Supporting Information (SI) Methods.

Compiling the databases. A list of viruses and associated mammalian hosts was obtained from an extensive database of virus-mammal associations published by Olival et al. (1). The Olival et al. (1) databases represent the most comprehensive meta-analysis of trait predictors of zoonotic potential to date, analyzing both host and viral traits associated with zoonosis across phylogenetic groups. Using the information provided in their databases, we extracted associations corresponding to directly transmitted viruses previously PCR-identified or isolated in both mammals and humans, and with evidence of animal-to-human spillover. We excluded viruses classified as zoonotic based on exclusively serological data, and viruses with vector-borne transmission or "spillback" from humans to animals. Viruses such as HIV that have zoonotic origins, but now maintain separate, genetically distinct animal and human transmission cycles (2) were also not included. We supplemented our initial list of virus-mammal associations by cross-referencing other existing virus databases (3–5) and conducting literature searches. For each virus, we confirmed the accuracy, detection quality, and completeness of the mammal associations, resolving any inconsistencies between the referenced database and scientific literature. Through our searches, we additionally identified virus and mammal associations that were missing from our list and added those that met the criteria we outlined above. We compiled a list of 420 virus-mammal associations, which included 278 unique host species and 67 unique zoonotic viruses (*SI Data and Results,* Table S1).

For each virus-mammal association in our database, we conducted a series of literature searches to collect two metrics of zoonotic risk: viruses' human case fatality rates (CFRs) and capacities for human-to-human transmission. We collected CFRs as a proxy for virulence, reporting the mean of the maximum and minimum recorded CFR per zoonosis in the literature. When different CFRs for a given virus could be linked to spillover events from different host species, we reported these distinct host-CFR associations as separate entries in our database. This distinction occurred in the case of two viruses only, Nipah virus, which spills over to humans from both pigs (6) and bats (7), and Marburg virus, which spills over to humans from both primates (8,9) and bats (10). We additionally collected information on each zoonosis' capacity for human-to-human transmission according to a four-point scale, adapted from previously defined classification schemes (5,11–13). We assigned a human transmissibility level of "1" to viruses for which human-to-human transmission had not been recorded; "2" to viruses for which human-to-human transmission had been recorded, but was described as atypical; "3" to viruses for which human-to-human transmission had occurred regularly, but was restricted to self-limiting outbreaks; and "4" to viruses for which endemic human transmission had been reported. We constructed this ranking system based on literature that has defined epidemic outcomes for different levels of human-to-human transmission (12–14). Previous meta-analyses have relied on binary categorization (i.e., pathogens are either capable or incapable of human-to-human transmission) which fails to capture critical nuance. Slight variations in viral capacity for between-human transmission can have a large impact on the outcomes of human epidemics (13). In capturing the full extent of variation in viruses' capacity to transmit between humans, our classification system provides a better foundation for identifying viruses that pose the greatest threat to human populations.

In addition to collecting our targeted metrics of virulence and transmissibility, we classified each virus-mammal association according to the mammal's role in the transmission of the virus. First, we used a binary code to distinguish between reservoir and secondary hosts ("reservoir status"), assigning "1" to mammal species that maintain viruses endemically (reservoir hosts) and "2" to species that harbor the virus but are not implicated in zoonotic

maintenance (secondary hosts). Thus, the "reservoir status" variable identifies the primary selective environment (i.e., reservoir) of viruses in zoonotic transmission cycles. We assigned a second binary code to define each host's role in zoonotic spillover to humans ("spillover capacity"), assigning "1" to mammal species that serve as a source of human infection and "0" to species that have no record of transmission to humans. Thus, the "spillover capacity" variable identifies host species implicated in infecting humans. Combining these two codes, we defined a third "spillover type" code to distinguish between "primary" spillover from a reservoir host species (reservoir status = 1, spillover capacity = 1) and "secondary" spillover from a secondary (bridge) host species (reservoir status = 2, spillover capacity = 1).

In addition to the virus-mammal association database, we compiled a database of all directly-transmitted mammalian viruses, including both zoonotic and non-zoonotic viruses. We extracted directly-transmitted viruses from the extensive database of mammalian viruses published by Olival et al. (1), again excluding viruses with vector-borne transmission, "spillback" from humans to animals, or exclusively human transmission. We collected 7 additional viruses from Geoghegan et al. (3), compiling a total of 345 unique viruses (*SI Data and Results,* Table S2).

Using previously published databases (1,3–5,15–18), we collected a series of host and viral predictor variables that based on the literature, we hypothesized might explain observed variation in zoonotic virus dynamics in human hosts. The literature has identified a suite of life history, ecological, immunological, and biological host factors that can influence host-pathogen coevolution (19–22). In this study, we focused on four life history traits that could be quantified across mammal species: body mass, litter size, gestation period duration, and lifespan. The literature has linked host body mass with the rate of disease progression (23), reservoir competence (24), and pathogen replication rate (25,26). Host reproductive effort trades off with investment in immunity and shapes the demography of the susceptible host population (27,28). Long-lived hosts are associated with heightened transmission in a population (29,30). Our analysis builds off previous work by Luis et al. (31), which considered host life history traits such as body mass, lifespan, and litters per year in a meta-analysis of zoonotic burden across bats and rodents. For viral traits, we first focused on traits previously linked to zoonotic infectivity by Olival et al. (1). In particular, we considered viruses' host ranges, collecting viruses' host phylogenetic breadths from Olival et al. (1), as a pathogen's degree of generalism has been posited to influence the evolution of virulence (32). However, we also included the position of a virus' host breadth relative to humans by considering the maximum host phylogenetic distance from humans across a virus' host range. We collected additional viral traits from the International Committee on Taxonomy of Viruses (ICTV) database (33), tracking genome and DNA/RNA composition, as different viral groups have been associated with different viral outcomes in humans (34).

We obtained these host traits primarily from the PanTHERIA database (16), supplementing missing trait information with the Animal Diversity Web (18), The Encyclopedia of Life (15), AnAge database of animal ageing and longevity (17), and literature searches (*SI Data and Results,* Table S3). We proxied unavailable trait data by averaging across other host species in the same genus, or borrowing data from species in the same family that had similar body masses. We obtained virus taxonomic classification and genome composition from the ICTV database (33). All additional host and viral traits were collected from Olival et al. (1). Viral trait data were unavailable for a small subset of viruses, particularly lesser known viruses; we did not proxy this missing viral data, instead representing unavailable information with NA values. All datasets with metadata and references are available in the *SI Data and Results,* Tables S1–4. Table 2 describes all predictor and response variables used in our analysis.

Olival et al. calculated hosts' phylogenetic distance from humans and viruses' host phylogenetic breadths from a matrix of phylogenetic distances between mammal species. The authors derived the distance values in this matrix from a maximum likelihood phylogenetic tree of mammalian cytochrome *b* sequences using the ape package in R (35,36). Cytochrome *b* is a mitochondrial protein with high sequence variability and availability across mammal species, and as a result, is commonly used to determine phylogenetic relationships between mammals (37). Cytochrome *b* has also been demonstrated to be the most effective mitochondrial genetic marker for reconstructing mammalian phylogenies (38). Many previous studies of parasite sharing between animals and plant species have quantified the degree of phylogenetic difference between host species as the number of years since species diverged in evolutionary history (39–42). Using time since divergence in this context implies that host species share pathogens due to mutual ancestry. However, spillover occurs when a pathogen overcomes genetic barriers to infect a novel host – it does not arise due to shared ancestry between two host species (43). Host genetic factors also determine how the host will respond to infection, which will influence the pathogen's virulence and capacity for transmission in a novel host population (44). Given the lack of a universally identified and available genetic loci for host immune traits, studies of cross-species pathogen emergence have often relied on mitochondrial genetic markers to measure the degree of phylogenetic difference between host species (45,46). Although cytochrome *b* sequences are available for the majority of mammal species, some species in our dataset lacked sufficient sequence data and thus, were excluded from the Olival et al. (1) matrix of phylogenetic distances. We calculated phylogenetic distance values for these missing species by averaging across other host species in the same genus or order.

Statistical analysis. We used generalized additive models (GAMs) in the mgcv package in R (47) to assess host and viral predictors of zoonotic risk because we expected to observe nonlinear relationships. GAMs are a class of flexible generalized linear models that use smooth functions to capture nonlinear relationships between a response and predictor variables as opposed to manually specifying higher order polynomial functions. We fit two sets of GAMs, assessing host predictors of zoonotic risk in one group of models and viral predictors in the other. In all cases, as recommended by the package author, we fixed the number of smoothing knots (k) at 7, which partitioned our nonlinear host and viral trait predictors into different regions for model fitting (47), and fit the GAM via restricted maximum likelihood (REML) estimation.

Our global models included all trait predictors outlined in Table 2, including a citation predictor to control for any potential publication bias—7 host traits for our host models and 8 viral traits for our viral models. We used automated term selection by double penalty smoothing for variable selection by setting select=TRUE within the gam function of mgcv. This method constructs an additional penalty for each GAM smooth function, effectively removing terms without predictive power, and has been recognized as superior or comparable to alternative approaches (48). We set an effective degree of freedom cutoff of 0.001 to identify which terms had been penalized and effectively removed from the model (1).

Host models

We restricted our analysis of host predictors of zoonotic risk to known reservoir host species with demonstrated evidence of animal-to-human spillover (reservoir status = 1, spillover capacity = 1). Thus, our host models only considered species implicated as both the primary selective environment and source of human infection for a given virus. However, because the specific host species responsible for a spillover event is not always identified, we were frequently unable to collect human case fatality rate and transmissibility data that varied

depending on the virus' mammal species of origin. Thus, to avoid pseudoreplication, we further restricted our analysis to include only unique entries for each host order per virus in a simplified dataset. In this simplified dataset, we summarized information across hosts encapsulated in each unique entry by taking the maximum value for each host trait metric. This simplified host order dataset thus included 63 unique viruses (4 of the original 67 viruses were excluded because their reservoir host species is unknown) and 78 unique host order-virus associations for analysis.

Using this simplified dataset, we first asked *what host variables best predict case fatality rates in human hosts following spillover?* We addressed this question using a GAM in the gaussian family. Specifically, we queried the predictive capacity of the host-specific traits outlined in Table 2 (but summarized at the order level) on the response variable of mean case fatality rate in a human host. Typically, case fatality rate response variables were the same across multiple entries for order-level predictors (with the exception of Nipah virus in bats vs. pigs and Marburg virus in bats vs. primates).

We next asked, *what host variables best predict the extent of human-to-human transmission of a given virus following spillover?* Using the same simplified, order-level database adopted for the above question, we addressed this host question with a GAM in the 'ocat' ('ordered categorical data') family. In this case, we queried the predictive capacity of the same host traits outlined in Table 2 on the response variable of human transmissibility on a four-point scale (see 'Compiling the virus-mammal association database'). We defined the vector of categorical cut points, θ, to match our four-point scale for transmissibility (θ = 1,2,3,4).

In addition to the GAMs described above, we explored the relationship between host phylogenetic distance from humans and both case fatality rate and transmissibility as a function of 'spillover type' ("primary" vs. "secondary"). While our previous GAMs only included reservoir host species, here we considered both reservoir and secondary hosts with evidence of spillover to humans. Thus, this analysis considered all 67 viruses (i.e., included viruses for which only secondary host species are known) and 80 unique host order-virus associations. As outlined above in 'compiling the virus-mammal association database', spillover from reservoir hosts was defined as "primary" and spillover from secondary (bridge) hosts was defined as "secondary". To assess differences in the effect of spillover type on the relationship between host phylogenetic distance from humans and case fatality / human transmissibility, we fit a separate GAM for each response variable to this data subset. We queried the response variable of CFR / human transmissibility (respectively, in the gaussian and 'ocat' families), against the predictor variable of host phylogenetic distance, modeled as two distinct smoothers separated by spillover type (using the "by" term in mgcv) (47).

Virus models

For our analysis of viral predictors of zoonotic risk, we first asked *what viral traits best predict the probability that a virus is zoonotic*? Here, we recapitulated previous meta-analyses of viruses' zoonotic potential (i.e., whether an animal virus has the capacity to infect humans). However, unlike most previous work, we included two viral traits that required consideration of mammalian host associations: *(i)* phylogenetic host breadth, which encapsulates the maximum phylogenetic distance between the two most distantly related known hosts for a given virus and *(ii)* maximum host phylogenetic distance from humans, which corresponds to the phylogenetic distance from humans of the most distantly related known host for a given virus (thus serving as an anchor point for *(i)*). Olival et al. (1) likewise considered phylogenetic host breadth in their analysis of viral predictors, but to our knowledge, this is the first analysis of zoonoses to also consider the position of host breadth

relative to humans (i.e., our metric for maximum host phylogenetic distance from humans). For this analysis, we limited the virus database extracted from Olival et al. (1) (*SI Data and Results,* Table S2, N=345) to consider directly-transmitted mammal-infecting viruses only. We then constructed a binomial GAM, testing the predictive capacity of viral traits outlined in Table 2 against the response variable of zoonotic status (0-1, is versus is not).

Our analysis of viral predictors of case fatality and human transmissibility largely mirrored that of our host analysis. We again only considered unique entries to avoid pseudoreplication. Distinct from our host predictor models, which considered unique response values for each host order-virus association, we here grouped by discrete response variables per virus. The subsequent simplified virus dataset thus included 69 unique associations between discrete CFR / transmissibility values across 67 unique zoonotic viruses. Distinct CFRs resulting from distinct spillover events in different regions meant that both Nipah and Marburg virus were represented twice in the dataset (though note that viral traits were not unique within each paired entry). Distinct from our host predictor models, phylogenetic metrics were calculated without exclusion of secondary hosts with no recorded transmission to humans (spillover capacity = 0), since we assumed that viral infection of these hosts constituted an important component of a virus's evolutionary history.

As with host models, we then used this simplified viral dataset to ask, *what viral traits best predict case mortality rates in human hosts following spillover?* and *what viral traits best predict the extent of human-to-human transmission of a given virus following spillover?* We addressed these questions using, respectively, GAMs in the gaussian and ocat families.

References

1. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature. 2017 29;546(7660):646–50.

2. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. Cold Spring Harbor Perspectives in Medicine. 2011 Sep 1;1(1):a006841–a006841.

3. Geoghegan JL, Senior AM, Di Giallonardo F, Holmes EC. Virological factors that increase the transmissibility of emerging human viruses. Proceedings of the National Academy of Sciences. 2016 Apr 12;113(15):4170–5.

4. Walker JW, Han BA, Ott IM, Drake JM. Transmissibility of emerging viral zoonoses. Hsiao CK, editor. PLOS ONE. 2018 Nov 7;13(11):e0206926.

5. Woolhouse MEJ, Brierley L. Epidemiological characteristics of human-infective RNA viruses. Scientific Data. 2018 Feb 20;5:180017.

6. Epstein JH, Field HE, Luby S, Pulliam JRC, Daszak P. Nipah virus: Impact, origins, and causes of emergence. Current Infectious Disease Reports. 2006 Feb;8(1):59–65.

7. Luby S, Rahman M, Hossain M, Blum L, Husain M, Gurley E, et al. Foodborne Transmission of Nipah Virus, Bangladesh. Emerging Infectious Diseases. 2006;12(12):1888–94.

8. Kissling RE, Robinson RQ, Murphy FA, Whitfield SG. Agent of disease contracted from green monkeys. Science. 1968 May 24;160(3830):888–90.

9. Smith CEG, Simpson DIH, Bowen ETW, Zlotnik I. Fatal human disease from Vervet Monkeys. The Lancet. 1967 Nov 25;290(7526):1119–21.

10. Smith DH, Isaacson M, Johnson KM, Bagshawe A, Johnson BK, Swanapoel R. Marburg-virus disease in Kenya. The Lancet. 1982;319(8276):816–20.

11. Brierley L, Pedersen AB, Woolhouse MEJ. Tissue Tropism and Transmission Ecology Predict Virulence of Human RNA Viruses. bioRxiv. 2019 Mar 19. Available from: http://biorxiv.org/lookup/doi/10.1101/581512

12. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature. 2007 May;447(7142):279–83.

13. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, et al. Epidemic Dynamics at the Human-Animal Interface. Science. 2009 Dec 4;326(5958):1362–7.

14. Woolhouse MEJ, Gowtage-Sequeria S. Host Range and Emerging and Reemerging Pathogens. Emerging Infectious Diseases. 2005 Dec;11(12):1842–7.

15. Parr CS, Wilson N, Leary P, Schulz K, Lans K, Walley L, et al. The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. Biodiversity Data Journal. 2014 Apr 29;2:e1079.

16. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. 2009;90(9):2648–2648.

17. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. Human Ageing Genomic Resources: new and updated databases. Nucleic Acids Res. 2018 Jan 4;46(Database issue):D1083–90.

18. Myers P, Espinosa R, Parr CS, Jones T, Hammond GS, Dewey TA. The Animal Diversity Web (online). 2019; Available from: https://animaldiversity.org.

19. May RM, Anderson RM. Parasite—host coevolution. Parasitology. 1990 Jun;100(S1):S89–101.

20. Little TJ, Shuker DM, Colegrave N, Day T, Graham AL. The Coevolution of Virulence: Tolerance in Perspective. Manchester M, editor. PLoS Pathogens. 2010 Sep 9;6(9):e1001006.

21. Anderson RM, May RM. Coevolution of hosts and parasites. Parasitology. 1982 Oct;85(02):411.

22. Gandon S, Agnew P, Michalakis Y. Coevolution between Parasite Virulence and Host Life-History Traits. :15.

23. Cable JM, Enquist BJ, Moses ME. The Allometry of Host-Pathogen Interactions. Ratner A, editor. PLoS ONE. 2007 Nov 7;2(11):e1130.

24. Huang ZYX, de Boer WF, van Langevelde F, Olson V, Blackburn TM, Prins HHT. Species' Life-History Traits Explain Interspecific Variation in Reservoir Competence: A Possible Mechanism Underlying the Dilution Effect. Schneider BS, editor. PLoS ONE. 2013 Jan 24;8(1):e54341.

25. Banerjee S, Perelson AS, Moses M. Modelling the effects of phylogeny and body size on within-host pathogen replication and immune response. Journal of The Royal Society Interface. 2017 Nov;14(136):20170479.

26. Althaus CL. Of mice, macaques and men: scaling of virus dynamics and immune responses. Frontiers in Microbiology. 2015;6:3.

27. Gandon S, Jansen VAA, Van Baalen M. Host life history and the evolution of parasite virulence. Evolution. 2007 May 9;55(5):1056–62.

28. Koella JC, Restif O. Coevolution of parasite virulence and host life history. Ecology Letters. 2001 May 30;4(3):207–14.

29. Hily JM, García A, Moreno A, Plaza M, Wilkinson MD, Fereres A, et al. The Relationship between Host Lifespan and Pathogen Reservoir Potential: An Analysis in the System Arabidopsis thaliana-Cucumber mosaic virus. Melcher U, editor. PLoS Pathogens. 2014 Nov 6;10(11):e1004492.

30. Lidsky PV, Andino R. Protection from epidemics is a driving force for evolution of lifespan setpoints. bioRxiv. 2017 Nov 6. Available from: http://biorxiv.org/lookup/doi/10.1101/215202

31. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc Biol Sci. 2013 Apr 7;280(1756).

32. Leggett HC, Buckling A, Long GH, Boots M. Generalism and the evolution of parasite virulence. Trends in Ecology & Evolution. 2013 Oct;28(10):592–6.

33. Fauquet, C., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. Virus taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, 2018.

34. Han BA, Kramer AM, Drake JM. Global Patterns of Zoonotic Disease in Mammals. Trends in Parasitology. 2016 Jul;32(7):565–77.

35. Poulin R, Mouillot D. Parasite specialization from a phylogenetic perspective: a new index of host specificity. Parasitology. 2003 May;126(5):473–80.

36. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004 Jan 22;20(2):289–90.

37. Castresana J. Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals. Molecular Biology and Evolution. 2001 Apr 1;18(4):465–71.

38. Tobe SS, Kitchener AC, Linacre AMT. Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome B and cytochrome oxidase subunit I mitochondrial genes. PLoS ONE. 2010 Nov 30;5(11):e14156.

39. Parker IM, Saunders M, Bontrager M, Weitz AP, Hendricks R, Magarey R, et al. Phylogenetic structure and host abundance drive disease pressure in communities. Nature. 2015 Apr;520(7548):542–4.

40. Park AW, Farrell MJ, Schmidt JP, Huang S, Dallas TA, Pappalardo P, et al. Characterizing the phylogenetic specialism-generalism spectrum of mammal parasites. Proc Biol Sci. 2018 14;285(1874).

41. Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen–host range. PNAS. 2007 Mar 20;104(12):4979–83.

42. Cooper N, Griffin R, Franz M, Omotayo M, Nunn CL. Phylogenetic host specificity and understanding parasite sharing in primates. Fryxell J, editor. Ecology Letters. 2012 Dec;15(12):1370–7.

43. Warren CJ, Sawyer SL. How host genetics dictates successful viral zoonosis. PLOS Biology. 2019 Apr 19;17(4):e3000217.

44. Nolan D, Gaudieri S, Mallal S. Host genetics and viral infections: immunology taught by viruses, virology taught by the immune system. Current Opinion in Immunology. 2006 Aug 1;18(4):413–21.

45. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science. 2010 Aug 6;329(5992):676–9.

46. Washburne AD, Crowley DE, Becker DJ, Olival KJ, Taylor M, Munster VJ, et al. Taxonomic patterns in the zoonotic potential of mammalian viruses. PeerJ. 2018 Nov 28;6:e5979.

47. Wood SN. mgcv: GAMs and generalized ridge regression for R. R News. 1(2):20–5.

48. Marra G, Wood SN. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis. 2011 Jul;55(7):2372–87.