Taxonomic patterns in the zoonotic potential of mammalian viruses

Alex D Washburne Corresp., [1] , **Daniel E Crowley** [1] , **Daniel J Becker** [1] , **Kevin J Olival** [2] , **Matthew Taylor** [1] , **Vincent J Munster** [3] , **Raina K Plowright** [1]

[1] Department of Microbiology and Immunology, Montana State University, Bozeman, United States

[2] Ecohealth Alliance, New York, United States

[3] National Institute of Allergy and Infectious Disease, Hamilton, Montana, United States

Corresponding Author: Alex D Washburne
Email address: bigalculus@gmail.com

P r edicting and simplifying which pathogens may spill over from animals to humans is a major priority in infectious disease biology. Many efforts to determine which viruses are at risk of spillover use a subset of viral traits to find trait-based associations with spillover.

We adapt a new method - phylofactorization - to identify not traits but lineages of viruses at risk of spilling over. Phylofactorization is used to partition the ICTV viral taxonomy based on non-human host range of viruses and whether there exists evidence the viruses have infected humans. We identify clades on a range of taxonomic levels with high or low propensities to spillover, thereby simplifying the classification of zoonotic potential of mammalian viruses.

[p] P hylofactorization by whether a virus is zoonotic yields many disjoint clades of viruses containing few to no representatives that have spilled over to humans. Phylofactorization by non-human host breadth yields several clades with significantly higher host
breadth.

We connect the phylogenetic factors above with life-histories of clades, revisit trait-based analyses, and illustrate how cladistic coarse-graining of zoonotic potential can refine trait-based analyses by illuminating clade-specific determinants of spillover risk. [p]

Taxonomic patterns in the zoonotic potential of mammalian viruses

Alex D. Washburne[1*], Daniel E. Crowley[1], Daniel J. Becker[1,2], Kevin J. Olival[3], Matthew Taylor[1], Vincent J. Munster[4], Raina K. Plowright[1]

[1]*Department of Microbiology and Immunology, Montana State University, Bozeman MT*
[2]*Center for the Ecology of Infectious Disease, University of Georgia, Athens, GA*
[3]*EcoHealth Alliance New York, NY*
[4]*National Institute of Allergy and Infectious Diseases, Hamilton, MT*
[*]*Corresponding author: alex.d.washburne@gmail.com*

Abstract

Predicting and simplifying which pathogens may spill over from animals to humans is a major priority in infectious disease biology. Many efforts to determine which viruses are at risk of spillover use a subset of viral traits to find trait-based associations with spillover.

We adapt a new method - phylofactorization - to identify not traits but lineages of viruses at risk of spilling over. Phylofactorization is used to partition the ICTV viral taxonomy based on non-human host range of viruses and whether there exists evidence the viruses have infected humans. We identify clades on a range of taxonomic levels with high or low propensities to spillover, thereby simplifying the classification of zoonotic potential of mammalian viruses.

Phylofactorization by whether a virus is zoonotic yields many disjoint clades of viruses containing few to no representatives that have spilled over to humans. Phylofactorization by non-human host breadth yields several clades with significantly higher host breadth.

We connect the phylogenetic factors above with life-histories of clades, revisit trait-based analyses, and illustrate how cladistic coarse-graining of zoonotic potential can refine trait-based analyses by illuminating clade-specific determinants of spillover risk.

Introduction

Zoonotic spillover, the transmission of pathogens from non-human vertebrate animals to humans, is a major public health challenge yet the prediction and management of which pathogens spillover is poorly understood (Plowright et al. 2017). Recent, high-profile spillover events, such as that of Nipah virus (Chua et al. 2000; Pulliam et al. 2012), Hendra virus (Murray et al. 1995; Plowright et al. 2015), avian influenza virus (Fouchier et al. 2004; Li et al. 2014), Ebola virus (Baize et al. 2014), rabies virus (Messenger, Smith, and Rupprecht 2002; Schneider et al. 2009), Chikungunya virus (Powers and Logue 2007), Zika virus (Wikan and Smith 2016), and West Nile virus (Petersen and Roehrig 2001; Hayes and Gubler 2006), among others, strongly motivate a broader understanding of the trait determinants and viral clades at risk of pathogen spillover (Rosenberg 2015; Geoghegan et al. 2016; Plowright et al. 2016; Geoghegan and Holmes 2017; Lloyd-Smith et al. 2009).

Emerging infectious diseases arise from a pre-existing pool of viruses with traits, life-histories, and evolutionary histories, all of which interact to determine the propensity for pathogen spillover. With thousands of viruses in nature, understanding which clades of viruses are most prone to zoonotic spillover is critical for improving pathogen surveillance efforts and designing public health interventions across scales from reservoir hosts to humans.

A recent study examined viral traits that predicted a virus' zoonotic status; after controlling for research effort they found that predictors included the phylogenetic host breadth of the virus, genome length, and whether or not the virus was vector borne, reproduced in the cytoplasm, or was enveloped (Olival et al. 2017). Trait-based analyses such as these are insightful and useful, but can be limited by incomplete characterization of traits associated with zoonoses and confounded or biased by non-zoonotic clades – clades with dramatically lower rates of zoonosis - and whose homologous traits can all become negatively associated with zoonosis. Cladistic analyses complement trait-based analyses by identifying monophyletic clades or taxonomic groups with common patterns of zoonosis (Figure 1).

Cladistic analyses can enhance and focus trait-based analyses to refine an understanding of the causes of viral zoonoses. Viral life histories - the patterns and traits determining survival and reproduction - vary greatly across lineages with vastly different molecular machinery for host cell entry, viral gene expression, replication, assembly and dissemination (Jane Flint et al. 2015). General traits, such genetic material or presence of an envelope, may miss the specific molecular mechanisms, such as promiscuous receptor binding or immunoevasion, by which viral quasispecies from a particular clade can become capable of infecting humans. Furthermore, general traits found across all viruses may be less actionable for management interventions than clade-specific mechanisms. For example, knowing the particular species of mosquito, *Aedes aegypti*, responsible for transmitting many flaviviruses (Black et al. 2002) enables ecological interventions aimed at modifying *A. aegypti* susceptibility to *Flavivirus* infection, population density, and more (Morrison et al. 2008). Identifying viral clades with unusually high or low propensity for spillover can prioritize surveillance effort and generate detailed, clade-specific hypotheses about molecular mechanisms and pathogen life-history traits driving spillover risk. Furthermore, identification of non-zoonotic clades can allow refined trait-based analyses: instead of “which traits are associated with zoonosis”, one can ask “which traits are associated with zoonosis for those pathogens at risk of spilling over?” Identifying molecular mechanisms and hypothesized life-history traits for clade-specific determinants of spillover can revise the list of traits used in trait-based analyses (Figure 1).

To build off the recent study of viral traits determining spillover, we apply a new biological machine-learning method, phylofactorization (A. D. Washburne, Silverman, Leff, et al. 2017; A. D. Washburne, Silverman, Morton, et al. 2017), to identify clades of mammalian viruses with high within-clade similarity in zoonotic potential and high between-clade differences. Phylofactorization is a greedy, graph-partitioning algorithm which partitions a tree, in our case the International Committee on Taxonomy of Viruses (ICTV) taxonomy tree, by identifying edges with the most significant differences in the response variable. Using a cladistic taxonomy tree as a scaffold for identifying clades with common propensities for spillover can flexibly identify the range of taxonomic scales of interest and implicitly identify life history traits which enable viruses to spill over into humans. We analyze viruses from a recently published dataset of mammalian viruses (Olival et al. 2017) and replicate the paper’s trait-based analysis alongside our cladistic analysis to provide a side-by-side comparison of how a cladistic analysis informs, complements and, in some cases, changes the results of trait-based analyses.

We perform phylofactorization by whether or not a clade has infected humans, discover several non-zoonotic clades, and remove the non-zoonotic clades from a replicated trait-based analysis. We then perform phylofactorization by non-human host breadth, as it is a related

measurement of the host-tropism of viruses and a robust predictor of whether/not a virus can infect humans. Host-breadth phylofactorization yields a set of clades with high host breadth and high propensity to spillover. Having identified non-zoonotic clades and clades with high non-human host breadth and a high risk of zoonosis, we finish with a discussion of the life-histories of two somewhat opposite clades on the spectrum of spillover risk: the families *Togaviridae* and *Herpesviridae*. The life-history discussion is used to motivate a level of molecular and mechanistic detail which can enable clade-specific surveillance and management of pathogen spillover.

Methods

Data collection

Olival et al. (Olival et al. 2017) presented the most comprehensive analysis of mammalian host–virus relationships to date, spanning 2,805 mammal–virus associations (754 mammal species and 586 unique viral species from 28 viral families). We obtained these data and associated scripts for model selection from the EcoHealth Alliance HP3 github repository on September 20, 2017. The ICTV taxonomy tree was constructed with arbitrary branch lenghts using the R package ape version 5.4 (Paradis, Claude, and Strimmer 2004) in R version 3.4. Phylofactorization was implemented with the R package phylofactor version 0.0.1 available at https://github.com/reptalex/phylofactor.

Phylofactorization

Phylofactorization partitions the phylogeny by defining a contrast function along edges (A. D. Washburne, Silverman, Leff, et al. 2017; A. Washburne 2017), identifying the edge which maximizes an objective of the contrast function, and iteratively partitions the tree along objective-maximizing edges to ensure non-overlapping contrasts. Phylofactorization is used to identify edges (ancestral lineages) along which major changes in zoonotic propensity or non-human host-breadth occurred, allowing us to identify clades for focused natural history analysis and inclusion as categorical variables in predictions of zoonotic risk for discrete surveillance prioritization. While phylofactorization can be used in conjunction with time-reversible ancestral state reconstruction to avoid the nested dependence of other root-to-tip ancestral state reconstruction methods, the absence of well-defined branch lengths on the ICTV taxonomy would make dubious such methods dependent upon explicit models of evolution, and so we use two-sample tests to identify edges along which major changes occurred.

For phylofactorization of Bernoulli-distributed zoonosis data, we used Fisher's exact test as a contrast function and the inverse of its P-value as the objective function. For phylofactorization of real-valued host-breadth data, we used a Wilcox test, and similarly used its inverse-P-value as the objective function. Null simulations (n=350) were run for both phylofactorizations and only those factors whose objective function is larger than the 95% quantile from null simulations of the corresponding factor were kept. This resulted in 10 factors of zoonosis, and 9 factors for host-breadth.

To compare the effect of non-zoonotic clades on trait determinants of spillover, we removed the first 5 factors from zoonosis phylofactorization, all of which had significantly lower rates of zoonosis than the rest (*Papillomaviridae*, *Herpesviridae*, *Orthoretrovirinae*, *Nidovirales*

and *Parvovirus*). After removing non-zoonotic clades, we repeated the trait-based(Olival et al. 2017) classification of whether or not a virus has been observed as zoonotic using the best-fit, all data model from Olival et al. (Olival et al. 2017). Similarly, to determine whether/not the taxa from host-breadth phylofactorization predict zoonosis, we replicated the original model selection from Olival et al., but included the various taxonomic levels identified as a multilevel factor.

Classifying Zoonotic Transmission of Phylofactorization-derived Clades

Classification of host-breadth clades with zoonotic transmission traits required curation of additional trait data for the 170 zoonoses in the data. We followed a systematic process for trait data collection (Haddaway and Watson 2016). To determine human-to-human transmission, we supplemented data from Geoghegan et al. (Geoghegan et al. 2016) using eight reference textbooks in virology (supplemental material) and a targeted search for each zoonotic virus in GoogleScholar (see supplemental material for search strings). Following Plowright et al. (Plowright et al. 2017), we next determined zoonotic transmission through the three primary routes of pathogen release: vector-borne transmission, reservoir excretion, and reservoir host slaughter. We expanded datasets from (Olival et al. 2017), (Jones et al. 2008) and (Plourde et al. 2017) by screening the same texts and with a similar targeted GoogleScholar search. Pathogen release was recorded as three binary covariates for whether zoonotic transmission occurs through excretion, slaughter, or vectors; these categories are not mutually exclusive. For all zoonotic viruses within each host-breadth clade, we tabulated both the proportion of each trait (onward human transmission, vector transmission, transmission through excretion, and transmission through slaughter) and the proportion of records for which values were unknown.

Results

Zoonotic Phylofactorization & Comparative trait-based analysis

Phylofactorization of whether or not a virus is zoonotic yielded 10 significant clades of various sizes and taxonomic levels (Figure 2). The viral family *Papillomaviridae*, family *Herpesviridae*, subfamily *Orthoretrovirinae*, order *Nidovirales*, genus *Parvovirus*, and families *Calciviridae*, *Adenoviridae*, and *Astroviridae* all had a lower fraction of zoonotic representatives compared to the paraphyletic remainder, which at the end of phylofactorization had 352 viruses of which 43.5% (n=153) were zoonotic. The genera *Alphavirus* and *Deltaretrovirus* had significantly high proportions of zoonotic representatives, with 64% of the 25 *Alphavirus* species being zoonotic and 100% of the 4 *Deltaretrovirus* species being zoonotic.

Next, we excluded the top-5 taxonomic factors found to be non-zoonotic clades - *Parvoviridae*, *Herpesviridae*, *Orthoretrovirinae*, *Nidovirales* and *Parvovirus* - and replicated the trait-based classification of zoonotic viruses from previous work (Olival et al. 2017). Excluding non-zoonotic clades affects the significance and explanatory power of vector-borne transmission, whether or not a virus is enveloped, and whether/not a virus replicates within the cytoplasm (Table 1). The sensitivity of results reported in previous trait-based analyses is due to strong associations between these traits and non-zoonotic clades identified through phylofactorization. Non-human host phylogenetic breadth remained a highly significant predictor of zoonosis, suggesting that among clades of viruses likely to spillover, the ability to infect

disparate mammalian hosts confers a significant ability to infect humans. To perform a similar cladistic analysis of this stable predictor of a virus' ability to infect humans, non-human host breadth was used for subsequent phylofactorization to identify any phylogenetic patterns in host breadth.

Host-Breadth Phylofactorization

Phylofactorization by host breadth yielded 9 significant clades with a range of host breadth, each clade being partitioned as having a higher host-breadth than the paraphyletic remainder. Specifically, the viral family *Togaviridae*, genus *Arenavirus*, genus *Flavirus*, genus *Orthopoxvirus*, family *Picornaviridae*, family *Bunyaviridae*, order *Mononegavirales*, family *Reoviridae* and subfamily *Alphaherpesvirinae* all had signifiicantly higher host breadth than the paraphyletic remainder (Figure 3). The family *Togaviridae* contains the genus *Alphavirus* which was also obtained in the zoonosis-phylofactorization. The subfamily *Alphaherpesvirinae*, in the family *Herpesviridae* also identified in the zoonosis-phylofactorization, was identified as having higher host breadth than the remainder of *Herpesviridae* yet the lowest host breadth of the significant, monophyletic clades obtained by phylofactorization.

In order to determine if the clades partitioned by host breadth are significant determinants of zoonotic spillover, the clades identified through host-breadth phylofactorization were then Included as a multilevel factor in a replicated trait-based analysis and model selection (Olival et al. 2017). Incorporating the multilevel factor of host-breadth clades dramatically affected the resulting model: vector-borne transmission, viral envelopes, and cytoplasmic replication - all traits identified in model selection when not controlling for host-breadth clades - are all dropped during model selection when host-breadth clades are included as a multilevel explanatory factor.

The linear predictors of zoonotic clades under the resulting generalized additive model suggest a prioritization scheme of host-breadth clades based on the log-odds of zoonosis within clades factored by host breadth (Figure 4). At the top of the list is *Togaviridae* and at the bottom of the list is the paraphyletic remainder. The *Togaviridae*, *Flavivirus*, and *Bunyaviridae* clades partitioned by phylogenetic breadth are at high risk of spillover into humans; the viruses in these clades are enveloped and almost exclusively transmitted to humans through arthropod vectors. The *Picornaviridae* and *Orthopoxvirus* had lower proportions of zoonoses but higher propensities for onwards transmission in humans. *Alphaherpesvirinae* was partitioned from the paraphyletic remainder as having a slightly higher phylogenetic breadth than the remainder, which included the *Beta*- and *Gammaherpesvirinae*; however, compared to other clades partitioned by host breadth, the *Alphapherpesvirinae* have a low propensity to spillover into people.

Discussion

Phylofactorization of mammalian viruses

Emerging infectious diseases arise from a pre-existing pool of viruses with traits, life-histories, and evolutionary histories, all of which interact to determine the propensity for spillover. Any one virus has myriad traits adapted to the virus' life cycle, ranging from

interactions with a host cell that facilitate infection and replication to the processes that facilitate transmission and environmental persistence. Life histories determine the evolutionary pressure and the constellation of traits defining a virus, and using the evolutionary tree as a scaffold for identifying clades with common propensities for spillover can implicitly identify traits and life-histories which enable viruses to spill over into humans. Modeling, prediction, and management of pathogen spillover requires a comprehensive approach that incorporates clade-specific life histories to identify clade-specific traits that enable pathogens to jump from wildlife to people.

In this paper, we have built on a recent study of traits determining spillover in mammalian viruses (Olival et al. 2017) by using cladistic algorithms to partition the viral taxonomy and classify viral zoonoses. Using a modern graph-partitioning technique built for evolutionary trees (A. D. Washburne, Silverman, Leff, et al. 2017; A. Washburne 2017), we have partitioned the ICTV taxonomy tree and flexibly identified 10 clades of viruses across taxonomic scales with significantly extreme (high or low) propensities for zoonosis.

Removing non-zoonotic clades from the existing trait-based prediction of zoonotic viruses decreased the importance of vector-borne and cytoplasmic replication traits, but increases the importance of enveloped viruses, due to strong, negative associations of these traits with non-zoonotic clades of viruses. One trait remains a robust predictor of zoonosis: the phylogenetic breadth of non-human hosts.

Phylofactorization of the viral taxonomy by non-human host breadth yields 9 clades of viruses at various taxonomic levels. These 9 clades classify viruses with a range of zoonotic potential, even when controlling for host breadth. Identifying phylogenetic factors which can predict zoonosis enables researchers to take a fresh look at the problem of pathogen spillover by focusing on the life-histories of these 9 clades.

Non-human host breadth is a particularly useful trait to use for phylofactorization. Zoonosis combines human:host exposure with the capacity of viruses to cross species barriers. Clades of viruses with molecular mechanisms enabling spillover into human populations, yet currently absent from humans due to infrequent exposure (i.e. viruses in remote regions or hosts, or viruses that are not shed from their hosts) will not be identified under an analysis of “zoonosis” as a response variable but can be identified by an analysis of host-breadth. Likewise, using zoonoses as the outcome variable for trait-based analyses may overemphasize the importance of traits of viruses with high human:host exposure, such as viruses that require large doses to initiate infections but manage to infect humans due to repeated exposure increasing the chance that a person is exposed to a sufficient dose to initiate infection. By performing phylofactorization on both human zoonosis and non-human host breadth, we have identified clades with risks of spillover in both humans and wildlife. There were converging lines of evidence on multiple clades clades in phylofactorization of both human zoonosis and host breadth: *Togaviridae* are highly zoonotic and have a broad host-range, while Herpesviridae have low rates of zoonosis and narrow host ranges; finer detail is observed in the alphaviruses’ propensity for zoonosis and the *Alphaherpesvirinae* having higher phylogenetic breadth than other Herpesviridae while, compared to other clades, still having a lower propensity to spillover. We examine possible life history traits of these and other clades below.

Life history traits of zoonotic and non-zoonotic clades

The family *Togaviridae*, containing the genus *Alphavirus*, is a family with a high host-breadth and a high propensity for spilling over into humans. Togaviruses are vector-borne through mosquitoes and ticks, and such life-history may have selected generalist pathogens and produced a collection of traits shared within the lineage which enable proliferation in the blood of numerous vertebrate hosts. The Togaviruses are generalist pathogens with high host-breadth and propensities for spillover, and their host breadth requires proteins which enable receptor-binding and replication in cells from different species. Identifying these determinants of host-breadth in Togaviruses may focus attention on the key traits enabling pathogen spillover in this highly zoonotic clade.

It's hypothesized that Chikungunya virus, an *Alphavirus*, binds receptors ubiquitously expressed among species and cell types (van Duijl-Richter et al. 2015), and that the Sindbis virus binds the laminin receptor (Wang et al. 1992), heparan sulfate (Wang et al. 1992; Byrnes and Griffin 1998), and the C-type lectins DC-SIGN and L-SIGN (Klimstra et al. 2003), all of which are found on most cell surfaces of most animal tissues. The broad range of cellular receptors utilized by alphaviruses is attributable to a common protein shared among alphaviruses: the E2 protein (Voss et al. 2010). The E2 protein contains residues critical for broad host range and tissue tropism. and thus may be a key trait underlying the zoonotic potential of Togaviruses. Togaviruses all contain 6 genes, and sequence-based surveillance targeting all Togaviruses can improve assessments of spillover risk among mammalian viruses. Furthermore, targeted sequencing of genes particular to Togavirus life history - in particular, variation in the E2 protein - may improve assessments of spillover risk within *Togaviridae* (Klimstra et al. 2003).

At the opposite end of the spillover spectrum from *Togaviridae* is the the viral family *Herpesviridae*, which has a low propensity to be zoonotic. The subfamily *Alphaherpesvirinae* has low non-human host breadth relative to the other clades pulled out in phylofactorization, but it was identified as having a higher host breadth than the *Beta*- and *Gammaherpesvirinae* and the paraphyletic remainder of viruses from which it was partitioned.

Herpesviruses have a range of modes of transmission, often establish lifelong infections, and have several methods to evade the immune system (Roizman 1982). Cross-species transmission events of Herpesviruses have been documented for closely related host species (Woźniakowski and Samorek-Salamonowicz 2015), such as cases of fatal human herpesvirus infection in wild primates (Heldstab et al. 1981) and the converse lethal infections of humans exposed to Herpes B from macaques (Huff and Barry 2003). However, herpesviruses generally show relatively strict species specificity and difficulty with cross-species transmission, including when non-human primates consume the tissues of other non-human primates (Murthy et al. 2013). The reasons for the species-specificity of Herpesviridae are multifactorial, but careful understanding of their life history illuminates some compelling hypotheses for clade-specific barriers to spillover.

One possibility for the species specificity of Herpesviridae is tissue tropism. *Herpesviridae* infections are often associated with terminally-differentiated cells, like neurons and immune cells. Entry into these cells may require binding receptors which are well-known to have undergone positive selection among mammalian lineages (Kosiol et al. 2008). However, the generalization that *Herpesviridae* are tissue-tropic and infect immune cells is not universal; it is particularly true for *Gammaherpesvirinae*, somewhat true for *Betaherpesvirinae*, and not at all

true for *Alphaherpesvirinae*. Interestingly, the *Alphaherpesvirinae* were identified from phylofactorization as having a higher host-breadth than the remaining *Herpesviridae*, suggesting that tissue-tropism may be a clade-specific barrier to spillover in the *Herpesviridae*.

Immune evasion also plays a major role in Herpesviridae life history and the mechanisms of immune evasion in Herpesviruses may limit their host range. For example, the ICP47 gene of Herpes Simplex virus type 1 (HSV-1) is involved in immune evasion through the inhibition of TAP-mediated antigen transport; the ICP47 protein exhibits a 100-fold decrease in its binding affinity for mouse TAP compared to human TAP (K. Ahn et al. 1996). TAP-inhibition has also been documented in human cytomegalovirus (Kwangseog Ahn et al. 1997), equine herpesvirus 1 (Ambagala, Gopinath, and Srikumaran 2004), and bovine herpesvirus 1 (Koppers-Lalic et al. 2003). TAP inhibition is only one of many mechanisms of immunoevasion within the *Herpesviridae*. Host-specificity for each member of the *Herpesviridae* might be related to mechanisms for immunoevasion, like TAP-binding, which require viral proteins interfacing with the conserved pathways yet species-specific enzyme structure underlying mammalian immune systems. The immunological barriers to spillover have been documented not just in the family *Herpesviridae*, but also in Ebola virus (Groseth et al. 2012) and Zika virus (Xia et al. 2018) - in all cases, the species-specificity of protein:protein interactions can be a barrier to spillover. For the many *Herpesviridae* whose life cycles involve latent or chronic infections, unsuccessful immunoevasion in new hosts may be especially costly.

Similar life-history case studies of viral clades partitioned here, and the clade-specific barriers to spillover, may greatly improve our understanding and surveillance of viral spillover. Clades such as the flaviviruses and bunyaviruses - both predominantly vector-borne and enveloped (Figure 4) - corroborate the notion of generalist pathogens and each clade may have genes determining spillover which are particular to that clade. Arenaviruses have high host breadth (Figure 3) yet comparatively low zoonotic propensity (Figure 4). Arenaviruses bind to α-dystroglycan protein as a receptor, enabling broad cell tropism (Meyer, De La Torre, and Southern 2002), yet arenaviruses contain a low proportion of viruses which have spilled over into humans. Unlike the togaviruses, bunyaviruses and flaviviruses, arenaviruses are transmitted to humans primarily from direct or aerosolized exposure to rodent excreta (Emonet et al. 2007; Gonzalez et al. 2007). On one hand, the limited frequency of such exposures makes rodent-human transmission inefficient and less likely to result in a human-rodent zoonosis (Bausch and Mills 2014); on the other hand, the high incidence of Lassa hemorrhagic fever and onward human-human transmission of the Lassa virus suggests such transmission-related bottlenecks may be bypassed with appropriate adaptations. If exposure from reservoirs to humans is the main bottleneck for the zoonosis of many arenaviruses, the sensitivity of aerosolized arenaviruses to ultraviolet radiation, pH, and temperature may modulate the likelihood for cross-species transmission (Stephenson, Larson, and Dominik 1984; Gonzalez et al. 2007; Sagripanti and Lytle 2011), and thus genetic determinants of environmental persistence may be important for spillover surveillance in arenaviruses.

The case studies and hypothesized barriers to spillover expounded here are not exhaustive nor intended to be authoritative; they are intended to remind of the complexity of viral lineages' life histories and illustrate pipeline from cladistic analysis of spillover, life history classification of identified clades, and consideration of new traits and molecular determinants clade-specific risks of zoonosis (Figure 1). More broadly, bottlenecks to pathogen spillover may

occur at one or several of the stages in the viral life cycle and the contact process between reservoir hosts and humans. Entry, viral gene expression, replication, assembly, dissemination, and human-reservoir contact are all processes which may produce clade-specific bottlenecks to spillover. Additionally, there may be multiple stages of the viral life cycle for which viruses contain adaptations which allow them to specialize or generalize across hosts.

Conclusions

Partitioning viruses into monophyletic clades with common propensities for spillover can simplify the problem of pathogen spillover, assist primer design for genomic surveillance (Gardy and Loman 2017) and increase our resolution for understanding clade-specific barriers to spillover. We've reduced a dataset of 586 viruses into 9 clades with different patterns of spillover and, likely, common sequences that can be used for primer design in sequencing-based surveillance. The different clades have different life-histories, ranging from acute infections, vector-borne transmission, envelope-mediated immune evasion and promiscuous receptor-binding in *Togaviridae* to the persistent-recurrent infections, non-vector-borne transmission and species-specific immune evasion mechanisms of *Herpesviridae*.

Understanding life-histories in a handful of clades can focus surveillance efforts on clades most likely to spill-over and target particular genes believed to underlie spillover risk within the focal clade. In vitro studies of molecular mechanisms within high-risk clades, such as the togaviruses, flaviviruses and bunyaviruses, can determine the risk of jump-capable pathogens arising from viral variants in reservoir hosts and identify clade-specific life-history bottlenecks to spillover. Future trait-based analyses can incorporate life-history details obtained from cladistic analyses. Similarly, future cladistic analyses can focus study on viral life histories to provide a more complete understanding of the traits correlated with pathogen spillover.

Acknowledgements

This research was developed with funding from the U.S. Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. We thank members of the Plowright and Cross groups at Montana State University and the USGS for feedback.

Figure Legends

Figure 1: Cladistic analyses complement trait-based analyses. Understanding which clades of viruses spillover can improve our understanding of life-history and trait-based predictors of zoonosis. A thought experiment of predicting flight in a subset of 10 vertebrates, including reptiles, birds and mammals reveals the importance of cladistic analyses. Trait-based analyses aim to identify explanatory traits correlated with a focal response trait, such as identifying feathers and hollow bones as associated with flight in the species observed. We propose a method outlined by the arrows. First, cladistic analysis can identify clades with a common pattern in the response variable. The clade highlighted in green - birds - has more representatives who fly than others. Then, cladistic analyses can allow life-history classification of clades to generate hypotheses about which traits (e.g. wings) and even clade labels (e.g. birds) should be included for amend trait-based analyses. Some traits, such as feathers, hollow bones, wings, and living in trees are dependent upon one-another and a consequence of life histories, and life-histories can reveal the more nuanced sets of traits, including clade-specific traits, which may underlie pathogen spillover. In this paper, for example, we find *Togaviridae* to be at risk for spilling over and discuss how previously unconsidered traits, such as variants of the E2 protein shared among *Togaviridae*, may be clade-specific determinants of spillover.

Figure 2: Phylogenetic factors of zoonosis in mammalian viruses. Phylogenetic factorization iteratively partitions a phylogeny along edges separating species with meaningful differences. Phylofactorization of the ICTV taxonomy by a Fisher test on the fraction of zoonotic viruses identified clades with different numbers of species (n) and different rates of zoonosis (% Zoonotic). 32% of the viruses in the original dataset of mammalian viruses are zoonotic, indicated as black dots on the tip of the tree. The first 8 phylogenetic factors identify clades whose distinction is a significantly low rate of zoonosis relative to other viruses. Deltaretroviruses and Alphaviruses are then identified as having an unusually high fraction of zoonotic viruses. Cladistic structure in zoonotic potential can be used to identify previously ignored traits and prioritize surveillance programs.

Table 1: Effect sizes and significance of traits pre- and post-cladistic analysis. Traits such as enveloped, vector-borne, and cytoplasmic replication are sensitive due to their cladistic signal. Here, effect sizes (difference in linear predictors) are presented for the all viruses in the original trait-based analysis, the zoonotic clades (removal of non-zoonotic clades found in Figure 2), and non-zoonotic clades. The direction of the effect of all three traits is conserved, but the magnitude and relative significance changes, as these traits are strongly associated with non-zoonotic clades. Asterisks indicate significance levels $P<0.05$ (*), $P<0.01$ (**), $P<0.001$ (***) from Chi-squared tests of the deviance from generalized additive models.

Figure 3: Phylogenetic factors of non-human mammalian host breadth. (A) Non-human phylogenetic distance on a cytochrome B phylogeny is a reliable predictor of zoonotic viruses, even upon removal of non-zoonotic clades identified in figure 2. Zoonotic prediction is the linear predictor from generalized additive modelling, plotted here against phylogenetic breadth. The exclusion of major non-zoonotic clades did not dramatically affect the relationship found in the original analysis by Olival et al. (2017). (B) Phylofactorization of viruses by mammalian host breadth identifies clades with high host breadth. The family Togaviridae, which contains the alphaviruses, was identified as having a significantly higher host breadth than other viruses.

Figure 4: Phylofactorization of viruses by host range suggests a surveillance prioritization scheme. Clades identified by host-range phylofactorization vary in their fraction of zoonotic representatives, and inclusion of clades into model selection of trait-based analysis produces updated predictions for zoonosis in clades while controlling for other traits. Traits classifying the host-range clades suggest different pathways to spillover, from the vector-borne transmission of *Togaviridae* to environmental persistence of excreted *Picornaviridae*. Additional traits were pulled for a subsample of viruses with a meta-analysis described in the methods. Total clade side indicates the size of the clade in the Olival et al. (2017) dataset, sample size (n) the number of viruses for which we could identify additional traits related to route of transmission and the pathway to spillover within each clade.

References

Ahn, K., T. H. Meyer, S. Uebel, P. Sempé, H. Djaballah, Y. Yang, P. A. Peterson, K. Früh, and R. Tampé. 1996. “Molecular Mechanism and Species Specificity of TAP Inhibition by Herpes Simplex Virus ICP47.” *The EMBO Journal* 15 (13):3247–55.

Ahn, Kwangseog, Albrecht Gruhler, Begona Galocha, Thomas R. Jones, Emmanuel J. H. J. Wiertz, Hidde L. Ploegh, Per A. Peterson, Young Yang, and Klaus Früh. 1997. “The ER-Luminal Domain of the HCMV Glycoprotein US6 Inhibits Peptide Translocation by TAP.” *Immunity* 6 (5):613–21.

Ambagala, Aruna P. N., Raju S. Gopinath, and S. Srikumaran. 2004. “Peptide Transport Activity of the Transporter Associated with Antigen Processing (TAP) Is Inhibited by an Early Protein of Equine Herpesvirus-1.” *The Journal of General Virology* 85 (Pt 2):349–53.

Baize, Sylvain, Delphine Pannetier, Lisa Oestereich, Toni Rieger, Lamine Koivogui, N ’faly Magassouba, Barrè Soropogui, et al. 2014. “Emergence of Zaire Ebola Virus Disease in Guinea.” *The New England Journal of Medicine* 371 (15):1418–25.

Bausch, Daniel G., and James N. Mills. 2014. “Arenaviruses: Lassa Fever, Lujo Hemorrhagic Fever, Lymphocytic Choriomeningitis, and the South American Hemorrhagic Fevers.” In *Viral Infections of Humans*, 147–71.

Black, William C., Kristine E. Bennett, Norma Gorrochótegui-Escalante, Carolina V. Barillas-Mury, Ildefonso Fernández-Salas, María de Lourdes Muñoz, José A. Farfán-Alé, Ken E. Olson, and Barry J. Beaty. 2002. “Flavivirus Susceptibility in Aedes Aegypti.” *Archives of Medical Research* 33 (4):379–88.

Byrnes, A. P., and D. E. Griffin. 1998. “Binding of Sindbis Virus to Cell Surface Heparan Sulfate.” *Journal of Virology* 72 (9):7349–56.

Chua, K. B., W. J. Bellini, P. A. Rota, B. H. Harcourt, A. Tamin, S. K. Lam, T. G. Ksiazek, et al. 2000. “Nipah Virus: A Recently Emergent Deadly Paramyxovirus.” *Science* 288 (5470):1432–35.

Duijl-Richter, Mareike K. S. van, Tabitha E. Hoornweg, Izabela A. Rodenhuis-Zybert, and Jolanda M. Smit. 2015. “Early Events in Chikungunya Virus Infection-From Virus Cell Binding to Membrane Fusion.” *Viruses* 7 (7):3647–74.

Emonet, Sébastien, Karine Retornaz, Jean-Paul Gonzalez, Xavier de Lamballerie, and Rémi N. Charrel. 2007. “Mouse-to-Human Transmission of Variant Lymphocytic Choriomeningitis Virus.” *Emerging Infectious Diseases* 13 (3):472–75.

Fouchier, R. A. M., P. M. Schneeberger, F. W. Rozendaal, J. M. Broekman, S. A. G. Kemink, V. Munster, T. Kuiken, et al. 2004. “Avian Influenza A Virus (H7N7) Associated with Human Conjunctivitis and a Fatal Case of Acute Respiratory Distress Syndrome.” *Proceedings of the National Academy of Sciences* 101 (5):1356–61.

Gardy, Jennifer L., and Nicholas J. Loman. 2017. “Towards a Genomics-Informed, Real-Time, Global Pathogen Surveillance System.” *Nature Reviews. Genetics*, November. https://doi.org/10.1038/nrg.2017.88.

Geoghegan, Jemma L., and Edward C. Holmes. 2017. “Predicting Virus Emergence amid Evolutionary Noise.” *Open Biology* 7 (10). https://doi.org/10.1098/rsob.170189.

Geoghegan, Jemma L., Alistair M. Senior, Francesca Di Giallonardo, and Edward C. Holmes. 2016. “Virological Factors That Increase the Transmissibility of Emerging Human Viruses.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (15):4170–75.

Gonzalez, J. P., S. Emonet, X. de Lamballerie, and R. Charrel. 2007. “Arenaviruses.” In *Current Topics in Microbiology and Immunology*, 253–88.

Groseth, Allison, Andrea Marzi, Thomas Hoenen, Astrid Herwig, Don Gardner, Stephan Becker, Hideki Ebihara, and Heinz Feldmann. 2012. “The Ebola Virus Glycoprotein Contributes to but Is Not Sufficient for Virulence in Vivo.” *PLoS Pathogens* 8 (8):e1002847.

480 Haddaway, Neal R., and Maggie J. Watson. 2016. “On the Benefits of Systematic Reviews for
481 Wildlife Parasitology.” *International Journal for Parasitology. Parasites and Wildlife* 5
482 (2):184–91.
483 Hayes, Edward B., and Duane J. Gubler. 2006. “West Nile Virus: Epidemiology and Clinical
484 Features of an Emerging Epidemic in the United States.” *Annual Review of Medicine*
485 57:181–94.
486 Heldstab, A., D. Rüedi, W. Sonnabend, and F. Deinhardt. 1981. “Spontaneous Generalized
487 Herpesvirus Hominis Infection of a Lowland Gorilla (Gorilla Gorilla Gorilla).” *Journal of*
488 *Medical Primatology* 10 (2-3):129–35.
489 Huff, Jennifer L., and Peter A. Barry. 2003. “B-Virus (Cercopithecine herpesvirus1) Infection in
490 Humans and Macaques: Potential for Zoonotic Disease.” *Emerging Infectious Diseases* 9
491 (2):246–50.
492 Jane Flint, S., Vincent R. Racaniello, Glenn F. Rall, Lynn W. Enquist, and Anna-Marie Skalka.
493 2015. *Principles of Virology: Volume 1: Molecular Biology*.
494 Jones, Kate E., Nikkita G. Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L.
495 Gittleman, and Peter Daszak. 2008. “Global Trends in Emerging Infectious Diseases.”
496 *Nature* 451 (7181):990–93.
497 Klimstra, W. B., E. M. Nangle, M. S. Smith, A. D. Yurochko, and K. D. Ryman. 2003. “DC-SIGN
498 and L-SIGN Can Act as Attachment Receptors for Alphaviruses and Distinguish between
499 Mosquito Cell- and Mammalian Cell-Derived Viruses.” *Journal of Virology* 77 (22):12022–
500 32.
501 Koppers-Lalic, D., M. Rychlowski, D. van Leeuwen, F. A. M. Rijsewijk, M. E. Ressing, J. J.
502 Neefjes, K. Bienkowska-Szewczyk, and E J H. 2003. “Bovine Herpesvirus 1 Interferes with
503 TAP-Dependent Peptide Transport and Intracellular Trafficking of MHC Class I Molecules in
504 Human Cells.” *Archives of Virology* 148 (10):2023–37.
505 Kosiol, Carolin, Tomáš Vinař, Rute R. da Fonseca, Melissa J. Hubisz, Carlos D. Bustamante,
506 Rasmus Nielsen, and Adam Siepel. 2008. “Patterns of Positive Selection in Six Mammalian
507 Genomes.” *PLoS Genetics* 4 (8):e1000144.
508 Li, Qun, Lei Zhou, Minghao Zhou, Zhiping Chen, Furong Li, Huanyu Wu, Nijuan Xiang, et al.
509 2014. “Epidemiology of Human Infections with Avian Influenza A(H7N9) Virus in China.”
510 *The New England Journal of Medicine* 370 (6):520–32.
511 Lloyd-Smith, J. O., D. George, K. M. Pepin, V. E. Pitzer, J. R. C. Pulliam, A. P. Dobson, P. J.
512 Hudson, and B. T. Grenfell. 2009. “Epidemic Dynamics at the Human-Animal Interface.”
513 *Science* 326 (5958):1362–67.
514 Messenger, Sharon L., Jean S. Smith, and Charles E. Rupprecht. 2002. “Emerging
515 Epidemiology of Bat-Associated Cryptic Cases of Rabies in Humans in the United States.”
516 *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of*
517 *America* 35 (6):738–47.
518 Meyer, B. J., J. C. De La Torre, and P. J. Southern. 2002. “Arenaviruses: Genomic RNAs,
519 Transcription, and Replication.” In *Current Topics in Microbiology and Immunology*, 139–
520 57.
521 Morrison, Amy C., Emily Zielinski-Gutierrez, Thomas W. Scott, and Ronald Rosenberg. 2008.
522 “Defining Challenges and Proposing Solutions for Control of the Virus Vector Aedes
523 Aegypti.” *PLoS Medicine* 5 (3):e68.
524 Murray, K., P. Selleck, P. Hooper, A. Hyatt, A. Gould, L. Gleeson, H. Westbury, L. Hiley, L.
525 Selvey, and B. Rodwell. 1995. “A Morbillivirus That Caused Fatal Disease in Horses and
526 Humans.” *Science* 268 (5207):94–97.
527 Murthy, S., E. Couacy-Hymann, S. Metzger, K. Nowak, H. De Nys, C. Boesch, R. Wittig, M. A.
528 Jarvis, F. H. Leendertz, and B. Ehlers. 2013. “Absence of Frequent Herpesvirus
529 Transmission in a Nonhuman Primate Predator-Prey System in the Wild.” *Journal of*
530 *Virology* 87 (19):10651–59.

Olival, Kevin J., Parviez R. Hosseini, Carlos Zambrana-Torrelio, Noam Ross, Tiffany L. Bogich, and Peter Daszak. 2017. “Host and Viral Traits Predict Zoonotic Spillover from Mammals.” *Nature* 546 (7660):646–50.

Paradis, E., J. Claude, and K. Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2):289–90.

Petersen, L. R., and J. T. Roehrig. 2001. “West Nile Virus: A Reemerging Global Pathogen.” *Emerging Infectious Diseases* 7 (4):611–14.

Plourde, Benjamin T., Tristan L. Burgess, Evan A. Eskew, Tara M. Roth, Nicole Stephenson, and Janet E. Foley. 2017. “Are Disease Reservoirs Special? Taxonomic and Life History Characteristics.” *PloS One* 12 (7):e0180716.

Plowright, Raina K., Peggy Eby, Peter J. Hudson, Ina L. Smith, David Westcott, Wayne L. Bryden, Deborah Middleton, et al. 2015. “Ecological Dynamics of Emerging Bat Virus Spillover.” *Proceedings. Biological Sciences / The Royal Society* 282 (1798):20142124.

Plowright, Raina K., Colin R. Parrish, Hamish McCallum, Peter J. Hudson, Albert I. Ko, Andrea L. Graham, and James O. Lloyd-Smith. 2017. “Pathways to Zoonotic Spillover.” *Nature Reviews. Microbiology* 15 (8):502–10.

Plowright, Raina K., Alison J. Peel, Daniel G. Streicker, Amy T. Gilbert, Hamish McCallum, James Wood, Michelle L. Baker, and Olivier Restif. 2016. “Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir–Host Populations.” *PLoS Neglected Tropical Diseases* 10 (8):e0004796.

Powers, Ann M., and Christopher H. Logue. 2007. “Changing Patterns of Chikungunya Virus: Re-Emergence of a Zoonotic Arbovirus.” *The Journal of General Virology* 88 (Pt 9):2363–77.

Pulliam, Juliet R. C., Jonathan H. Epstein, Jonathan Dushoff, Sohayati A. Rahman, Michel Bunning, Aziz A. Jamaluddin, Alex D. Hyatt, et al. 2012. “Agricultural Intensification, Priming for Persistence and the Emergence of Nipah Virus: A Lethal Bat-Borne Zoonosis.” *Journal of the Royal Society, Interface / the Royal Society* 9 (66):89–101.

Roizman, Bernard. 1982. “The Family Herpesviridae: General Description, Taxonomy, and Classification.” In *The Herpesviruses*, 1–23.

Rosenberg, Ronald. 2015. “Detecting the Emergence of Novel, Zoonotic Viruses Pathogenic to Humans.” *Cellular and Molecular Life Sciences: CMLS* 72 (6):1115–25.

Sagripanti, Jose-Luis, and C. David Lytle. 2011. “Sensitivity to Ultraviolet Radiation of Lassa, Vaccinia, and Ebola Viruses Dried on Surfaces.” *Archives of Virology* 156 (3):489–94.

Schneider, Maria Cristina, Phyllis Catharina Romijn, Wilson Uieda, Hugo Tamayo, Daniela Fernandes da Silva, Albino Belotto, Jarbas Barbosa da Silva, and Luis Fernando Leanes. 2009. “Rabies Transmitted by Vampire Bats to Humans: An Emerging Zoonotic Disease in Latin America?” *Revista Panamericana de Salud Publica = Pan American Journal of Public Health* 25 (3):260–69.

Stephenson, E. H., E. W. Larson, and J. W. Dominik. 1984. “Effect of Environmental Factors on Aerosol-Induced Lassa Virus Infection.” *Journal of Medical Virology* 14 (4):295–303.

Voss, James E., Marie-Christine Vaney, Stéphane Duquerroy, Clemens Vonrhein, Christine Girard-Blanc, Elodie Crublet, Andrew Thompson, Gérard Bricogne, and Félix A. Rey. 2010. “Glycoprotein Organization of Chikungunya Virus Particles Revealed by X-Ray Crystallography.” *Nature* 468 (7324):709–12.

Wang, K. S., R. J. Kuhn, E. G. Strauss, S. Ou, and J. H. Strauss. 1992. “High-Affinity Laminin Receptor Is a Receptor for Sindbis Virus in Mammalian Cells.” *Journal of Virology* 66 (8):4992–5001.

Washburne, Alex. 2017. “Phylofactorization - Theory and Challenges.” https://doi.org/10.1101/196378.

Washburne, Alex D., Justin D. Silverman, Jonathan W. Leff, Dominic J. Bennett, John L. Darcy, Sayan Mukherjee, Noah Fierer, and Lawrence A. David. 2017. “Phylogenetic Factorization

of Compositional Data Yields Lineage-Level Associations in Microbiome Datasets." *PeerJ* 5. PeerJ Inc.:e2969.

Washburne, Alex D., Justin D. Silverman, Jamie T. Morton, Daniel Becker, Daniel Crowley, Sayan Mukherjee, Lawrence A. David, and Raina K. Plowright. 2017. "Phylofactorization - a Graph Partitioning Algorithm to Identify Phylogenetic Scales of Ecological Data." https://doi.org/10.1101/235341.

Wikan, Nitwara, and Duncan R. Smith. 2016. "Zika Virus: History of a Newly Emerging Arbovirus." *The Lancet Infectious Diseases* 16 (7):e119–26.

Woźniakowski, Grzegorz, and Elżbieta Samorek-Salamonowicz. 2015. "Animal Herpesviruses and Their Zoonotic Potential for Cross-Species Infection." *Annals of Agricultural and Environmental Medicine: AAEM* 22 (2):191–94.

Xia, Hongjie, Huanle Luo, Chao Shan, Antonio E. Muruato, Bruno T. D. Nunes, Daniele B. A. Medeiros, Jing Zou, et al. 2018. "An Evolutionary NS1 Mutation Enhances Zika Virus Evasion of Host Interferon Induction." *Nature Communications* 9 (1):414.

Figure 1

Cladistic analyses complement and refine trait-based analyses of viral zoonotic potential

Understanding which clades of viruses spillover can improve our understanding of life-history and trait-based predictors of zoonosis. A thought experiment of predicting flight in a subset of 10 vertebrates, including reptiles, birds and mammals reveals the importance of cladistic analyses. Trait-based analyses aim to identify explanatory traits correlated with a focal response trait, such as identifying feathers and hollow bones as associated with flight in the species observed. We propose a method outlined by the arrows. First, cladistic analysis can identify clades with a common pattern in the response variable. The clade highlighted in green - birds - has more representatives who fly than others. Then, cladistic analyses can allow life-history classification of clades to generate hypotheses about which traits (e.g. wings) and even clade labels (e.g. birds) should be included for amend trait-based analyses. Some traits, such as feathers, hollow bones, wings, and living in trees are dependent upon one-another and a consequence of life histories, and life-histories can reveal the more nuanced sets of traits, including clade-specific traits, which may underlie pathogen spillover. In this paper, for example, we find *Togaviridae* to be at risk for spilling over and discuss how previously unconsidered traits, such as variants of the E2 protein shared among *Togaviridae*, may be clade-specific determinants of spillover.

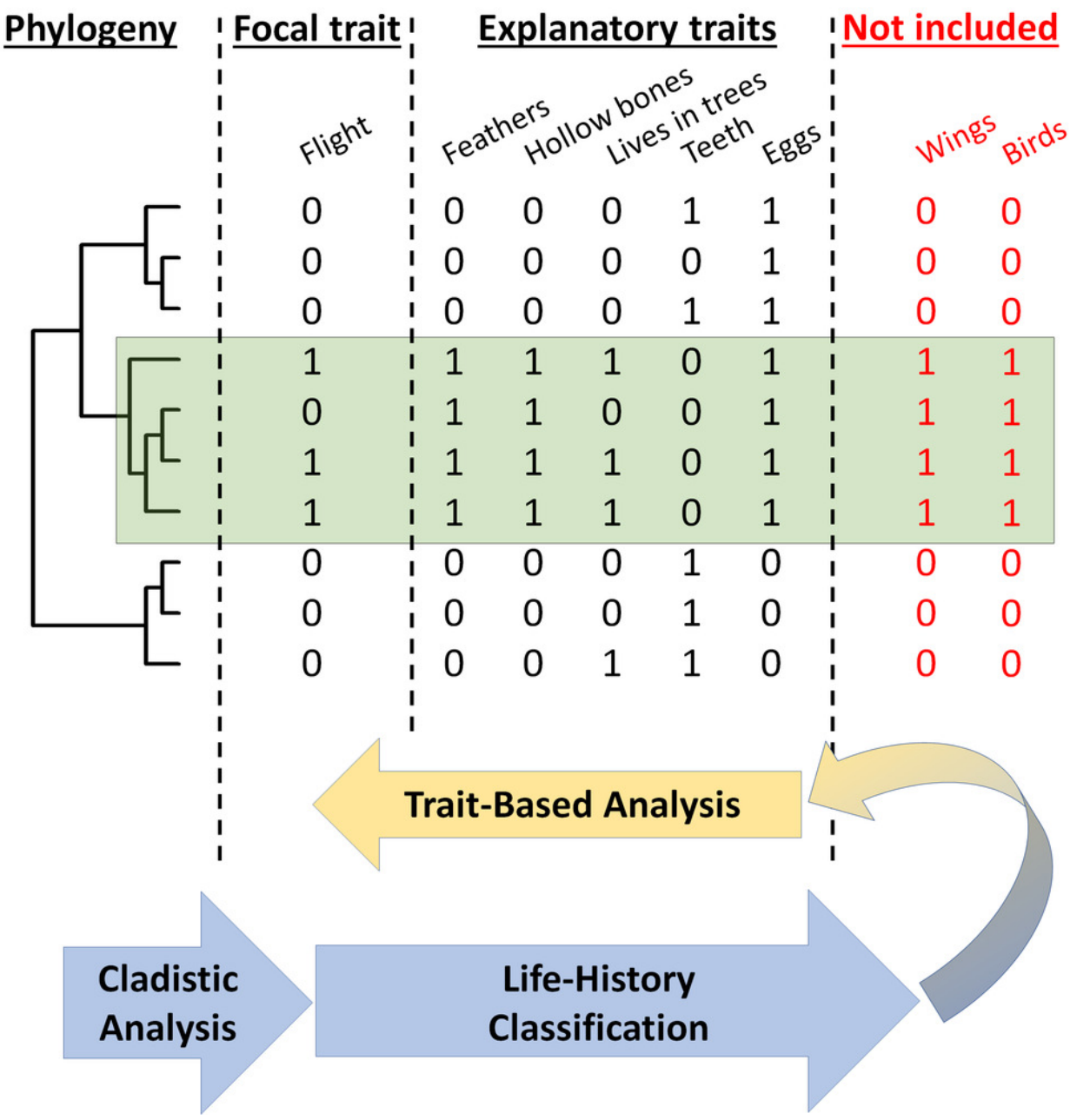
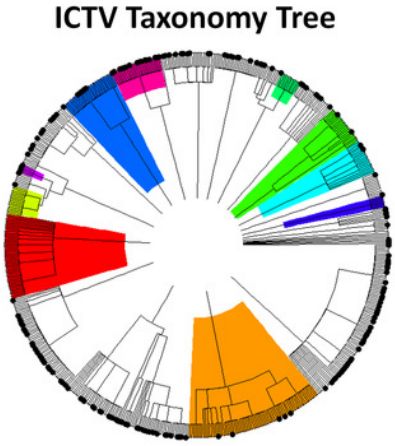

Phylogeny
Focal trait
Explanatory traits
Not included
Flight
Feathers
Hollow bones
Lives in trees
Teeth
Eggs
Wings
Birds
0 0 0 0 1 1 0 0
0 0 0 0 0 1 0 0
0 0 0 0 1 1 0 0
1 1 1 1 0 1 1 1
0 1 1 0 0 1 1 1
1 1 1 1 0 1 1 1
1 1 1 1 0 1 1 1
0 0 0 0 1 0 0 0
0 0 0 0 1 0 0 0
0 0 0 1 1 0 0 0
Trait-Based Analysis
Cladistic Analysis
Life-History Classification

Figure 2

Phylogenetic factors of zoonosis in mammalian viruses

Phylogenetic factorization iteratively partitions a phylogeny along edges separating species with meaningful differences. Phylofactorization of the ICTV taxonomy by a Fisher test on the fraction of zoonotic viruses identified clades with different numbers of species (n) and different rates of zoonosis (% Zoonotic). 32% of the viruses in the original dataset of mammalian viruses are zoonotic, indicated as black dots on the tip of the tree. The first 8 phylogenetic factors identify clades whose distinction is a significantly low rate of zoonosis relative to other viruses. Deltaretroviruses and Alphaviruses are then identified as having an unusually high fraction of zoonotic viruses. Cladistic structure in zoonotic potential can be used to identify previously ignored traits and prioritize surveillance programs.



Factor #	Taxonomic Level	Name	n	% Zoonotic
1	Family	Papillomaviridae	42	0
2	Family	Herpesviridae	68	7.4
3	Subfamily	Orthoretrovirinae	14	0
4	Order	Nidovirales	20	10
5	Genus	Parvovirus	10	0
6	Family	Caliciviridae	17	11.8
7	Family	Adenoviridae	28	21.4
8	Family	Astroviridae	6	0
9	Genus	Deltaretrovirus	4	100
10	Genus	Alphavirus	25	64
Remainder	Paraphyletic	NA	352	43.5

Figure 3

Phylogenetic factors of non-human mammalian host breadth

(A) Non-human phylogenetic distance on a cytochrome B phylogeny is a reliable predictor of zoonotic viruses, even upon removal of non-zoonotic clades identified in figure 2. Zoonotic prediction is the linear predictor from generalized additive modelling, plotted here against phylogenetic breadth. (B) Phylofactorization of viruses by mammalian host breadth identifies clades with high host breadth. The family Togaviridae, which contains the alphaviruses, was identified as having a significantly higher host breadth than other viruses.

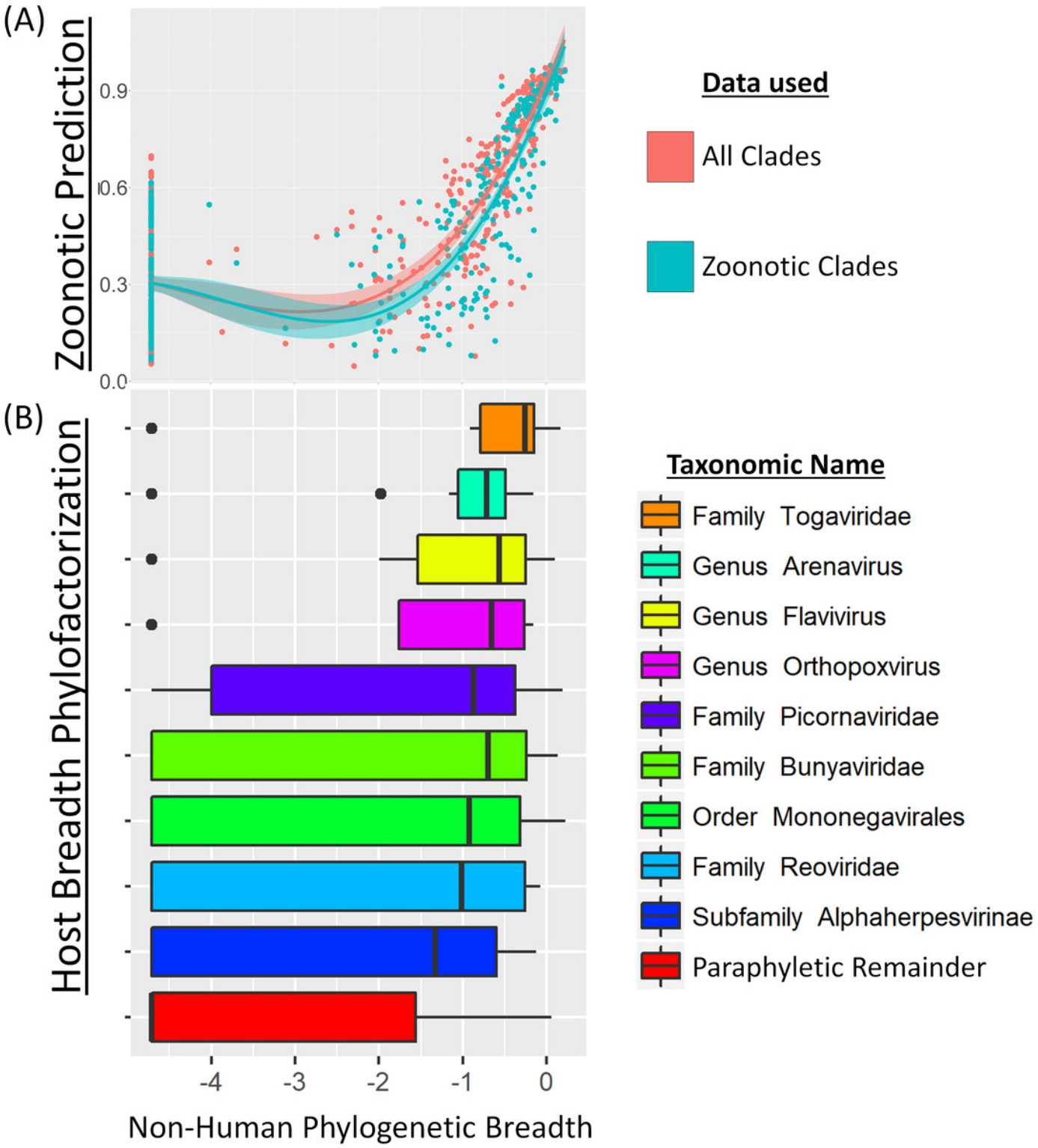
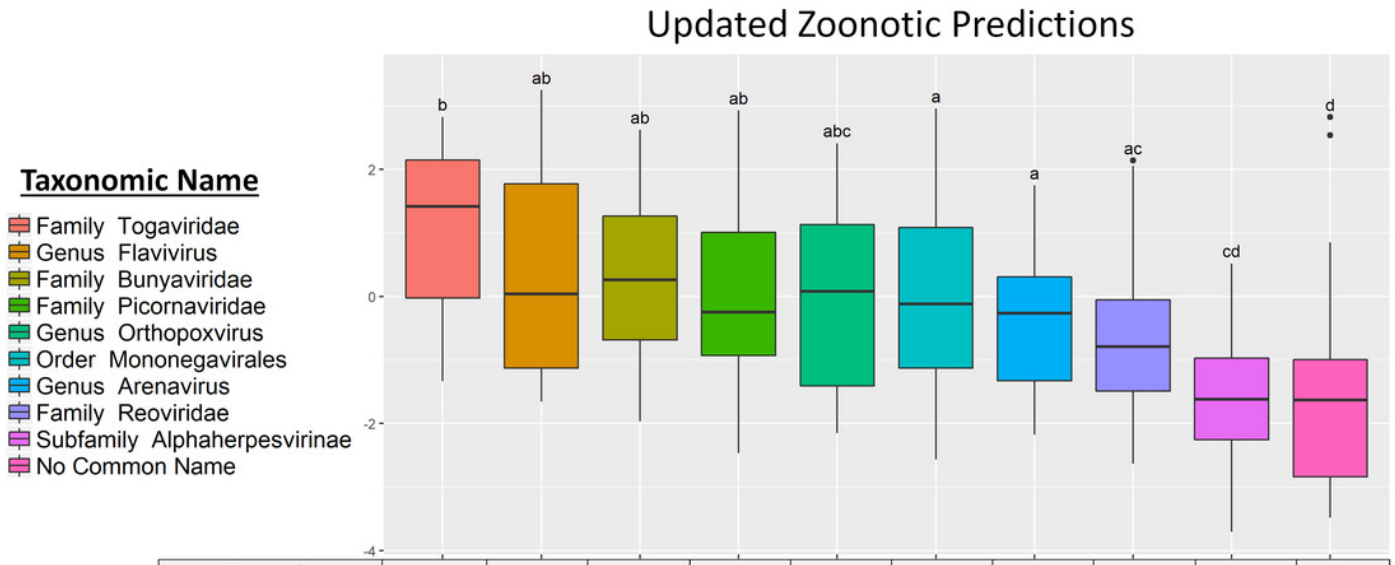

(A)
Zoonotic Prediction
0.0
0.3
0.6
0.9
(B)
Host Breadth Phylofactorization
-4
-3
-2
-1
0
Non-Human Phylogenetic Breadth
Data used
All Clades
Zoonotic Clades
Taxonomic Name
Family Togaviridae
Genus Arenavirus
Genus Flavivirus
Genus Orthopoxvirus
Family Picornaviridae
Family Bunyaviridae
Order Mononegavirales
Family Reoviridae
Subfamily Alphaherpesvirinae
Paraphyletic Remainder

Figure 4

Phylofactorization of viruses by host range suggests a taxonomic-based surveillance prioritization scheme.

Clades identified by host-range phylofactorization vary in their fraction of zoonotic representatives, and inclusion of clades into model selection of trait-based analysis produces updated predictions for zoonosis in clades while controlling for other traits. Traits classifying the host-range clades suggest different pathways to spillover, from the vector-borne transmission of Togaviridae to environmental persistence of excreted Picornaviridae.



Zoonotic transmission traits	Toga.	Flav.	Buny.	Pico.	Orth.	Mon.	Aren.	Reo.	Alpha.	Rest
Total clade size	26	42	83	20	9	66	23	29	27	261
Sample size (*n*)	16	23	42	7	4	32	9	11	3	41
Enveloped (%)	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	0.59
Vector-borne (%)	1.00	0.96	0.67	0.00	0.00	0.22	0.11	0.73	0.00	0.12
• Unknown (%)	0.00	0.00	0.02	0.14	0.00	0.19	0.00	0.00	0.00	0.07
Excretion (%)	0.06	0.13	0.38	1.00	1.00	0.62	1.00	0.27	1.00	0.85
• Unknown (%)	0.00	0.04	0.12	0.00	0.00	0.22	0.00	0.09	0.00	0.10
Slaughter (%)	0.00	0.00	0.05	0.29	0.25	0.16	0.00	0.00	0.00	0.49
• Unknown (%)	0.00	0.04	0.14	0.29	0.00	0.25	0.00	0.09	0.00	0.10
Human-human (%)	0.19	0.13	0.07	0.71	0.50	0.34	0.22	0.18	0.67	0.56
• Unknown (%)	0.19	0.22	0.33	0.14	0.00	0.19	0.22	0.27	0.33	0.17

Table 1(on next page)

Effect sizes and significance of traits pre- andpost-cladistic analysis

Traits such as enveloped, vector-borne, and cytoplasmic replication are sensitive due to their cladistic signal. Here, effect sizes (difference in linear predictors) are presented for the all viruses in the original trait-based analysis, the zoonotic clades (removal of non-zoonotic clades found in Figure 2), and non-zoonotic clades. The direction of the effect of all three traits is conserved, but the magnitude and relative significance changes, as these traits are strongly associated with non-zoonotic clades. Asterisks indicate significance levels $P<0.05$ (*), $P<0.01$ (**), $P<0.001$ (***) from Chi-squared tests of the deviance from generalized additive models.

1

	All Clades	Zoonotic Clades	Non-Zoonotic Clades
Vector	1.4*	0.62	-0.16**
Envelope	0.87	1.2*	-0.36**
Cytoplasmic Replication	1.8**	0.88	-0.38***