HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2020 January 29.

Published in final edited form as:

Nat Microbiol. 2019 November ; 4(11): 1778–1780. doi:10.1038/s41564-019-0523-5.

Virus Structures Constrain Transmission Modes

Frederic D. Bushman[1], **Kevin McCormick**[1], **Scott Sherrill-Mix**[1]

[1]University of Pennsylvania Perelman School of Medicine, Department of Microbiology, 3610 Hamilton Walk, Philadelphia, PA 19104-6076

Abstract

This study investigates links between virus structures and routes of transmission. Viruses show a wide range of structures, and transmission of viruses between vertebrate hosts can take place by many routes. We compiled a database of 243 virus/host combinations, and report a statistical analysis documenting associations between structures and routes of transmission—for example, viruses that are transmitted by the fecal-oral route are rarely enclosed in a lipid envelope.

Main text

To develop a data set to compare virus structures and transmission routes, we focused on viruses causing diseases in humans and domestic animals, for which the most complete information is available (Supplementary Table 1). A database of 104 human viruses was compiled from standard sources and validated with extensive searches of PubMed and other resources[1–6]. To query a wide range of vertebrates, while focusing on animals with closely studied viruses, we also analyzed viruses of dogs (n=16), cats (n=14), cows (n=36), horses (n=27), pigs (n=31), and chickens (n=17)[4,5].

To allow statistical analysis, virus structures and modes of transmission were grouped into categories. Virus structures were described as 1) enveloped (enclosed in a lipid membrane) or not, 2) comprised of capsids that were spherical, bullet-form, filamentous, or pleomorphic, and 3) whether the genome was comprised of RNA or DNA[1–5].

All viruses were annotated with their modes of transmission. To allow statistical analysis, transmission routes were summarized as 12 categories: 1) fecal-oral, 2) arthropod-borne (arboviruses), 3) inhalation of aerosols (liquid droplets), 4) inhalation of dried material (dust

Correspondence: Correspondence can be directed to bushman@mail.med.upenn.edu.

Author contributions

FDB conceived the work, assembled the initial database, and wrote the paper. SS-M and KM developed the electronic versions of the data, helped reference the database, and carried out statistical analyses. SS-M developed the statistical analytical strategy. KM developed the automated reference search strategy.

Data Availability

All data used is presented in tabular form in the Supplementary Material.

Code Availability

Code used is available at https://doi.org/10.5281/zenodo.3246601

Competing interests

The authors declare that they have no competing interests.

or feces), 5) sexual, 6) eating (without passage through the full gastrointestinal tract), 7) oral to bloodstream (typically biting), 8) breastfeeding, 9) vertical from mother to offspring, 10) vertical via germ line integration, 11) via blood products (direct introduction into the bloodstream via blood products, organ transplantation, needle sharing, and cuts), and 12) contact of skin or eye with bodily fluids or fomites. Figure 1 shows the data for human viruses; heat maps for the other vertebrates are in Supplementary Figures 1–6. Some of these categories are broad and somewhat heterogeneous, requiring judgment calls in borderline cases. For example, several non-enveloped viruses may spread within an organism in a "quasi-enveloped" form, and possibly even between organisms, though usually with lower efficiency[7]. Here we only considered the non-enveloped forms. In some cases infection may have different efficiency via different routes—we did not attempt to distinguish inefficient and efficient transmission mechanisms.

A complication in the analysis is that measurements on individual viruses are not fully independent. Viruses are more likely to share characteristics with near neighbors on a phylogenetic tree than more distant neighbors. For this reason, we obtained a phylogenetic tree from the International Committee on Taxonomy of Viruses (ICTV) and carried out statistical analyses that respected the phylogenetic framework.

We used phylogenetic logistic regression[8] to assess associations between transmission mode and structural traits while accounting for phylogenetic correlation (Fig. 2). Significant associations between structural features and transmission modes are shown in a heatmap format, with the direction of the effects indicated by the color code. A disadvantage of this approach is that some candidate associations could not be queried when quasi-complete separation prevented model convergence (Fig. 2, grey squares). We thus performed additional pairwise comparisons using Pagel's correlation method[9] to assess correlated evolution in the viral phylogeny, and found that none of the associations excluded from the logistic regression for quasi-complete separation were statistically significant by Pagel's method.

We found that fecal-oral transmission was strongly associated with absence of a lipid envelope (Fig. 2). Of the 16 viruses known to be transmitted by the fecal-oral route in humans analyzed here (Fig. 1), none have lipid envelopes. This was strongly though not universally supported in the animal virus data as well (Supplementary Table 1; Supplementary Figures 1–6). These include viruses belonging to the families *Picornaviridae*, *Caliciviridae*, *Hepeviridae*, *Reoviridae*, *Parvoviridae*, *Polyomaviridae*, and *Adenoviridae*. The lack of envelope in fecal-oral transmitted viruses is readily rationalized by the known fragility of lipid bilayers. A lipid membrane is expected to be disrupted by drying, and material in the lumen of the human gut is dehydrated in the colon. In the upper GI tract, stomach acid provides another chemical barrier. Lipid bilayers are also disrupted by detergents, and bile acids serve as detergents in gut. In addition, in order to survive after shedding from gut into the environment, a viral particle must survive in its new chemical environment as well. Thus the strict lack of lipid membranes is readily understood as a requirement for survival in diverse chemically harsh environments.

Another strong pattern in the data indicates that viruses transmitted by casual contact, which includes contact with bodily fluids via the skin and eye, commonly have DNA genomes (Fig. 2). While both DNA and RNA viruses are included in this transmission mode, the DNA viruses are statistically enriched. DNA is a more stable polymer than RNA, possibly explaining more efficient infection by DNA viruses via a route potentially involving greater exposure to environmental stresses. Possibly RNA genomes are more readily detected by the host than DNA genomes at some site visited by viruses during this form of transmission. Viruses involved include most of the DNA viruses—*Adenoviridae*, *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, and *Poxviridae*.

Similarly, DNA viruses are also favored for sexual transmission (Fig. 2). These include *Herpesviridae*, *Papillomaviridae*, and *Poxviridae*. In this case as well, the relative stability of DNA compared to RNA may help support transmission through a chemically harsh environment, or the DNA genome may more easily dodge innate immune detection.

Further associations are seen in the data in Figure 2, offering additional opportunities for exploration.

These data may find several practical applications. A deeper understanding of the structure-transmission relationships for pathogenic viruses may help highlight vulnerabilities where intervention could break transmission chains. Conversely, these findings may be useful where viruses are repurposed as platforms for synthetic biology. Examples include use as vectors for gene delivery in human gene therapy, as antibacterial agents in phage therapy, and as highly polyvalent nanostructures for tissue staining and drug delivery. A careful matching of viral structural classes to the intended application based on information summarized here may help increase the chances of producing useful designs.

Methods

The database used in this analysis (Supplementary Table 1) was constructed in several steps. The list of viral select agents infecting humans was obtained from NIAID online resources (https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens). Further data on virus types, structures and transmission modes was taken from the Center for Disease Control (www.cdc.gov/), Merck Veterinary Manual (www.merckvetmanual.com), ViralZone[4], VIPR[6], and other standard resources[2,5]. Data on virus taxonomy was obtained from the ICTV. This yielded an initial database of 243 virus-host combinations, each characterized by six structural features and 12 transmission modes. Following this first pass analysis, automated searches were devised to capture references documenting each transmission mode, favoring where ever possible peer-reviewed literature. PubMed was searched with the name of each host vertebrate, the name of each virus, “transmission”, and the name of one of the 12 transmission modes (for example, one search included the words “human”, “human immunodeficiency virus”, “transmission”, and “sexual”). In some cases, searches were further refined to obtain optimal documentation. Search output was evaluated for each of the 2916 entries, yielding 463 TRUE associations of viruses with transmission modes, and 2453 false (i. e. negative in PubMed, denoted PubMedNeg in Supplementary Table 1). A total of 486 supporting citations were recorded, of which 379 were PubMed

entries, and the rest non-peer reviewed sources. Some edge cases were discussed and resolved by two or more virologists. “Oral/bloodstream” infection was used to describe transmission by vertebrate biting; transmission via biting insects was denoted “arbovirus”. “Sexual” encompassed homosexual and heterosexual transmission routes.

Features of virus structure were compared to modes of transmission using phylogenetic logistic regression[8] carried out using the R function phyloglm from the R package phylolm[10].

A complication is the fact that viral phylogeny is not fully resolved. High-level relationships are uncertain—for example, is there any relationship between a dsDNA virus and an ssRNA virus? Thus in our tree all viral families were connected by long branches to a hypothetical (and unlikely) common ancestor, so that the uncertainty in high level classifications would have minimal effects on the outcome. The evolutionary distance/time between taxa is also uncertain for many viruses, so all other branches in our global viral phylogeny were set to an equal length. In addition, some of our viruses could infect more than one of the vertebrate hosts studied. To account for this, we made “host” a low-level branch off “species”, so that host was included in the analysis but not heavily weighted. We recognize that this phylogenetic tree is implausible in some respects, but it is a useful tool in controlling for the influence of phylogeny in the analysis.

Pagel’s model for correlated evolution[9] was evaluated using the function fitPagel in the R package phytools[11] and the pairwise p-values corrected for multiple comparisons by the Benjamini–Hochberg procedure[12]. Only transmission modes that were represented by at least 10 viral species-host combinations were considered in the analysis. Analyses were carried out in R v3.4.4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Arwa Abbas, Paul Bates, and Louis Taylor for suggestions and comments on the manuscript; and Laurie Zimmerman for help with the manuscript. This work was supported by NIH grants R61-HL137063 and R01-HL113252, the Penn Center for AIDS Research (P30 AI 045008), and the PennCHOP Microbiome Program.

References

1. Klug A & Caspar DL The structure of small viruses. Adv Virus Res 7, 225–325 (1960). [PubMed: 13756988]
2. Fields B Fields Virology, e. 6th edn, (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2013).
3. Branden CI John a. T.. Introduction to Protein Structure. 2nd edn, (Garland Science, 1999).
4. Hulo C et al. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res 39, D576–582, doi:10.1093/nar/gkq901 (2011). [PubMed: 20947564]
5. MacLachlan NJ & Dubovi EJ Fenner’s Veterinary Virology. (Elsevier, 2017).
6. Pickett BE et al. ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res 40, D593–598, doi:10.1093/nar/gkr859 (2012). [PubMed: 22006842]

7. Feng Z, Hirai-Yuki A, McKnight KL & Lemon SM Naked Viruses That Aren't Always Naked: Quasi-Enveloped Agents of Acute Hepatitis. Annu Rev Virol 1, 539–560, doi:10.1146/annurev-virology-031413-085359 (2014). [PubMed: 26958733]

8. Ives AR & Garland T Jr. Phylogenetic logistic regression for binary dependent variables. Syst Biol 59, 9–26, doi:10.1093/sysbio/syp074 (2010). [PubMed: 20525617]

9. Pagel M Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B 225, 37–45 (1994).

10. Ho L & Ane C A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst Biol 63, 397–408, doi:10.1093/sysbio/syu005 (2014). [PubMed: 24500037]

11. Revell LJ phylotools: and R package for phylogenetic comparative biology (and other things). Methods in Eco. and Evol. 3, 217–223 (2012).

12. Banjamini Y & Hochberg Y Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J. R. Statist. Soc. B 57, 289–300 (1995).

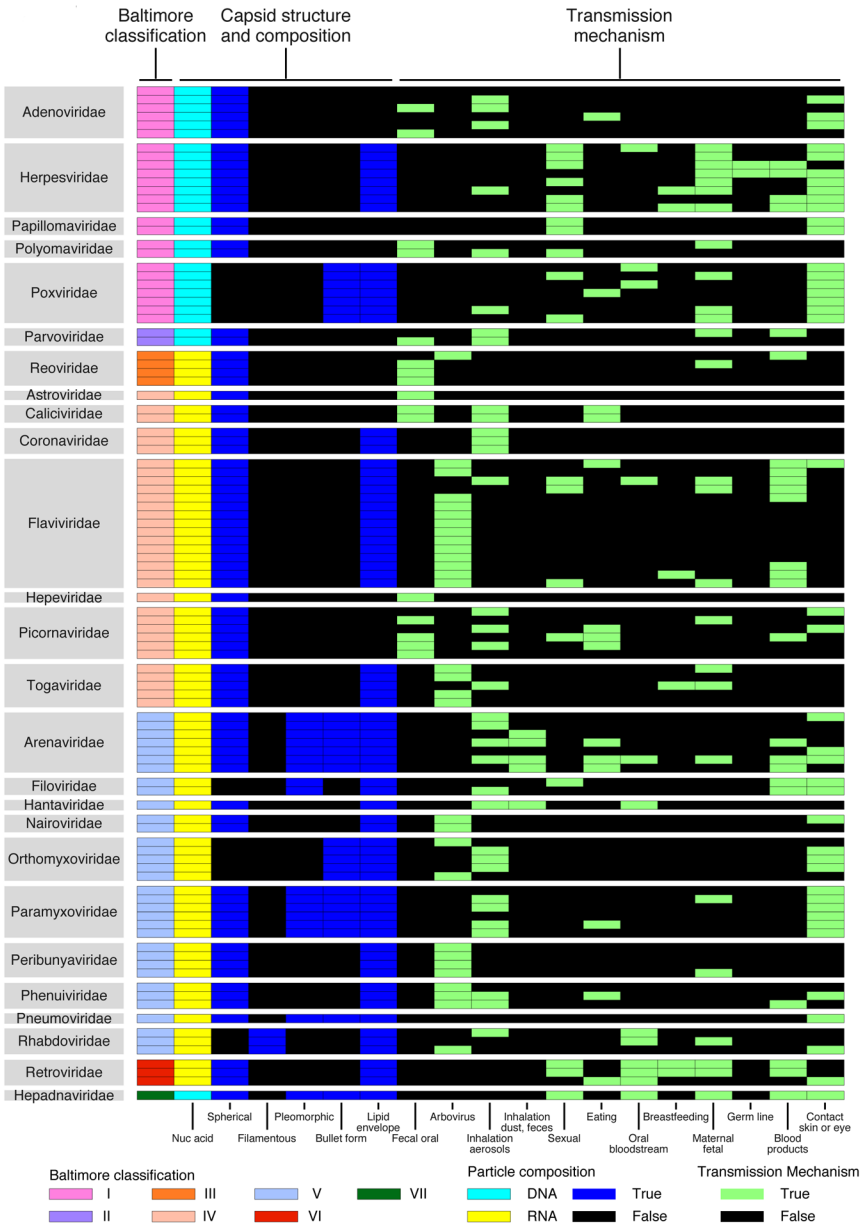

Figure 1.

Summary of structures of human viruses and their modes of transmission. Rows summarize information on individual human viral species. Columns summarize information on classification, capsid structure, particle composition, and transmission mechanism. Colored rectangles indicate positive evidence for the feature queried, black rectangles indicate lack of evidence. The color code is at the bottom. A full summary of all data on human and animal viruses is in Supplementary Table 1.

Figure 2.

Associations between virus structures and modes of transmission. Analysis was carried out over pooled human and animal viruses. Phylogenetic logistic regression was used for analysis. Colors indicate strength and direction of association where each row represents a phylogenetic logistic regression of the given transmission mode on structural traits. Each cell represents the bound of the 95% confidence interval closest to 1 (or 1 when the interval overlaps), where a fold change of 1 represents no association, for the fold change in odds of the transmission mode for viruses containing the specified structure. Thus all colored squares show associations achieving $p<0.05$. Grey indicates traits excluded due to quasi-complete separation preventing model convergence; checks using Pagel's method showed that none of these showed significant associations.