1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Delving Below the Species Level
To Characterize the Ecological Diversity within the Global Virome:
An Exploration of West Nile Virus

Tong Kong^{1*}, Kelly Mei^{2*}, Ammie Wang¹, Danny Krizanc¹, and Frederick M. Cohan²

Departments of Mathematics & Computer Science¹ and Biology², Wesleyan University,
Middletown, Connecticut, USA

*These authors contributed equally to this work.

Corresponding author: Frederick M. Cohan
Physical address: Department of Biology, Wesleyan University, Middletown, CT 06459, USA
Email: fcohan@wesleyan.edu
Phone: 860-810-5486

Running title: Host-association diversity within West Nile Virus

Keywords: Global Virome Project; host associations; ecotypes; genome sequence; microdiversity

Summary

Efforts to describe the diversity of viruses have largely focused on classifying viruses at the species level. However, substantial ecological diversity, both in virulence level and host range, is known within virus species. Here we demonstrate a proof of concept for easily discovering ecological diversity within a virus species taxon. We have focused on the West Nile Virus to take advantage of its broad host range in nature. We produced a genome-based phylogeny of world diversity of WNV and then used Ecotype Simulation 2 to hypothesize demarcation of genomes into 69 putative ecotypes (ecologically distinct populations), based only on clustering of genome sequences. Then we looked for evidence of ecological divergence among ecotypes based on differences in host bird associations within the Connecticut-New York region. Our results indicated significant heterogeneity among ecotypes for their associations with different bird hosts. Ecological diversity within other zoonotic viruses could be easily discovered using this approach. Opportunities for extending this line of research to human associations of virus ecotypes are limited by missing geographic metadata on human samples.

Introduction

Recent virus spillovers have revealed our species' vulnerability to infection from wild animals. Our domestic animals have recently conveyed several deadly bat viruses to humanity, including MERS (Drosten et al. 2014), Hendra (Plowright et al. 2011), and Nipah (Daszak et al. 2013); the bushmeat industry has brought us SARS (Yip et al. 2009); and human encroachment on wild lands has recurrently brought us Ebola (Rulli et al. 2017). Moreover, these infamous examples are not the whole story—on average, more than two new viruses are discovered in humans each year (Woolhouse et al. 2008). The undiscovered viruses residing in animal reservoirs are estimated to number 1.67 million, and about half of these are expected to have the capacity for infecting humans (Carroll et al. 2018).

Until recently, discovering the viruses that infect humanity from other animals has been a catch-up game. Starting when people get sick from a novel infection, virologists work urgently to characterize the virus, develop model systems, and produce vaccines and treatments. The Emerging Pandemic Threats project of USAID and the Global Virome Project have aimed to go beyond catch-up to *anticipate* future viral spillovers (Morse et al. 2012, Carroll et al. 2018). The Emerging Pandemic Threats project has aimed to identify the geographical regions, animal reservoirs, and human activities most likely to yield a future pandemic, and the GVP aims to discover, characterize, and sequence the genome of every virus on the planet that could possibly infect humans, all before the viruses can infect humanity. The GVP's proposers have argued that if we limit our focus to the virus families most likely to infect humans, as well as the host taxa with the greatest frequency of yielding spillovers to humans, the project's costs will in the end be much cheaper than paying for the damage from a pandemic (Carroll et al. 2018).

In its prospectus plan, the Global Virome Project focuses on finding new viruses that have important biological differences from already known species (Carroll et al. 2018). The GVP does not explicitly state the desired taxon rank at which new viruses should be sought, but it will likely aim to find viruses that are unique in their transmission, pathogenicity, cell and tissue tropism, host range, virion morphology, and antigenic relatedness—that is, viruses that meet the criteria of being distinct species taxa (Simmonds et al. 2017).

Virologists are increasingly discovering new virus species from molecular data, where a given level of sequence divergence suggests that an unknown virus is sufficiently distinct in its biological properties from other described species. For any given group of viruses, virologists have estimated a level of sequence divergence that correlates with significant differences in the properties of the viruses (Simmonds and Aiewsakun 2018). Recently, bioinformatic approaches have created algorithms to demarcate biologically significant virus groups on the basis of forming distinct sequence clusters (Lauber and Gorbalenya 2012, Yu et al. 2013, Bao et al. 2014). Sequence-based approaches hold promise for providing benchmark levels of sequence divergence that will allow discovery of biologically unique groups within a given virus group. The Global Virome Project may take advantage of these benchmarks in deciding whether an unknown virus is of sufficient interest.

Nevertheless, we cannot yet be certain that taxonomy of viruses at the species level is equipped to identify all the important ecological diversity among viruses. This is because extremely closely related viruses assigned to the same species taxon may be importantly different in their ecological features. Bacteriologists have long appreciated that a species taxon may contain important ecological diversity (Konstantinidis and Tiedje 2005, Denef et al. 2010,

Shapiro et al. 2012, Cohan 2017), and significant diversification within species is becoming evident in the world of virology as well (Marston and Amrich 2009).

This is readily apparent in the case of bacteriophages. A recent study of microdiversity of marine phages revealed that extremely close relatives differed in the depths and bacterial hosts they were adapted to (Warwick-Dugdale et al. 2019). Likewise, animal viruses within a single virus species taxon have diversified to infect different hosts. For example, lineages within the Rabies Virus species taxon have diverged in the bat species they infect (Streicker et al. 2010). Moreover, Rabies lineages associated with different bats have shown a history of divergent selection in proteins involved in host cell interaction (Streicker et al. 2012).

Viruses within a single species have also diverged in their levels of virulence toward their animal hosts. For example, among subgroups of HIV-1 Group M, the C subgroup reproduces and causes disease in humans much less aggressively than other subgroups (Ariën et al. 2007, Blanquart et al. 2016). Within West Nile Virus, the focus of the present paper, major clades have been shown to differ in the viremia level caused in their bird hosts (Duggal et al. 2014), as well as in transmission from their mosquito vectors (Ebel et al. 2004, Hadfield et al. 2019). If viruses within a species taxon typically differ in either host specificities or virulence, then the search for new viruses may benefit from efforts to delve below the species level.

Here we present a test case for routine discovery of potentially novel pathogens within a virus species. We have focused on West Nile Virus, a zoonotic virus that recurrently spills over to humans from birds by way of mosquitoes. WNV is a broadly generalist virus—it infects hundreds of species of birds and is transmitted by scores of mosquito species (Colpitts et al. 2012, Reisen 2013). Our study takes advantage of this virus’s ecological generalism.

We have investigated whether phylogenetic clades within Cluster 4 of West Nile Virus, which has spread throughout the Americas (Mann et al. 2013), differ in their likelihood to be sampled from different bird species. Our approach was to hypothesize ecologically distinct populations (ecotypes) of West Nile Virus, using the algorithm Ecotype Simulation 2 (Koeppel et al. 2008, Francisco et al. 2014). The algorithm analyzes the sequence diversity of a rarely sexual or entirely asexual group to identify sequence clusters that are likely to be ecologically distinct (ecotypes), without using any ecological data. When ES2 has previously been applied to closely related bacteria coexisting within the same region, the putative ecotypes have been confirmed to be ecologically distinct through differences in habitat associations and physiology (Connor et al. 2010, Koeppel et al. 2013, Becraft et al. 2015). The algorithm should be appropriate to apply to West Nile Virus, since this virus recombines extremely rarely (Pickett and Lefkowitz 2009).

We will demonstrate that within one geographical region, putative ecotypes of West Nile Virus identified by Ecotype Simulation are quantitatively different in the host bird species they tend to infect, suggesting ecological diversification within the virus species. We will argue that a within-species focus may be valuable in discovering viruses with differing ecological niches.

MATERIALS AND METHODS

Aligning genomes. We obtained 1590 full-genome sequences of West Nile Virus from the NCBI Virus Variation database (Hatcher et al. 2017) in September 2018, drawing on the global diversity of WNV genome sequences from all mosquito and vertebrate hosts sampled from 55 regions on 6 continents. Identical sequences were collapsed, leaving 1319 unique

sequences. The whole-genome sequences extracted varied in length from 10299 to 11355 bp. A built-in MUSCLE sequence alignment tool from the NCBI Virus Variation database constructed and exported an alignment file in FASTA format.

We used as an outgroup the whole genome sequence of Zika Virus strain KX893855, provided by NCBI. Zika and West Nile Virus are both in the *Flavivirus* genus, and Zika has a similar genome content and nucleotide length as WNV (Gupta et al. 2016, Rey et al. 2018). The Zika sequence was appended to the WNV sequence alignment through the “-add” function of MAFFT (Katoh and Standley 2013). We wrote a program, removeGaps.py, to remove gaps in the alignment file: <https://github.com/tkong233/Global-Virome-Project>. Applying the algorithm did not change the eventual demarcations of viruses into putative ecotypes. All programs written for this study are available through this Github link.

Constructing the phylogenetic Tree. We used FastTree to construct a maximum-likelihood phylogeny from the genome alignment in Newick format, which was visualized by MEGA7. We then rooted the tree to the Zika virus outgroup within MEGA7’s Tree Explorer window.

Demarcating putative ecotypes. We inputted both the alignment and Newick tree files into the Ecotype Simulation 2 algorithm. ES2 models the sequence diversity within an asexual or rarely-sexual clade to predict ecologically distinct populations from patterns of sequence clustering only, without using ecological data (Koeppel et al. 2008, Francisco et al. 2014) <https://github.com/sandain/ecosim/releases>. The log file of ES2 yielded ecotype demarcations, which were converted into an Excel file using our Python program dataframeGen.py. The unique sequences within each ecotype are listed in Supplementary Table 1.

Host associations of WNV putative ecotypes with taxonomic groupings of birds. One geographical region of particular interest comprised the adjacent states of Connecticut and New York in the USA, owing to sampling there of 362 birds providing West Nile Virus. Two bird species were abundantly sampled: the American crow (*Corvus brachyrhynchos*, 243 sequences), and the blue jay (*Cyanocitta cristata*, 80 sequences). Both these species are within the Corvidae family of the perching-bird order, Passeriformes. The remaining 39 birds sampled from Connecticut and New York included 20 other birds within the Passeriformes, 10 birds within the hawk-like order Accipitriformes, and 9 birds within the owl order Strigiformes. We tested for host association differences among putative ecotypes from a matrix of 5 taxonomic groups of birds (blue jays, American crows, other passeriforms, Accipitriformes, and Strigiformes) X 23 putative ecotypes, noting that putative ecotypes sampled fewer than 5 times from birds in Connecticut and New York were discarded from the analysis (Table 1). The matrix was analyzed by a contingency test of association using the G test in the DescTool package within RStudio. Finally, a phylogenetic tree containing only viruses sampled from Connecticut and New York birds was created with MAFFT, FastTree, and MEGA7 (Figure 1).

RESULTS

Demarcation of WNV sequences into putative ecotypes. Based on the 1319 unique whole-genome sequences of WNV sampled worldwide, the Ecotype Simulation algorithm hypothesized and demarcated 69 putative ecotypes (Table 1, Supplementary Table 1).

Associations of WNV putative ecotypes with bird taxa sampled in Connecticut and New York. From the Connecticut-New York region, we focused on WNV sequences from five

taxonomic groups: the two bird species within the Passeriformes order that were most abundantly sampled (American crows and blue jays), various species less well sampled within the Passeriformes (“other passeriforms”), and two groups constituting the Accipitriformes and Strigiformes orders (Table 1). A contingency test yielded a highly significant heterogeneity among putative ecotypes for their associations with these host groups ($G = 133.15$, $df = 84$, $P = 0.000514$) (Figure 1). Among the most salient ecotype differences was the range of relative frequencies of American crows and blue jays, with for example, American crows ranging from 25% to 100% among the 23 putative ecotypes analyzed (Table 1).

DISCUSSION

Differences among WNV putative ecotypes in their associations with bird host species.

Our aim was to identify important ecological and evolutionary divergence within a virus species taxon already known to be pathogenic for humans. We focused on West Nile Virus because it infects a variety of hosts and hence may have diversified by host specificity. We found that West Nile clades differed in their associations with bird hosts within the Connecticut-New York region. For example, the frequency of WNV samples in American crows varied from 25% to 100% among the 23 putative ecotypes analyzed. Because this study was circumscribed to one geographical region, where the bird hosts, mosquito vectors, and the viruses could disperse, the host associations are likely due to some intrinsic ecological difference (or differences) among the ecotypes.

One possible explanation for the host associations may be that some WNV lineages have evolved to reach higher blood titers in some bird species than in others, such that virologists

are better able to detect them there. Alternatively, WNV lineages may have diversified by the mosquitoes they can most successfully infect (as seen for lineages within a fungal species) (Zebold et al. 1979). Provided that mosquito species differ in their likelihood to feed on one species of bird over another, virus diversification by mosquito vector could lead to the host associations we have observed.

Limited opportunity to extend this analysis to associations of WNV lineages with humans. It is tempting to extend this line of analysis to compare putative ecotypes for their tendencies to associate with humans. Sufficient genome sequence data are available for this purpose, with WNV genome sequences from 569 birds and 69 humans (nearly all asymptomatic) sampled in the USA. Unfortunately, geographical metadata are entirely lacking for the human samples, beyond that they stem from the USA. The omission of the metadata is exacerbated by the fact that WNV overwinters throughout the continental USA (Reisen 2013) and so has developed some endemicity of lineages in geographic regions within the country (Hepp et al. 2018). Therefore, one could not absolutely ensure that any observed associations of putative ecotypes with humans did not emerge as an artifact of differences in where humans and birds were sampled.

Omission of important details in environmental metadata has repeatedly limited microbial ecologists' ability to study the genetics of microbial adaptations to environments and hosts (Cohan 2012). We suggest that virus genome sequences should report at least minimal geographical and environmental metadata about where each sample was collected, as guided by Minimum Information about any(x) Sequence (MIxS) (Yilmaz et al. 2011).

Nevertheless, we tentatively explored differences in our WNV putative ecotypes in their tendencies to be sampled from humans versus birds. To try to reduce geographical bias, we focused only on ecotypes that were present in regions where both humans and birds were sampled (that is, ecotypes containing at least one bird and one human sample). The 18 putative ecotypes meeting this criterion were considerably different in their frequencies of human samples, ranging from 4.8% to 65%, with a highly significant heterogeneity in a contingency test ($G = 76.828$, $df = 17$, $P = 1.40e-09$) (Supplementary Table 2).

One possible explanation may be that some virus ecotypes reproduce more aggressively within humans, leading to a greater probability of being detected in blood samples in asymptomatic humans. Alternatively, some ecotypes may reproduce more successfully within the peridomestic mosquitos that most frequently bite humans, such as *Aedes albopictus*. This would lead to greater rates of infection of these ecotypes in humans. Finally, and unavoidably with the present data set, the ecotype differences could reflect a geographical bias in sampling humans.

Potential implications of WNV bird associations for human spillovers. Evolutionary diversification of WNV ecotypes to different bird species could possibly impact the likelihood of spillover to humans. This is because some animal species serve better as an “ecological bridge” to transmit viruses to humans. For example, humans would never have become infected by Hendra Virus directly from its reservoir hosts in the *Pteropus* fruit bat genus. However, when the bats transmit the infection to horses, the virus is sufficiently amplified in horses’ saliva to infect the horses’ caretakers (NSW Government 2017). Also, pigs have served as an ecological bridge for infection of pigs’ caretakers by Nipah virus, again from *Pteropus* bats (National

Center for Emerging and Zoonotic Infectious Diseases 2017). We suggest that evolution of a WNV ecotype that is particularly good at infecting a suburban backyard bird at high blood titers may result more frequently in human infections.

In some cases, a virus can more easily evolve adaptation to a new host species if it has previously evolved to infect an “evolutionary bridge” host. This is seen recurrently in Influenza, where pigs serve as a vessel for bird-adapted influenza to become adapted to mammals (Webster et al. 1992). Also, the Parvovirus lineage that eventually became Canine Parvovirus II is thought to have utilized its brief time infecting raccoons in evolving from a cat-infecting to a dog-infecting virus (Allison et al. 2012). Likewise, some bird species may possibly serve as an evolutionary bridge for WNV to more successfully infect humans, facilitating spillover by some WNV ecotypes more than others.

Implications for the Global Virome Project. The GVP has generally sought to look for novel viruses that are outside of existing species taxa. However, virologists have found occasional evidence of important ecological divergence within virus species taxa (Ebel et al. 2004, Ariën et al. 2007, Streicker et al. 2010, Blanquart et al. 2016). Our present study presents a proof of concept that we can easily hypothesize putative ecotypes within a virus species taxon from genome sequence data, and then test the putative ecotypes for ecological distinctness by exploring for differences in host associations.

258 **Acknowledgements**

259 We are grateful for funding from Wesleyan University, the John and Rosemarie Dooley
260 Foundation, and from the Huffington Foundation.

261

262 **Data Availability Statement**

263 The data we have created are available in Table 1 and Supplementary Tables 1 and 2.

264

265 **Ethical Statement**

266 An ethical statement is not applicable as the data we have analyzed were collected from
267 the NCBI database.

268

269 **Table 1. Host associations of WNV putative ecotypes with five taxonomic groups of birds in**
 270 **the Connecticut-New York region.** Listed are the total numbers of viruses from each ecotype
 271 infecting each host category, without regard to whether the sequences were unique. Included
 272 are those ecotypes with at least five viruses sampled from the Connecticut-New York region.

273

	American crows	Blue jays	Other Passeriforms	Accipitriformes	Strigiformes
Ecotype 2	9	3	1	0	0
Ecotype 3	10	0	0	0	0
Ecotype 4	35	9	1	1	0
Ecotype 15	3	3	0	0	1
Ecotype 20	9	1	0	0	0
Ecotype 21	5	2	0	0	0
Ecotype 23	7	4	0	1	0
Ecotype 24	7	0	3	1	0
Ecotype 30	6	4	0	0	0
Ecotype 35	3	6	3	0	0
Ecotype 38	13	1	0	0	0
Ecotype 39	4	6	0	0	0
Ecotype 40	19	8	6	0	3
Ecotype 41	9	2	1	0	0
Ecotype 47	7	4	1	0	0
Ecotype 48	11	5	0	0	0
Ecotype 50	8	10	0	0	1
Ecotype 51	8	2	1	1	0

Ecotype 52	13	1	0	1	0
Ecotype 53	8	2	0	1	2
Ecotype 56	13	4	0	1	2
Ecotype 58	11	2	1	0	0
Ecotype 59	23	1	2	3	0

274

275

Figure 1. Host associations of WNV putative ecotypes with five taxonomic groups of birds in the Connecticut-New York region. The Zika virus strain KX893855 served as the outgroup.

280 **SUPPORTING INFORMATION**

281

282 **Supplementary Table 1. Sequence names of the 1319 unique WNV whole-genome**
 283 **sequences, as demarcated into putative ecotypes by Ecotype Simulation 2.**

284 **Supplementary Table 2. Host associations of WNV putative ecotypes with birds versus**
 285 **humans in the USA.** Listed are the total numbers of viruses from each ecotype infecting each
 286 host category, without regard to whether the sequences were unique. Included are those
 287 ecotypes with at least one human and one bird sample from the USA.

288

289

LITERATURE CITED

- Allison, A. B., C. E. Harbison, I. Pagan, K. M. Stucker, J. T. Kaelber, J. D. Brown, M. G. Ruder, M. K. Keel, E. J. Dubovi, E. C. Holmes, and C. R. Parrish. 2012. Role of multiple hosts in the cross-species transmission and emergence of a pandemic parvovirus. *J Virol* **86**:865-872.
- Ariën, K. K., G. Vanham, and E. J. Arts. 2007. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* **5**:141-151.
- Bao, Y., V. Chetvernin, and T. Tatusova. 2014. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. *Archives of Virology* **159**:3293-3304.
- Becraft, E. D., J. M. Wood, D. B. Rusch, M. Köhl, S. I. Jensen, D. A. Bryant, D. W. Roberts, F. M. Cohan, and D. M. Ward. 2015. The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Front Microbiol* **6**:590.
- Blanquart, F., M. K. Grabowski, J. Herbeck, F. Nalugoda, D. Serwadda, M. A. Eller, M. L. Robb, R. Gray, G. Kigozi, O. Laeyendecker, K. A. Lythgoe, G. Nakigozi, T. C. Quinn, S. J. Reynolds, M. J. Wawer, and C. Fraser. 2016. A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *eLife* **5**:e20492.
- Carroll, D., P. Daszak, N. D. Wolfe, G. F. Gao, C. M. Morel, S. Morzaria, A. Pablos-Mendez, O. Tomori, and J. A. K. Mazet. 2018. The Global Virome Project. *Science* **359**:872-874.
- Cohan, F. M. 2012. Science needs more *Moneyball*. *American Scientist* **100**:182-185.
- Cohan, F. M. 2017. Transmission in the Origins of Bacterial Diversity, From Ecotypes to Phyla. *Microbiology spectrum* **5**.
- Colpitts, T. M., M. J. Conway, R. R. Montgomery, and E. Fikrig. 2012. West Nile Virus: biology, transmission, and human infection. *Clin Microbiol Rev* **25**:635-648.
- Connor, N., J. Sikorski, A. P. Rooney, S. Kopac, A. F. Koepfel, A. Burger, S. G. Cole, E. B. Perry, D. Krizanc, N. C. Field, M. Slaton, and F. M. Cohan. 2010. The ecology of speciation in *Bacillus*. *Applied and Environmental Microbiology* **76**:1349-1358.
- Daszak, P., C. Zambrana-Torrel, T. L. Bogich, M. Fernandez, J. H. Epstein, K. A. Murray, and H. Hamilton. 2013. Interdisciplinary approaches to understanding disease emergence: The past, present, and future drivers of Nipah virus emergence. *Proceedings of the National Academy of Sciences of the United States of America* **110**:3681-3688.
- Denef, V. J., L. H. Kalnejais, R. S. Mueller, P. Wilmes, B. J. Baker, B. C. Thomas, N. C. VerBerkmoes, R. L. Hettich, and J. F. Banfield. 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc Natl Acad Sci U S A* **107**:2383-2390.
- Drosten, C., P. Kellam, and Z. A. Memish. 2014. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med* **371**:1359-1360.
- Duggal, N. K., A. Bosco-Lauth, R. A. Bowen, S. S. Wheeler, W. K. Reisen, T. A. Felix, B. R. Mann, H. Romo, D. M. Swetnam, A. D. Barrett, and A. C. Brault. 2014. Evidence for co-evolution of West Nile Virus and house sparrows in North America. *PLoS Negl Trop Dis* **8**:e3262.
- Ebel, G. D., J. Carricaburu, D. Young, K. A. Bernard, and L. D. Kramer. 2004. Genetic and phenotypic variation of West Nile Virus in New York, 2002-2003. *The American Journal of Tropical Medicine and Hygiene* **71**:493-500.
- Francisco, J. C., F. M. Cohan, and D. Krizanc. 2014. Accuracy and efficiency of algorithms for the demarcation of bacterial ecotypes from DNA sequence data. *Int. J. Bioinformatics Research and Applications* **10**:409-425.

Gupta, A. K., K. Kaur, A. Rajput, S. K. Dhanda, M. Sehgal, M. S. Khan, I. Monga, S. A. Dar, S. Singh, G. Nagpal, S. S. Usmani, A. Thakur, G. Kaur, S. Sharma, A. Bhardwaj, A. Qureshi, G. P. S. Raghava, and M. Kumar. 2016. ZikaVR: An Integrated Zika Virus Resource for Genomics, Proteomics, Phylogenetic and Therapeutic Analysis. *Scientific Reports* **6**:32713-32713.

Hadfield, J., A. F. Brito, D. M. Swetnam, C. B. F. Vogels, R. E. Tokarz, K. G. Andersen, R. C. Smith, T. Bedford, and N. D. Grubaugh. 2019. Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. *PLoS Pathogens* **15**:e1008042.

Hatcher, E. L., S. A. Zhdanov, Y. Bao, O. Blinkova, E. P. Nawrocki, Y. Ostapchuck, A. A. Schäffer, and J. R. Brister. 2017. Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic acids research* **45**:D482-D490.

Hepp, C. M., J. H. Cocking, M. Valentine, S. J. Young, D. Damian, K. E. Samuels-Crow, K. Sheridan, V. Y. Fofanov, T. N. Furstenau, J. D. Busch, D. E. Erickson, R. C. Lancione, K. Smith, J. Will, J. Townsend, P. S. Keim, and D. M. Engelthaler. 2018. Phylogenetic analysis of West Nile Virus in Maricopa County, Arizona: Evidence for dynamic behavior of strains in two major lineages in the American Southwest. *PLoS One* **13**:e0205801-e0205801.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**.

Koeppel, A., E. B. Perry, J. Sikorski, D. Krizanc, W. A. Warner, D. M. Ward, A. P. Rooney, E. Brambilla, N. Connor, R. M. Ratcliff, E. Nevo, and F. M. Cohan. 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences* **105**:2504-2509.

Koeppel, A. F., J. O. Wertheim, L. Barone, N. Gentile, D. Krizanc, and F. M. Cohan. 2013. Speedy speciation in a bacterial microcosm: New species can arise as frequently as adaptations within a species. *ISME J* **7**:1080-1091.

Konstantinidis, K. T., and J. M. Tiedje. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**:2567-2572.

Lauber, C., and A. E. Gorbalenya. 2012. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. *Journal of virology* **86**:3890-3904.

Mann, B. R., A. R. McMullen, D. M. Swetnam, and A. D. Barrett. 2013. Molecular epidemiology and evolution of West Nile virus in North America. *Int J Environ Res Public Health* **10**:5111-5129.

Marston, M. F., and C. G. Amrich. 2009. Recombination and microdiversity in coastal marine cyanophages. *Environ Microbiol* **11**:2893-2903.

Morse, S. S., J. A. Mazet, M. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrel, W. I. Lipkin, and P. Daszak. 2012. Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**:1956-1965.

National Center for Emerging and Zoonotic Infectious Diseases. 2017. Nipah virus: transmission.

NSW Government. 2017. Communicable Diseases Factsheet: Hendra virus. *in Health*, editor.

Pickett, B. E., and E. J. Lefkowitz. 2009. Recombination in West Nile Virus: minimal contribution to genomic diversity. *Virology Journal* **6**:165-165.

Plowright, R. K., P. Foley, H. E. Field, A. P. Dobson, J. E. Foley, P. Eby, and P. Daszak. 2011. Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (*Pteropus* spp.). *Proc Biol Sci* **278**:3703-3712.

Reisen, W. K. 2013. Ecology of West Nile virus in North America. *Viruses* **5**:2079-2105.

Rey, F. A., K. Stiasny, M.-C. Vaney, M. Dellarole, and F. X. Heinz. 2018. The bright and the dark side of human antibody responses to flaviviruses: lessons for vaccine design. *EMBO reports* **19**:206-224.

Rulli, M. C., M. Santini, D. T. S. Hayman, and P. D'Odorico. 2017. The nexus between forest fragmentation in Africa and Ebola virus disease outbreaks. *Scientific Reports* **7**:41613.

- Shapiro, B. J., J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabo, M. F. Polz, and E. J. Alm. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**:48-51.
- Simmonds, P., M. J. Adams, M. Benkő, M. Breitbart, J. R. Brister, E. B. Carstens, A. J. Davison, E. Delwart, A. E. Gorbalenya, B. Harrach, R. Hull, A. M. Q. King, E. V. Koonin, M. Krupovic, J. H. Kuhn, E. J. Lefkowitz, M. L. Nibert, R. Orton, M. J. Roossinck, S. Sabanadzovic, M. B. Sullivan, C. A. Suttle, R. B. Tesh, R. A. van der Vlugt, A. Varsani, and F. M. Zerbini. 2017. Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* **15**:161.
- Simmonds, P., and P. Aiewsakun. 2018. Virus classification – where do you draw the line? *Archives of Virology* **163**:2037-2046.
- Streicker, D. G., S. M. Altizer, A. Velasco-Villa, and C. E. Rupprecht. 2012. Variable evolutionary routes to host establishment across repeated rabies virus host shifts among bats. *Proc Natl Acad Sci U S A* **109**:19715-19720.
- Streicker, D. G., A. S. Turmelle, M. J. Vonhof, I. V. Kuzmin, G. F. McCracken, and C. E. Rupprecht. 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329**:676-679.
- Warwick-Dugdale, J., N. Solonenko, K. Moore, L. Chittick, A. C. Gregory, M. J. Allen, M. B. Sullivan, and B. Temperton. 2019. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* **7**:e6800.
- Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza A viruses. *Microbiological reviews* **56**:152-179.
- Woolhouse, M. E., R. Howey, E. Gaunt, L. Reilly, M. Chase-Topping, and N. Savill. 2008. Temporal trends in the discovery of human viruses. *Proc Biol Sci* **275**:2111-2115.
- Yilmaz, P., R. Kottmann, D. Field, R. Knight, J. R. Cole, L. Amaral-Zettler, J. A. Gilbert, I. Karsch-Mizrachi, A. Johnston, G. Cochrane, R. Vaughan, C. Hunter, J. Park, N. Morrison, P. Rocca-Serra, P. Sterk, M. Arumugam, M. Bailey, L. Baumgartner, B. W. Birren, M. J. Blaser, V. Bonazzi, T. Booth, P. Bork, F. D. Bushman, P. L. Buttigieg, P. S. Chain, E. Charlson, E. K. Costello, H. Huot-Creasy, P. Dawyndt, T. DeSantis, N. Fierer, J. A. Fuhrman, R. E. Gallery, D. Gevers, R. A. Gibbs, I. San Gil, A. Gonzalez, J. I. Gordon, R. Guralnick, W. Hankeln, S. Highlander, P. Hugenholtz, J. Jansson, A. L. Kau, S. T. Kelley, J. Kennedy, D. Knights, O. Koren, J. Kuczynski, N. Kyrpides, R. Larsen, C. L. Lauber, T. Legg, R. E. Ley, C. A. Lozupone, W. Ludwig, D. Lyons, E. Maguire, B. A. Methe, F. Meyer, B. Muegge, S. Nakielnny, K. E. Nelson, D. Nemergut, J. D. Neufeld, L. K. Newbold, A. E. Oliver, N. R. Pace, G. Palanisamy, J. Peplies, J. Petrosino, L. Proctor, E. Pruesse, C. Quast, J. Raes, S. Ratnasingham, J. Ravel, D. A. Relman, S. Assunta-Sansone, P. D. Schloss, L. Schriml, R. Sinha, M. I. Smith, E. Sodergren, A. Spo, J. Stombaugh, J. M. Tiedje, D. V. Ward, G. M. Weinstock, D. Wendel, O. White, A. Whiteley, A. Wilke, J. R. Wortman, T. Yatsunenko, and F. O. Glockner. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat Biotechnol* **29**:415-420.
- Yip, C. W., C. C. Hon, M. Shi, T. T. Lam, K. Y. Chow, F. Zeng, and F. C. Leung. 2009. Phylogenetic perspectives on the epidemiology and origins of SARS and SARS-like coronaviruses. *Infect Genet Evol* **9**:1185-1196.
- Yu, C., T. Hernandez, H. Zheng, S. C. Yau, H. H. Huang, R. L. He, J. Yang, and S. S. Yau. 2013. Real time classification of viruses in 12 dimensions. *PLoS One* **8**:e64328.
- Zebold, S. L., H. C. Whisler, J. A. Shemanchuk, and L. B. Travland. 1979. Host Specificity and Penetration in the Mosquito Pathogen *Coelomomyces-Psorophorae*. *Canadian Journal of Botany-Revue Canadienne De Botanique* **57**:2766-2770.