar. *PLoS ONE* **15**, e0230802 (2020).
108. Wacharapluesadee, S. et al. Diversity of coronavirus in bats from Eastern Thailand. *Virol. J.* **12**, 57 (2015).
109. Wacharapluesadee, S. et al. Longitudinal study of age-specific pattern of coronavirus infection in Lyle’s flying fox (*Pteropus lylei*) in Thailand. *Virol. J.* **15**, 38 (2018).
110. Wang, L. et al. Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China. *Emerg. Microbes Infect.* **6**, e14 (2017).
111. Wang, N. et al. Characterization of a new member of alphacoronavirus with unique genomic features in *Rhinolophus* bats. *Viruses* **11**, 379 (2019).
112. Waruhiu, C. et al. Molecular detection of viruses in Kenyan bats and discovery of novel astroviruses, caliciviruses and rotaviruses. *Viol. Sin.* **32**, 101–114 (2017).
113. Watanabe, S. et al. Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg. Infect. Dis.* **16**, 1217–1223 (2010).
114. Woo, P. C. Y. et al. Molecular diversity of coronaviruses in bats. *Virology* **351**, 180–187 (2006).
115. Woo, P. C. Y. et al. Rapid detection of MERS coronavirus-like viruses in bats: potential for tracking MERS coronavirus transmission and animal origin. *Emerg. Microbes Infect.* **7**, 18 (2018).
116. Xu, L. et al. Detection and characterization of diverse alpha- and betacoronaviruses from bats in China. *Viol. Sin.* **31**, 69–77 (2016).

117. Yadav, P. D. et al. Detection of coronaviruses in Pteropus & Rousettus species of bats from different States of India. *Indian J. Med. Res.* **151**, 226–235 (2020).
118. Yang, L. et al. MERS-related betacoronavirus in Vespertilio superans bats, China. *Emerg. Infect. Dis.* **20**, 1260–1262 (2014).
119. Yuan, J. et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *J. Gen. Virol.* **91**, 1058–1062 (2010).
120. Yuen, K. Y., Lau, S. K. P. & Woo, P. C. Y. Wild animal surveillance for coronavirus HKU1 and potential variants of other coronaviruses. *Hong. Kong Med. J.* **18**, 25–26 (2012).
121. Zhou, P. et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **556**, 255–258 (2018).
122. Poon, L. L. M. et al. Identification of a novel coronavirus in bats. *J. Virol.* **79**, 2001–2009 (2005).
123. Woo, P. C. Y. et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**, 884–895 (2005).
124. de Souza Luna, L. K. et al. Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J. Clin. Microbiol.* **45**, 1049–1052 (2007).
125. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494 (2019).
126. Washburne, A. D., Silverman, J. D. & Morton, J. T. Phylofactorization: a graph partitioning algorithm to identify phylogenetic scales of ecological data. *Ecol. Monogr.* <https://doi.org/10.1002/ecm.1353> (2019).
127. Cinar, O., Nakagawa, S. & Viechtbauer, W. Phylogenetic multilevel meta-analysis: a simulation study on the importance of modelling the phylogeny. *Methods Ecol. Evol.* **13**, 383–395 (2022).
128. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
129. Latinne, A. et al. Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235 (2020).
130. Wacharapluesadee, S. et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021).
131. Becker, D. J. et al. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe* [https://doi.org/10.1016/s2666-5247\(21\)00245-7](https://doi.org/10.1016/s2666-5247(21)00245-7) (2022).
132. Ruiz-Aravena, M. et al. Ecology, evolution and spillover of coronaviruses from bats. *Nat. Rev. Microbiol.* **20**, 299–314 (2022).
133. The Verena Consortium. Building a global atlas of wildlife disease data. *The Verena Blog* <https://www.viralemergence.org/blog/building-a-global-atlas-of-wildlife-disease-data> (2022).
134. Alves, R. S. et al. Detection of coronavirus in vampire bats (*Desmodus rotundus*) in southern Brazil. *Transbound. Emerg. Dis.* <https://doi.org/10.1111/tbed.14150> (2021).
135. Bergner, L. M., Orton, R. J. & Streicker, D. G. Complete genome sequence of an *Alphacoronavirus* from common vampire bats in Peru. *Microbiol. Resour. Announc.* **9**, e00742 (2020).
136. Becker, D. J. et al. Serum proteomics identifies immune pathways and candidate biomarkers of coronavirus infection in wild vampire bats. *Front. Virol.* <https://doi.org/10.3389/fviro.2022.862961> (2022).
137. Kettenburg, G. et al. Full genome *Nobecovirus* sequences from Malagasy fruit bats define a unique evolutionary history for this coronavirus clade. *Front. Public Health* **10**, 786060 (2022).
138. Hoarau, A. O. G. et al. Investigation of astrovirus, coronavirus and paramyxovirus co-infections in bats in the western Indian Ocean. *Virol. J.* **18**, 205 (2021).
139. Drexler, J. F., Corman, V. M. & Drosten, C. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antivir. Res.* **101**, 45–56 (2014).
140. Plowright, R. K., Becker, D. J., McCallum, H. & Manlove, K. R. Sampling to elucidate the dynamics of infections in reservoir hosts. *Phil. Trans. R. Soc. Lond. B* **374**, 20180336 (2019).
141. Jeong, J. et al. Persistent infections support maintenance of a coronavirus in a population of Australian bats (*Myotis macropus*). *Epidemiol. Infect.* **145**, 2053–2061 (2017).
142. Becker, D. J., Eby, P., Madden, W., Peel, A. J. & Plowright, R. K. Ecological conditions predict the intensity of Hendra virus excretion over space and time from bat reservoir hosts. *Ecol. Lett.* <https://doi.org/10.1111/ele.14007> (2022).
143. Thompson, C. W. et al. Preserve a voucher specimen! The critical need for integrating natural history collections in infectious disease studies. *mBio* **12**, e02698-20 (2021).
144. Moratelli, R. Wildlife biologists are on the right track: a mammalogist's view of specimen collection. *Zoologia* **31**, 413–417 (2014).
145. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Brit. Med. J.* **339**, b2535 (2009).
146. Meteorological versus astronomical seasons. *National Centers for Environmental Information* <https://www.ncei.noaa.gov/news/meteorological-versus-astronomical-seasons> (NOAA, 2016).
147. Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P. & Makowski, D. performance: an R package for assessment, comparison and testing of statistical models. *J. Open Source Softw.* <https://doi.org/10.21105/joss.03139> (2021).
148. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
149. Miller-Butterworth, C. M. et al. A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, *Miniopterus*. *Mol. Biol. Evol.* **24**, 1553–1561 (2007).
150. Orme, D. et al. *Caper: Comparative Analyses of Phylogenetics and Evolution in R*. R Package v.0.5.2 (ScienceOpen, 2012).
151. Fritz, S. A. & Purvis, A. Phylogenetic diversity does not capture body size variation at risk in the world's mammals. *Proc. R. Soc. B* **277**, 2435–2441 (2010).
152. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
153. Crowley, D., Becker, D., Washburne, A. & Plowright, R. Identifying suspect bat reservoirs of emerging infections. *Vaccines* **8**, 228 (2020).
154. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. Stat. Theory Appl.* **6**, 65–70 (1979).
155. Senior, A. M. et al. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology* **97**, 3293–3299 (2016).
156. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).
157. López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W. & Viechtbauer, W. Estimation of the predictive power of the model in mixed-effects meta-regression: a simulation study. *Br. J. Math. Stat. Psychol.* **67**, 30–48 (2014).

Acknowledgements

This work was supported by funding to the Viral Emergence Research Initiative (VERENA) consortium, including NSF BII 2021909 and NSF BII 2213854, as well as by the National Institute of General Medical Sciences of the National Institutes of Health (P20GM134973). L.E.C. received funding from the Ramon Murphy Program for Global Health Education in the Department of Medical Education at the Icahn School of Medicine at Mount Sinai. We thank N. Simmons for helpful feedback on our manuscript.

Author contributions

D.J.B., C.J.C. and L.E.C. devised the study. L.E.C., A.C.F. and B.C. performed the data collection. D.J.B. conducted the geographic and taxonomic analyses. L.E.C. conducted the phylogenetically controlled meta-analysis. L.E.C. and D.J.B. generated all figures and tables. L.E.C., A.C.F., C.J.C. and D.J.B. interpreted the results. L.E.C., A.C.F., C.J.C. and D.J.B. wrote the manuscript. All authors reviewed the manuscript and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-023-01375-1>.

Correspondence and requests for materials should be addressed to Lily E. Cohen.

Peer review information *Nature Microbiology* thanks Ricardo Moratelli, Melville Fenton, Clifton McKee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|--|
| Data collection | Data was collected in Airtable (version 1.5.3) and RStudio (version 2021.9.2 "Ghost Orchid"). The unprocessed data and scripts to generate the primary dataset (and all other derived datasets) are available at www.github.com/viralemergence/batgap ; DOI: 10.5281/zenodo.6644081. |
| Data analysis | Data was analyzed in RStudio (version 2021.9.2 "Ghost Orchid"). The unprocessed data and scripts to generate the primary dataset (and all other derived datasets) and to replicate all analyses and visualizations are available at www.github.com/viralemergence/batgap ; DOI: 10.5281/zenodo.6644081). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The primary dataset is available on Github (www.github.com/viralemergence/datacov; DOI: 10.5281/zenodo.6644163) and is comprised of data extracted from

papers obtained during a systematic search of PubMed (<https://pubmed.ncbi.nlm.nih.gov>), Web of Science (<https://www.webofscience.com>), and Global Health (<https://www.cabdirect.org/globalhealth>).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol, we generate and analyze the first standardized, open database of coronavirus surveillance in bats that provides disaggregated prevalence data (including negative results) at the finest methodological, spatiotemporal, and phylogenetic level of detail possible from public records.
Research sample	A database with over 2,200 estimates of coronavirus infection prevalence or seroprevalence in bats from 110 published studies and 89,752 samples.
Sampling strategy	We followed the PRISMA protocol (Figure 1). We systematically searched Web of Science, PubMed, and Global Health (a database comprising publications from the Public Health and Tropical Medicine database and CAB Abstracts). PubMed searches used the following string: (bat* OR Chiroptera*) AND (coronavirus* OR CoV*). Web of Science and Global Health (comprised of CAB Abstracts and Public Health and Tropical Medicine database) searches used the following string: (bat* OR Chiroptera*) AND (coronavirus* OR CoV*) AND (wild*). Searches were performed on September 24, 2020.
Data collection	L.E.C., A.F., and B.C. collected the data from the primary literature and compiled it in a shared Airtable.
Timing and spatial scale	The database includes studies published between 2005 and 2020. The spatial scale is global.
Data exclusions	Initially, studies were excluded if they did not provide estimate(s) of bat coronavirus prevalence or seroprevalence (e.g., authors only report detection of a novel sequence but not the total bats sampled nor number positive); this criterion was pre-established. During data collection, we excluded studies that provided contradictory data that could not be resolved through communication with authors and studies which did not provide either bat genus or species information. Numbers of studies and records excluded are provided in the PRISMA (Figure S1).
Reproducibility	Unprocessed data and scripts to generate the primary dataset and to reproduce all analyses and visualizations are provided at www.github.com/viralemergence/batgap ; DOI: 10.5281/zenodo.6644081.
Randomization	Study inclusion followed a systematic inclusion and exclusion process and can be considered a random sample of the literature.
Blinding	Blinding was not relevant to the meta-analysis or included in the PRISMA protocol.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging