1 **Title**: Reticulate evolution is favored in microbial niche switching

2 **Authors**: Eric J. Ma, Nichola J. Hill, Justin Zabilansky, Kyle Yuan, Jonathan A. Runstadler

3 **Affiliations**: Department of Biological Engineering, MIT

4

Abstract

Reticulate evolution is thought to accelerate the process of evolution beyond simple genetic drift and selection, helping to rapidly generate novel hybrids with combinations of adaptive traits. However, the long-standing dogma that reticulate evolutionary processes are likewise advantageous for switching ecological niches, as in microbial pathogen host switch events, has not been explicitly tested. We use data from the influenza genome sequencing project and a phylogenetic heuristic approach to show that reassortment, a reticulate evolutionary mechanism, predominates over mutational drift in transmission between different host species. Moreover, as host evolutionary distance increases, reassortment is increasingly favored. We conclude that the greater the quantitative difference between ecological niches, the greater the importance of reticulate evolutionary processes in overcoming niche barriers.

16 **Manuscript**

17 Reticulate evolutionary processes, such as horizontal gene transfer (HGT) and genomic
18 reassortment, have been proposed as a major mechanism for microbial evolution (*1*), aiding in
19 the diversification into new ecological niches (*2*). In contrast to clonal adaptation through genetic
20 drift over time, reticulate evolutionary processes allow an organism to acquire independently
21 evolved genetic material that can confer new fitness-enhancing traits. Examples include the
22 acquisition of cell surface receptor adaptations (point mutations) in viruses (*3*), and antibiotic
23 resistance (single genes) (*4*) and pathogenicity islands (or gene clusters) in bacteria (*5*). Host
24 switching, defined as a pathogen moving from one host species into another, represents a strong
25 fitness barrier to microbial pathogens. The acquisition of adaptations through reticulate processes
26 either prior to or after transmission from one species to another may serve to aid successful
27 pathogen host switches by improving fitness and the likelihood of continued transmission (*6*). In
28 this sense, reticulate evolution may be viewed as an ecological strategy for switching between
29 ecological niches (such as different host species), complementing but also standing in contrast to
30 the clonal adaptation of a microbial pathogen by genetic drift under selection. In order to test this
31 idea and its importance in host switch events, which are critical for (re-)emerging infectious
32 disease, we provide a quantitative assessment of the relative importance of reticulate processes
33 versus clonal adaptation in aiding the ecological niche switch of a viral pathogen

34 Data yielded from the influenza genome sequencing projects provides a unique opportunity for
35 quantitatively testing this concept, and is suitable for the following reasons. Firstly, the influenza
36 A virus (IAV) has a broad host tropism (*7*), and is capable of infecting organisms spanning
37 millennia of divergence on the tree of life. With different host-specific restriction factors forming
38 an adaptive barrier, each host species may then be viewed as a unique ecological niche for the

39 virus (*8*). Secondly, IAV is capable of and frequently undergoes reassortment, which is a well
40 documented reticulate evolutionary process (*9-12*). Finally, due to surveillance efforts over the
41 past two decades, whole genome sequences have been intensively sampled over a long time
42 frame, with corresponding host species metadata, available in an easily accessible and structured
43 format (*13*). Because reassortant viruses are the product of two or more genetically distinct
44 viruses co-infecting the same host, a more complex process than clonal transmission and
45 adaptation, they are expected to occur less frequently. Hence, the comprehensive IAV dataset,
46 which stretches over time and space with large sample numbers, provides the necessary scope to
47 detect reassortant viruses at a scale required to quantitatively assess the relative importance of
48 reticulate events in viral host switching.

49 In order to identify reassortment events and the hosts species involved, we adapted a
50 phylogenetic heuristic method (*14*), and mapped out a network of clonal and reassortment
51 descent relationships from a comprehensive set of completely sequenced IAV (18,632 viral
52 genomes) downloaded from the Influenza Research Database (*13*). Briefly, the core logic of the
53 method is as such: for every isolate in the dataset, we look for genomic sources such that the
54 sources found are of maximal similarity across all 8 genomic segments (Materials and Methods).
55 Clonal descent involves single sources, while reassortment descent involves source pairs. Where
56 multiple sources or source pairs correspond to the maximal similarity, all are kept as plausible
57 sources. In the resulting network, nodes are individual viral isolates, and edges are the clonal or
58 reassortment descent relationships.

59 In this network of viral isolates, clonal descent is mostly structured by host species, with known
60 global patterns of human-to-human (H3N2 & H1N1, and rarer H5N1 & H7N9), chicken-to-
61 chicken (H9N2, H7N9, H5N1) and swine-to-swine (H3N2, H1N1, H1N2) viral circulation

62 captured in the network reconstruction (Supplementary Figure 1). Edges in the network
63 connected viral isolates with a median genetic similarity of 99.7%, indicating high degrees of
64 phylogenetic similarity (Supplementary Figure 2). As expected, no clonal descent was identified
65 between viruses of different subtypes. Moreover, the network recreates the phylogeny of known
66 reassortant viruses, including the 2009 pandemic H1N1 and the recent 2013 H7N9 viruses,
67 further validating the accuracy of our reconstruction. Small-world simulation studies and
68 phylogenetic comparisons validated our method as being accurate in detecting reassortment
69 events (Supplementary Figures 3 and 4).

70 To test whether reassortment or clonal descent was an advantageous strategy when switching
71 hosts, we computed the weighted proportion of reassortant edges (out of all edges) occurring
72 between hosts of the same or different species. When host species were different, reassortant
73 edges were over-represented at 19 percentage points above a null permutation model
74 (permutation test described in Materials & Methods) (Figure 1a), and when host species were the
75 same, reassortant edges were under-represented by 7 percentage points relative to our null
76 model. Thus, reassortment is a strongly favored strategy when influenza crosses between
77 different host species.

78 We further sought to explore whether the predominant use of reticulate evolutionary processes in
79 host switch events were correlated with host phylogenetic relatedness. To do this, we first
80 computed the proportion of reassortment when switching between birds, non-human mammals,
81 or humans, which are 3 divergent host groupings. We further sub-divided avian and mammalian
82 categories into wild and domestic, to assess the impact of anthropological activity on the relative
83 importance of reassortment in host switch interfaces (see Materials and Methods for how AIV
84 was classified as domestic or wild). To ensure that the dataset was sufficient in scope to detect

reassortant viruses, we only considered host group transitions with at least 1000 descent events (both clonal and reassortant), or at least 10 reassortment events (dashed lines in Figures 1b & c). Nonetheless, all data are displayed for completeness.

Here, reassortment is over-represented relative to the null when host groups are different. Only two exceptions occur. The first is between wild birds, where reassortment is over-represented but host groups are not different. In this case, the "wild bird" label encompasses a wide range of host species, and as the natural reservoir for many diverse influenza viral subtypes, we expect to detect reassortment events more frequently between diverse species that may be distantly evolutionarily related. The second is the human-domestic mammal interface, where reassortment is not over-represented even though the host groups are different. In the case of human to domestic mammal host switches (reverse zoonosis), these are mostly well-documented reverse zoonotic events between human and swine hosts (*15*)), where shared cellular receptors for viral entry (*16*) facilitates zoonotic and reverse zoonotic transmission. This may be a case of host convergent evolution inadvertently lowering the adaptive barrier to host switching. Under representation of reassortment at human-to-human transitions is expected because of the limited number of highly similar viral subtypes circulating in human populations, which likely obscures the distinction between reassortment and clonal descent. However, we also expect disease control measures to further limit the frequency of co-infection and likelihood of reassortment events. Thus, despite exceptions which are explained by known influenza biology (e.g. human to swine transmissions), reassortment is strongly favored over clonal evolution when crossing between evolutionarily distant hosts.

To further explore the relationship between host evolutionary divergence and the predominance of reassortment in transmission events between species, we compared a common phylogenetic

measure of species divergence, the cytochrome oxidase I (COI) gene, to the use of reassortment in host switch events. A subset of viral hosts, encompassing a variety of bird and mammal species, have had their cytochrome oxidase I (COI) gene sequenced as part of the barcode of life project (*17*). For the subset of edges in the network for which both the source and sink hosts have a COI gene sequence that fulfilled our criteria for consideration (as described above), we computed the percentage evolutionary distance between the two hosts (Materials and Methods). Applying a similar permutation test and assessment criteria as described for host groups above, we found a trend of increasing over-representation at higher evolutionary distances (Figure 1c). Thus, as host evolutionary distance, or (more broadly, quantitative niche dissimilarity) increases, reticulate evolution becomes increasingly favored for influenza virus niche switch events.

We have quantitatively defined the importance of reticulate evolutionary events in switching ecological niches, using an infectious disease data set with characteristics that are particularly well suited for answering this question. Beyond the viral world, recent reviews have asserted the importance of reticulate evolutionary events as a driver of speciation and niche diversification (*18*, *19*), and recent studies have illustrated heightened fitness effects in hybrid populations (*20*, *21*). However, none have quantitatively tested the importance of reticulate evolutionary strategies in enabling ecological niche switches at a global scale, especially in comparison to clonal adaptation under drift and selection, a task feasible only in fast evolving organisms. To date, no studies have examined reticulate evolutionary processes in the context of quantified niche differences, as we have done here by measuring reassortment in the context of host evolutionary distance. Our study provides strong quantitative evidence supporting the hypothesis that reticulate evolutionary processes are advantageous relative to adaptation by drift for pathogen transfer between host species, and therefore more broadly, ecological niche switching.

We recognize that in this study, we have considered only a single pathogen for which abundant genomic data are available, and whose genomic and host tropic characteristics are suitable for this analysis. To specifically answer whether reticulate processes are favored over clonal transmission for other organisms, using these methods, depends on being able to acquire genome sequences with matched ecological niche metadata. We also note that the global influenza dataset will have unavoidable sampling biases. For example, human isolates predominate in the dataset, and consequently the human-associated subtypes H3N2 and H1N1 also dominate the dataset. Sequences from viral outbreaks will also be over-represented relative to isolates collected through routine surveillance sampling, and will unavoidably lead to a heightened detection of clonal descent in a single host species. In order to deal with this sampling bias, our permutation tests (for the host species and group labels) involve class labels of equal sizes. This allows us to calculate an expected distribution of proportions under ideal assumptions of equal sampling, which in turn forms the baseline for our conclusions.

In summary, using data available from a model zoonotic viral pathogen, we have shown that a reticulate evolutionary strategies are important in enabling pathogen host switches. For the influenza virus, reticulate evolution predominates when crossing between hosts. More broadly, the greater the quantitative difference between ecological niches, the greater the importance of reticulate evolutionary processes in enabling niche switches. While the quantitative importance of reticulate evolution may differ for different organisms evolving in different niches, we expect that further sequencing efforts from across broad domains of microbial life, and a further characterization and definition of their ecological niches, will elucidate whether this principle holds more broadly. Beyond its relevance to evolutionary ecology, reticulate evolution also has public health consequences. Reassortant influenza viruses have been implicated in all past

154 human pandemic strains (*22-25*), and the ancestry of HIV-1 involved a hybrid SIV (*26*). Hence,
155 knowing how reticulate events shape disease emergence may help the ecology and evolution of
156 infectious disease become a more predictive science, leading to insight important to disease
157 prevention and mitigation (*27*).

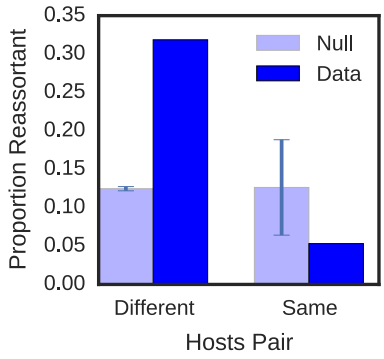
158 **Acknowledgments**

159 The authors would like to thank Dr. Nan Li for technical assistance on graph computation during
160 the earlier stages of this work. We would like to acknowledge funding from the NIH/NIAID
161 CEIRS Program (Contract no.: HHSN272014000008C) and the MIT Department of Biological
162 Engineering. The authors would also like to thank William R. Hesse of the MIT BE
163 Communications Lab for assistance in reviewing the manuscript.

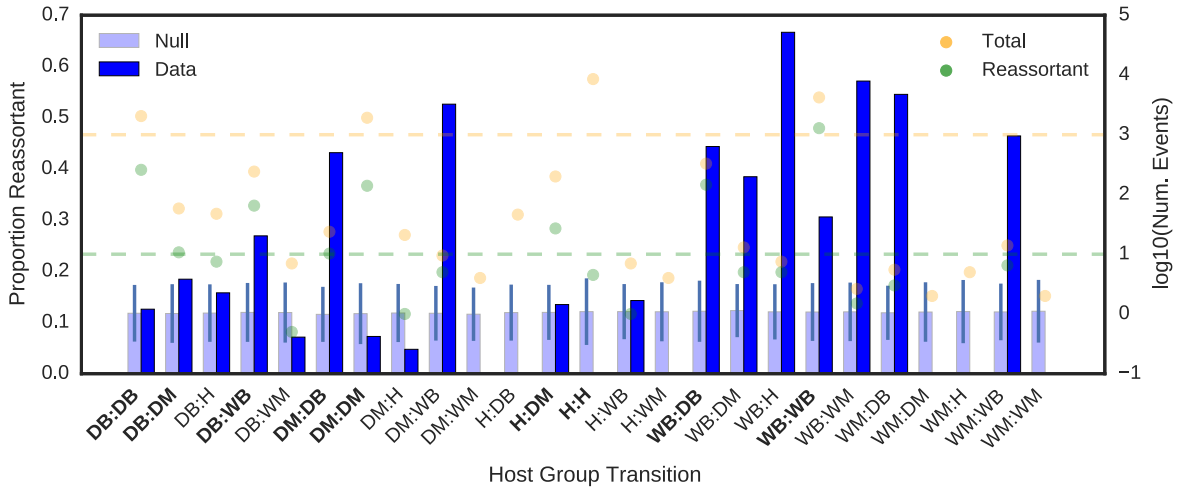
164

Figures

a.



b.



c.

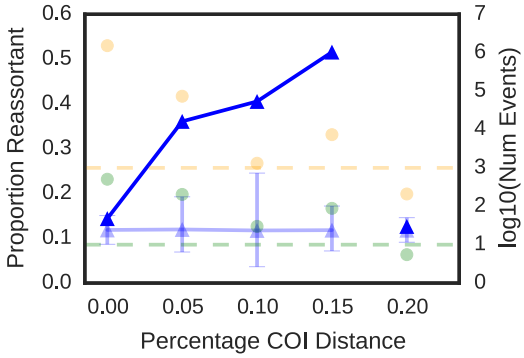


Figure 1. Proportion of reassortment events when crossing between (a) different or same hosts, (b) different host groups, and (c) hosts of differing evolutionary distance as measured by divergence in the cytochrome oxidase I (COI) gene. Reassortment is over-represented relative to

172 clonal descent in transmission across host barriers. (b) D: Domestic animal, H: Human, W: Wild,
173 B: Bird, M: Mammal. Donor host is labeled first. Bolded x-axis tick labels indicate data for
174 which the weighted sum of all edges exceeded 1000, or the weighted sum of reassortant edges
175 exceeded 10. (a, b, c) Vertical error bars on the null permutation model represent 3 standard
176 deviations from the mean from 100 simulations (a, b), or 95% density intervals from 500
177 simulations (c). (b, c) Translucent dots indicate the weighted sum of clonal descent (yellow) and
178 reassortment (green) events detected in the network under each host group transition. Horizontal
179 yellow and green lines indicates threshold of values of 1000 and 10 respectively.

180

References

1. A. Hernández-López *et al.*, To tree or not to tree? Genome-wide quantification of recombination and reticulate evolution during the diversification of strict intracellular bacteria. *Genome Biol Evol*. **5**, 2305–2317 (2013).

2. D. Peris *et al.*, Population structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids. *Molecular Ecology*. **23**, 2031–2045 (2014).

3. R. J. Garten *et al.*, Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science*. **325**, 197–201 (2009).

4. C. S. Smillie *et al.*, Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*. **480**, 241–244 (2011).

5. U. Antonenka, C. Nölting, J. Heesemann, A. Rakin, Horizontal transfer of Yersinia high-pathogenicity island by the conjugative RP4 attB target-presenting shuttle plasmid. *Mol. Microbiol.* **57**, 727–734 (2005).

6. S. K. Remold, A. Rambaut, P. E. Turner, Evolutionary genomics of host adaptation in vesicular stomatitis virus. *Mol. Biol. Evol.* **25**, 1138–1147 (2008).

7. R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, Y. Kawaoka, Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179 (1992).

8. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12**, 687–695 (2012).

9. C. Li *et al.*, Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4687–4692 (2010).

10. A. Mehle, V. G. Dugan, J. K. Taubenberger, J. A. Doudna, Reassortment and mutation of the avian influenza virus polymerase PA subunit overcome species barriers. *J. Virol.* **86**, 1750–1757 (2012).

11. T. T.-Y. Lam *et al.*, Reassortment Events among Swine Influenza A Viruses in China: Implications for the Origin of the 2009 Influenza Pandemic. *J. Virol.* **85**, 10279–10285 (2011).

12. H. Tao, J. Steel, A. C. Lowen, Intra-host dynamics of influenza virus reassortment. *J. Virol.* **88**, JVI.00715–14–7492 (2014).

13. R. B. Squires *et al.*, Influenza Research Database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza and Other Respiratory Viruses*. **6**, 404–416 (2012).

14. T. Jombart, R. M. Eggo, P. J. Dodd, F. Balloux, Reconstructing disease outbreaks from

genetic data: a graph approach. *Heredity*. **106**, 383–390 (2011).

15. M. I. Nelson, A. L. Vincent, Reverse zoonosis of influenza to swine: new perspectives on the human-animal interface. *Trends Microbiol.* **23**, 142–153 (2015).

16. W. Ma, R. E. Kahn, J. A. Richt, The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *J Mol Genet Med*. **3**, 158–166 (2008).

17. S. Ratnasingham, P. D. N. Herbert, bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes*. **7**, 355–364 (2007).

18. K. M. Dlugosch, S. R. Anderson, J. Braasch, F. A. Cang, H. D. Gillette, The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*. **24**, 2095–2111 (2015).

19. J. Molofsky, S. R. Keller, S. Lavergne, M. A. Kaproth, M. B. Eppinga, Human-aided admixture may fuel ecosystem transformation during biological invasions: theoretical and experimental evidence. *Ecol Evol*. **4**, 899–910 (2014).

20. K. J. F. Verhoeven, M. Macel, L. M. Wolfe, A. Biere, Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc. R. Soc. B*. **278**, 2–8 (2011).

21. S. R. Keller, P. D. Fields, A. E. Berardi, D. R. Taylor, Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species. *Journal of Evolutionary Biology*. **27**, 616–627 (2014).

22. A. Wu *et al.*, Sequential Reassortments Underlie Diverse Influenza H7N9 Genotypes in China. *Cell Host Microbe*. **14**, 446–452 (2013).

23. J. Pu *et al.*, Evolution of the H9N2 influenza genotype that facilitated the genesis of the novel H7N9 virus. *Proc. Natl. Acad. Sci. U.S.A.*, 201422456 (2014).

24. T. T.-Y. Lam *et al.*, The genesis and source of the H7N9 influenza viruses causing human infections in China. *Nature*. **502**, 241–244 (2013).

25. R. Gao *et al.*, Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).

26. E. Bailes *et al.*, Hybrid origin of SIV in chimpanzees. *Science*. **300**, 1713–1713 (2003).

27. B. R. Wasik, P. E. Turner, On the Biological Success of Viruses. *http://dx.doi.org/10.1146/annurev-micro-090110-102833*. **67**, 519–541 (2013).