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

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

**Affiliations**: Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA, 02139

**Corresponding Authors:** Jonathan A. Runstadler, jrun@mit.edu; Eric J. Ma, ericmajinglong@gmail.com

**Classification:**

Major: Biological Sciences

Minor: Ecology, Evolution

**Keywords:** Ecology, Reticulate Evolution, Influenza,

**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.

23 **Significance Statement:**

24 Are the processes that result in the exchange of genes between microbes quantitatively
25 advantageous for those microbes when switching between ecological niches? We consider the
26 influenza A virus as a model microbe, with its ability to infect multiple host species (ecological
27 niches) and undergo reassortment (exchange genes) with one another. Through our analysis of
28 sequence data from the Influenza Research Database and the Barcode of Life Database, we find
29 that the greater the quantitative difference between influenza hosts, the greater the proportion of
30 reassortment events were found. More broadly, for microbes, reticulate evolutionary processes
31 are quantitatively favoured when switching between ecological niches.

32

33 \body

34 **Manuscript**

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

52 Data yielded from the influenza genome sequencing projects provides a unique opportunity for
53 quantitatively testing this concept, and is suitable for the following reasons. Firstly, the influenza
54 A virus (IAV) has a broad host tropism (*7*), and is capable of infecting organisms spanning
55 millennia of divergence on the tree of life. With different host-specific restriction factors forming

56 an adaptive barrier, each host species may then be viewed as a unique ecological niche for the
57 virus (*8*). Secondly, IAV is capable of and frequently undergoes reassortment, which is a well
58 documented reticulate evolutionary process (*9-12*). Finally, due to surveillance efforts over the
59 past two decades, whole genome sequences have been intensively sampled over a long time
60 frame, with corresponding host species metadata, available in an easily accessible and structured
61 format (*13*). Because reassortant viruses are the product of two or more genetically distinct
62 viruses co-infecting the same host, a more complex process than clonal transmission and
63 adaptation, they are expected to occur less frequently. Hence, the comprehensive IAV dataset,
64 which stretches over time and space with large sample numbers, provides the necessary scope to
65 detect reassortant viruses at a scale required to quantitatively assess the relative importance of
66 reticulate events in viral host switching.

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

77 In this network of viral isolates, clonal descent is mostly structured by host species, with known
78 global patterns of human-to-human (H3N2 & H1N1, and rarer H5N1 & H7N9), chicken-to-

79 chicken (H9N2, H7N9, H5N1) and swine-to-swine (H3N2, H1N1, H1N2) viral circulation
80 captured in the network reconstruction (Supplementary Figure 1). Edges in the network
81 connected viral isolates with a median genetic similarity of 99.7%, indicating high degrees of
82 phylogenetic similarity (Supplementary Figure 2). As expected, no clonal descent was identified
83 between viruses of different subtypes. Moreover, the network recreates the phylogeny of known
84 reassortant viruses, including the 2009 pandemic H1N1 and the recent 2013 H7N9 viruses,
85 further validating the accuracy of our reconstruction. Small-world simulation studies and
86 phylogenetic comparisons validated our method as being accurate in detecting reassortment
87 events (Supplementary Figures 3 and 4).

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

96 We further sought to explore whether the predominant use of reticulate evolutionary processes in
97 host switch events were correlated with host phylogenetic relatedness. To do this, we first
98 computed the proportion of reassortment when switching between birds, non-human mammals,
99 or humans, which are 3 divergent host groupings. We further sub-divided avian and mammalian
100 categories into wild and domestic, to assess the impact of anthropological activity on the relative
101 importance of reassortment in host switch interfaces (see Materials and Methods for how AIV

102 was classified as domestic or wild). To ensure that the dataset was sufficient in scope to detect
103 reassortant viruses, we only considered host group transitions with at least 1000 descent events
104 (both clonal and reassortant), or at least 10 reassortment events (dashed lines in Figures 1b & c).
105 Nonetheless, all data are displayed for completeness.

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

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

136 We have quantitatively defined the importance of reticulate evolutionary events in switching
137 ecological niches, using an infectious disease data set with characteristics that are particularly
138 well suited for answering this question. Beyond the viral world, recent reviews have asserted the
139 importance of reticulate evolutionary events as a driver of speciation and niche diversification
140 (*18*, *19*), and recent studies have illustrated heightened fitness effects in hybrid populations (*20*,
141 *21*). However, none have quantitatively tested the importance of reticulate evolutionary
142 strategies in enabling ecological niche switches at a global scale, especially in comparison to
143 clonal adaptation under drift and selection, a task feasible only in fast evolving organisms. To
144 date, no studies have examined reticulate evolutionary processes in the context of quantified
145 niche differences, as we have done here by measuring reassortment in the context of host
146 evolutionary distance. Our study provides strong quantitative evidence supporting the hypothesis

147 that reticulate evolutionary processes are advantageous relative to adaptation by drift for
148 pathogen transfer between host species, and therefore more broadly, ecological niche switching.

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

162 In summary, using data available from a model zoonotic viral pathogen, we have shown that a
163 reticulate evolutionary strategies are important in enabling pathogen host switches. For the
164 influenza virus, reticulate evolution predominates when crossing between hosts. More broadly,
165 the greater the quantitative difference between ecological niches, the greater the importance of
166 reticulate evolutionary processes in enabling niche switches. While the quantitative importance
167 of reticulate evolution may differ for different organisms evolving in different niches, we expect
168 that further sequencing efforts from across broad domains of microbial life, and a further
169 characterization and definition of their ecological niches, will elucidate whether this principle

170 holds more broadly. Beyond its relevance to evolutionary ecology, reticulate evolution also has
171 public health consequences. Reassortant influenza viruses have been implicated in all past
172 human pandemic strains (*22-25*), and the ancestry of HIV-1 involved a hybrid SIV (*26*). Hence,
173 knowing how reticulate events shape disease emergence may help the ecology and evolution of
174 infectious disease become a more predictive science, leading to insight important to disease
175 prevention and mitigation (*27*).

**Acknowledgments**

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

## Figures

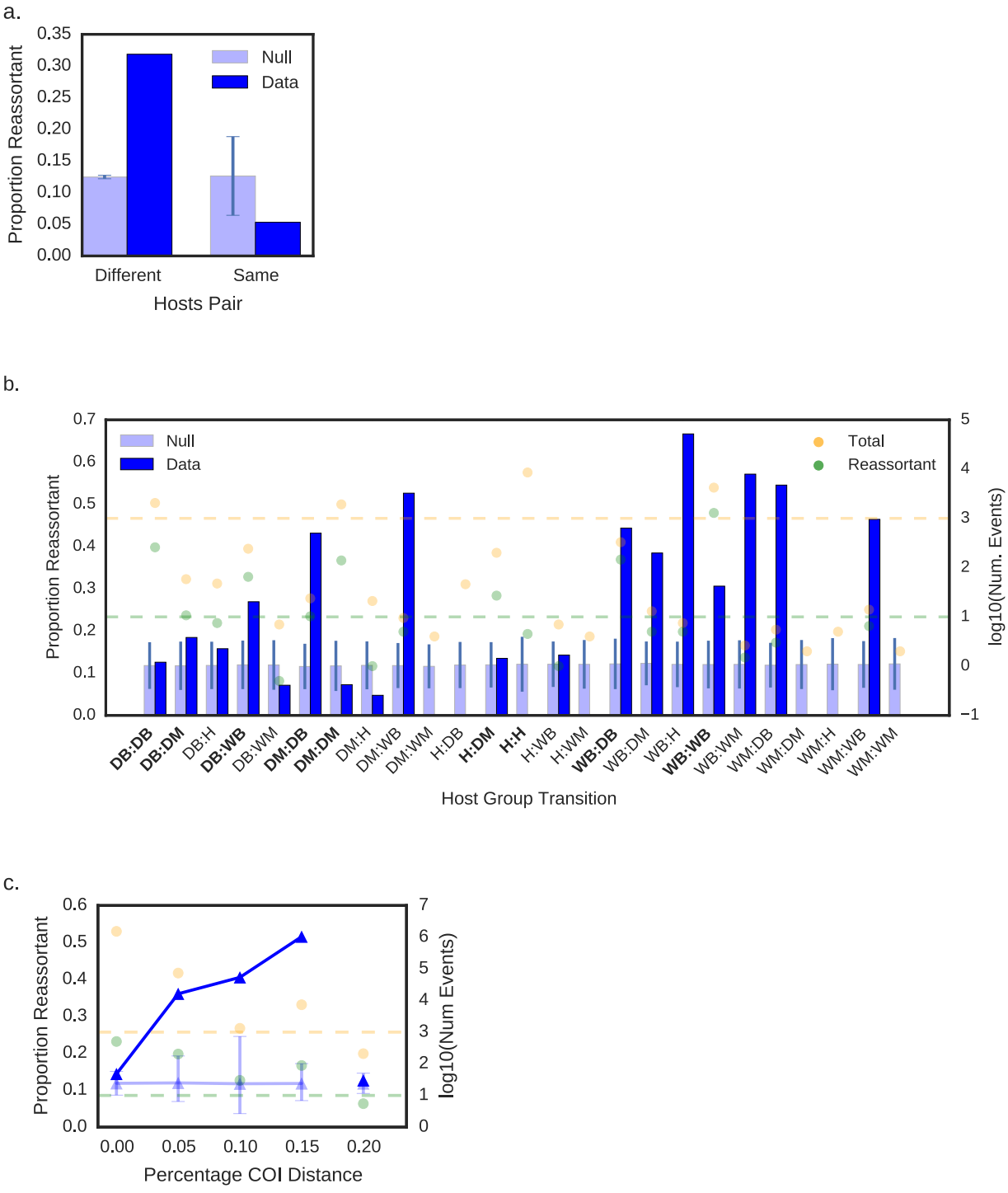


Figure 1. **Reassortment is over-represented relative to clonal descent in transmission across host barriers.** Proportion of reassortment events when crossing between (a) different or same hosts, (b) different host groups, and (c) hosts of differing evolutionary distance as measured by

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

## Materials and Methods

**Sequence data source and preprocessing.** Sequence and metadata was downloaded from the Influenza Research Database on 2 September 2015. Search parameters on the IRD were as such: Segment/Nucleotide data, Virus Type A, All Segments, All Hosts, All Geographic Groupings, Complete Genomes Only, Include pH1N1 sequences, Date Range 1980-Present. Advanced options included: Laboratory Strains excluded. Segment FASTA files were downloaded, with a Custom Format where only Accession Numbers were included.

**Source code and Data.** Digital Object Identifiers (DOIs) hosted on zenodo.org can be found for the following source code.

- Source code for graph construction: 10.5281/zenodo.33421.
- Source code for analysis, derivative data (network node and edge lists, computed threshold values) and figure construction as part of a series of Jupyter notebooks: 10.5281/zenodo.33422.
- Source code for the patristic distance tests: 10.5281/zenodo.33425.
- Source code and Jupyter notebooks for simulation studies: 10.5281/zenodo.33427

**Reassortant virus identification.** We adapted the SeqTrack algorithm (*14*) to perform graph construction. Sequences were aligned using Clustal Omega 1.2.1 (*28*), and the resultant distance matrix was converted into a similarity matrix by taking $1 - distance$. Affinity Propagation (*29*) clustering was done on each segment's similarity matrix to determine a threshold cut-off similarity value, defined as the minimum of minimum in-cluster pairwise identities. Because the Affinity Propagation algorithm does not scale well with sample size, we treated the threshold

221 computation as an estimation problem, computed as the median threshold of 50 random sub-
222 samples of 500 isolates.

223 We then thresholded each segment's similarity matrix based on its segment's threshold value,
224 summed all 8 thresholded similarity matrices, and then for each isolate, we identified the most
225 similar isolate that occurred prior to it in time. This yielded the initial "full complement" graph
226 without reassortant viruses. Each edge in this graph has an attached PWI, which is the sum of
227 PWIs across all 8 segments. Within this graph, there are isolates for which no "full complement"
228 of segments could be identified, which are candidate reassortant viruses. Additionally, amongst
229 the isolates for which a full complement of segments could be found from another source, we
230 identified those whose in-edges were weighted at the bottom 10% of all edges present in the
231 graph, which we also conservatively identified as candidate reassortant viruses. For these
232 viruses, we performed source pair searches, where we identified sources for a part of the genome
233 from one virus and sources for the complementary part of the genome from another virus. If the
234 summed PWI across the segments for the two viruses was greater than the single-source search,
235 we accepted the source pair as the candidate reassortant.

236 **Proportion Reassortment Calculations.** The proportion of reassortment events was calculated
237 by first weighting each incoming edge to every virus. The weighting procedure is described here:
238 If the virus is detected to be plausibly descended from n other viruses, as determined by maximal
239 similarity, it is given a weight of $\frac{1}{n}$. If the virus is detected to be a reassortant, then edges are
240 weighted by the the fraction of times that it is involved in a max similarity source pair. For
241 example, if a given node A has a plausible source of segments in B, C, D, with (B and C) and (B
242 and D) being plausible sources, then the edge B-A would be given a weight of 0.5, and the edge
243 C-A and D-A would be given a weight of 0.25 each.

244 **Permutation Tests.** Null models were constructed by permuting node labels; equal class size
245 permutation was performed for host species (for Figure 1a) and host group (for Figure 1b). For
246 example, if there were 9 'human', 4 'swine' and 2 'chicken' nodes (15 total), labels would be
247 randomly shuffled amongst the nodes such that each label were equally represented (5 each).

248 **Host Group Labelling.** Host species were manually classified into the "human", "domestic
249 animal" and "wild animal" groups, based on the country of isolation. For example, "ducks"
250 would be considered a "wild animal" in North America, while it would be considered a
251 "domestic animal" in East Asian countries. Ambiguous host species, while remaining in the
252 dataset, were excluded from the analysis.

253 **Host Evolutionary Distance.** Host species' scientific names were sourced from the Tree of Life
254 database (www.tolweb.org). Only host species with unambiguous scientific names recorded were
255 considered. Cytochrome oxidase I genes were sourced from the Barcode of Life Database
256 (www.boldsystems.org) on 31 October 2015. Sequences had to be at least 600 n.t. long to be
257 considered, and only positions with fewer than 3 gap characters were concatenated into the final
258 trimmed alignment. Evolutionary distance was computed from the trimmed alignment as the
259 proportion of mismatched nucleotides. Further details are available in the Jupyter notebooks.

260 **Simulation Studies.** In our simulations, we sought to model the process of obtaining sequences
261 as simply as possible. Therefore, we used a kinetic Monte Carlo-style process to simulate the
262 generation of new viral sequences under the processes of replication, mutation and reassortment.
263 Briefly, we simulated simple two-segment model viruses, with each of the two segments having
264 a different substitution rate. Each simulation run was initialized with anywhere between 1 and 5
265 viruses. At each time step, one virus was chosen at random to replicate (with 0.75 probability), or
266 reassort with another virus (with 0.25 probability). Simulations were run for 50 time steps.

267 Regardless of replication or reassortment, the progeny virus is subjected to mutations, with the
268 number of mutations in each segment being drawn from a Binomial distribution with probability
269 equal to the segment's substitution rate, and the exact positions drawn uniformly across the
270 segment. Specific implementation is viewable in the IPython HTML notebooks and Python class
271 definitions available on Github, available at the URL provided in Code and Data Deposition.
272 This simulation process thus gives rise to a fully known ground-truth graph, which all
273 reconstructions can be compared against.

274 The number of unique starting genotypes and total number of viral isolates being considered was
275 much smaller than the real-world data. Therefore, our graph reconstruction procedure captured
276 the essential parts of the method employed in the global analysis, but differed in the details.
277 Here, "full complements" involve only two segments. We did not perform affinity propagation
278 clustering as we started with completely randomly generated sequences of equal length. Our
279 "null model" graph is where source isolates are chosen uniformly at random from the set of
280 nodes occurring prior to the sink isolates.

281 In order to assess the accuracy of our reconstruction, we defined the path accuracy, and
282 reassortant path identification accuracy metrics. Edge accuracy, which is not used for evaluation
283 here, is whether a particular reconstruction transmission between two isolates exists in the
284 simulation. Path accuracy is a generalization of edge accuracy, where a path existing between the
285 source and sink nodes (without considering the direction of edges) in the reconstruction is
286 sufficient for being considered accurate. Reassortant path identification accuracy measures how
287 accurately we identified the reassortant paths, analogous to the regular path accuracy.

288 **Phylogenetic Reconstruction and Patristic Distance Comparison.** Phylogenetic
289 reconstruction was done for a subset of H3N8 viruses isolated from Minto Flats, AK, between
290 2009 and 2010 as part of a separate study. Briefly, each segment of the viral genomes were
291 aligned using Clustal Omega (*28*), and the genealogies reconstructed using BEAST 1.8.0 (*30*). A
292 minimum of 3 MCMC runs that converged on a single optimal tree were chosen to compute the
293 maximum clade credibility (MCC) tree. Burn-in ranged from 10 to 39 million steps out of 40,
294 with median 24 million steps. Patristic distances were calculated using the DendroPy package
295 (*31*). In the graph reconstruction on the Minto Flats study, we extracted the edges and nodes
296 involving only the H3N8 isolates, and computed the tree patristic distances between isolate pairs
297 linked by an edge in the graph.

298 **References**


299 1. Hernández-López A, et al. (2013) To tree or not to tree? Genome-wide quantification of
300 recombination and reticulate evolution during the diversification of strict intracellular
301 bacteria. *Genome Biol Evol* 5(12):2305–2317.

302 2. Peris D, et al. (2014) Population structure and reticulate evolution of Saccharomyces
303 eubayanus and its lager-brewing hybrids. *Molecular Ecology* 23(8):2031–2045.

304 3. Garten RJ, et al. (2009) Antigenic and genetic characteristics of swine-origin 2009
305 A(H1N1) influenza viruses circulating in humans. *Science* 325(5937):197–201.

306 4. Smillie CS, et al. (2011) Ecology drives a global network of gene exchange connecting the
307 human microbiome. *Nature* 480(7376):241–244.

308 5. Antonenka U, Nölting C, Heesemann J, Rakin A (2005) Horizontal transfer of Yersinia
309 high-pathogenicity island by the conjugative RP4 attB target-presenting shuttle plasmid.
310 *Mol Microbiol* 57(3):727–734.

311 6. Remold SK, Rambaut A, Turner PE (2008) Evolutionary genomics of host adaptation in
312 vesicular stomatitis virus. *Mol Biol Evol* 25(6):1138–1147.

313 7. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and
314 ecology of influenza A viruses. *Microbiol Rev* 56(1):152–179.

315 8. Duggal NK, Emerman M (2012) Evolutionary conflicts between viruses and restriction
316 factors shape immunity. *Nat Rev Immunol* 12(10):687–695.

317 9. Li C, et al. (2010) Reassortment between avian H5N1 and human H3N2 influenza viruses
318 creates hybrid viruses with substantial virulence. *Proc Natl Acad Sci USA* 107(10):4687–
319 4692.

320 10. Mehle A, Dugan VG, Taubenberger JK, Doudna JA (2012) Reassortment and mutation of
321 the avian influenza virus polymerase PA subunit overcome species barriers. *J Virol*
322 86(3):1750–1757.

323 11. Lam TT-Y, et al. (2011) Reassortment Events among Swine Influenza A Viruses in
324 China: Implications for the Origin of the 2009 Influenza Pandemic. *J Virol* 85(19):10279–
325 10285.

326 12. Tao H, Steel J, Lowen AC (2014) Intra-host dynamics of influenza virus reassortment. *J*
327 *Virol* 88(13):JVI.00715–14–7492.

328 13. Squires RB, et al. (2012) Influenza Research Database: an integrated bioinformatics
329 resource for influenza research and surveillance. *Influenza and Other Respiratory Viruses*
330 6(6):404–416.

331 14. Jombart T, Eggo RM, Dodd PJ, Balloux F (2011) Reconstructing disease outbreaks from

332 genetic data: a graph approach. *Heredity* 106(2):383–390.

333 15. Nelson MI, Vincent AL (2015) Reverse zoonosis of influenza to swine: new perspectives
334 on the human-animal interface. *Trends Microbiol* 23(3):142–153.

335 16. Ma W, Kahn RE, Richt JA (2008) The pig as a mixing vessel for influenza viruses:
336 Human and veterinary implications. *J Mol Genet Med* 3(1):158–166.

337 17. Ratnasingham S, Herbert PDN (2007) bold: The Barcode of Life Data System
338 (http://www.barcodinglife.org). *Mol Ecol Notes* 7(3):355–364.

339 18. Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the
340 details: genetic variation in introduced populations and its contributions to invasion.
341 *Molecular Ecology* 24(9):2095–2111.

342 19. Molofsky J, Keller SR, Lavergne S, Kaproth MA, Eppinga MB (2014) Human-aided
343 admixture may fuel ecosystem transformation during biological invasions: theoretical and
344 experimental evidence. *Ecol Evol* 4(7):899–910.

345 20. Verhoeven KJF, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological
346 invasions and the balance between local adaptation and inbreeding depression. *Proc R Soc*
347 *B* 278(1702):2–8.

348 21. Keller SR, Fields PD, Berardi AE, Taylor DR (2014) Recent admixture generates
349 heterozygosity–fitness correlations during the range expansion of an invading species.
350 *Journal of Evolutionary Biology* 27(3):616–627.

351 22. Wu A, et al. (2013) Sequential Reassortments Underlie Diverse Influenza H7N9
352 Genotypes in China. *Cell Host Microbe* 14(4):446–452.

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

355 24. Lam TT-Y, et al. (2013) The genesis and source of the H7N9 influenza viruses causing
356 human infections in China. *Nature* 502(7470):241–244.

357 25. Gao R, et al. (2013) Human infection with a novel avian-origin influenza A (H7N9) virus.
358 *N Engl J Med* 368(20):1888–1897.

359 26. Bailes E, et al. (2003) Hybrid origin of SIV in chimpanzees. *Science* 300(5626):1713–
360 1713.

361 27. Wasik BR, Turner PE (2013) On the Biological Success of Viruses.
362 *http://dxdoiorg/101146/annurev-micro-090110-102833* 67(1):519–541.

363 28. Sievers F, Dineen D, Wilm A, Higgins DG (2013) Making automated multiple alignments
364 of very large numbers of protein sequences. *Bioinformatics* 29(8):989–995.

29. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science* 315(5814):972–976.

30. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

31. Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26(12):1569–1571.