Multi-Strain Vaccination Strategies on Viral Genotype Networks

P. Alex Burnham
University of Vermont,
Department of Biology
Burlington, Vermont
pburnham@uvm.edu

Thomas S. O'Leary
University of Vermont,
Department of Biology
Burlington, Vermont
tsoleary@uvm.com

Blake J. M. Williams
University of Vermont,
Vermont Complex Systems Center
Burlington, Vermont
bwilli16@uvm.edu

ABSTRACT

We implement a novel approach for vaccine strain selection based on a genotype network of viral strains. Current strategies for selecting vaccine strains of multi-strain pathogens involve present and forecasted incidence of particular strains. Here we emphasize the effects of transcending immunity, and exploit the genetic similarity between strains to determine optimal strategies in the case of multi-strain vaccination. We employ a genetic algorithm (GA) to find optimal strategies in the $\binom{N}{k}$ search space of k vaccines on N strains, seeking to reduce the number of strains that may be reached in an outbreak. We tested the strategy on toy networks of varying size and structure, before searching optimal strategies for multiple real-world *Influenza A* (H3N2) genotype networks. This approach consistently reduced the mean expected outbreak size, with significant improvements on random searches. Evolved solutions were evaluated on *Influenza A virus* networks that grew beyond the time of solution computation, simulating the 6 month delay between strain selection and distribution. Despite ignorance toward future states of the genotype network, GA-evolved strategies consistently outperformed even the best random solutions after a year of novel strain emergence. Our approach suggests that knowledge of the genotype network can provide useful insight for vaccine strain selection.

KEYWORDS

network, genetic algorithm, vaccine, influenza

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
P. Alex Burnham, Thomas S. O'Leary, and Blake J. M. Williams. 2019. Multi-Strain Vaccination Strategies on Viral Genotype Networks. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

High mutation rates in RNA viruses such as *Zaire ebolavirus* [2], *Influenza A virus* [7], and *Rabies lyssavirus* [15] lead to numerous contemporaneous strains [16]. Vaccines are developed based on the antigenic properties of such viruses, however vaccine effectiveness can be less than ideal: influenza vaccine efficacy has been approximately 40% since 2005 [3, 5, 6, 11, 18]. Effective vaccination is challenged by: (i) rapid evolution of viruses away from the antigenic properties of strain(s) used for vaccines [17], and (ii) properly selecting strains for vaccines such that antibodies have a wide-reaching effect on prevalent and future strains [4, 9].

Here, we address the problem of selecting vaccination strains that provide antibody maximal coverage, in the case where multiple vaccination strains may be used. This problem has time complexity $O\binom{N}{k}$ for N strains in the population and k chosen vaccination strains. For large enough N and even modest increments in k, the time to brute-force an optimal combination of vaccination strains could be infeasible, especially with the use of in-depth modeling with compartmental or agent-based models, let alone laboratory viral inhibition assays.

Each spring and fall, the World Heath Organization (WHO) makes recommendations for specific strains to be included in the influenza vaccine for each hemisphere. WHO bases their recommendations largely on the current and forecasted incidence of a particular strain in the upcoming flu season, as well as the availability of similar vaccine viruses [1]. Although some attention is given to the genetic similarity between strains by incorporating phylogenetic analysis, the WHO might not fully exploit the information contained within genotype networks and complementary network analyses to inform their recommendations. While the WHO typically only recommends one or two vaccine strains per subtype of influenza, we explore the implementation of multiple

vaccine strains ($k \geq 3$) per genotype network, given the history of poor vaccine efficacy. Our approach suggests that choosing multiple strains based on knowledge of the network structure can greatly increase the efficacy of a vaccine.

We developed an approximation of vaccine efficacy through suppression of outbreak potential in the presence of vaccinated strains. Transcending effects of immunity, observed in viruses such as influenza [14], allow for genetically similar strains to be influenced by nearby vaccines. A genotype network was used to model the genetic similarity between strains, allowing for real-world and simulated network structure to be evaluated.

In this paper we implement a genetic algorithm to find ideal vaccination strains for a given genotype network. In Section 2 we discuss the details of the GA implementation, including solution representation, fitness evaluation, and the use of genotype networks. In Section 3, we first test this approach on a series of simple toy networks and a small Erdős-Rényi random graph to provide a clear understanding of how the vaccination strategy evolves on relatively simple network structures. We then apply the GA to a series of influenza A H3N2 genotype networks of ranging in size and complexity from size 81 to 1430, to test the approach on complex and large real world genotype networks. Finally, we evaluated GA-evolved and random vaccination strategies on an influenza network that is growing through the addition of novel strains arising via mutation over time, to simulate the lag in time between the selection of the vaccination strain and the end of a flu season.

2 METHODS

Genotype Network

A given set of strains are related to one another through a genotype network. Each node in this network corresponds to a unique gene or protein sequence (defined as a strain), with edges existing between strains whose sequences differ by one base pair or amino acid (indicating a plausible mutation pathway). In this paper, sequences will be assumed to be the amino acids of a specified antigenic protein. In the real-world application, this will be the hemagglutinin (HA) surface protein of *influenza A*, H3N2.

Outbreak Fitness Function

In epidemic models the basic reproductive number $R_0 = \frac{\beta}{\lambda}$, where β is the number of new cases generated by a case in time step t, and $\frac{1}{\gamma}$ is the mean time steps of infectivity for a case. In an infinite population, R_0 is the expected number of new infections each individual case will produce. For $R_0 < 1$ a disease is expected to die out, but for $R_0 > 1$ sustained transmission is expected; thus $R_0 = 1$ represents the epidemic threshold.

Here we define R_0^{eff} as the normalized effective R_0 after the effects of vaccination, such that for strain i and set of vaccination strains V:

$$R_0^{eff}(i) = \begin{cases} 1 & \text{if } V = \emptyset \\ 0 & \text{if } i \in V \\ \prod_{v \in V} (1 - e^{-x_{iv}/\delta}) & \text{otherwise} \end{cases} \quad (1)$$

where x_{iv} is the genetic distance between strains i, v, and δ is the tunable transcendence of immunity parameter. x_{iv} is determined from the shortest path in the network, which was observed to closely approximate genetic distance in real-world influenza networks (in the evaluation of fitness on a growing network component in Section 3, we allow the final network distances to be used in the calculation of an incomplete network). R_0^{eff} equals 1 in the absence of any vaccines, but is reduced to 0 for directly vaccinated strains, and otherwise equals the product of immunity that transcends vaccinated strains as a log decaying function of genetic distance.

We define R_0^{crit} as the normalized epidemic threshold, constrained to $(0, 1)$ for all $R_0 > 1$:

$$R_0^{crit} = \frac{1}{R_0} \text{ , for } R_0 > 1 \quad (2)$$

In this paper we let $R_0 = 2$, a value comparable to that of Ebola and pandemic influenza, such that $R_0^{crit} = \frac{1}{2}$.

The fitness F for a given set of vaccination strains V on network G is found by: (i) removing subcritical strains ($R_0^{eff}(i) < R_0^{crit}$), which potentially (and ideally) fragments the network into multiple components, then (ii) computing the mean component size for each strain i:

$$F(V, G) = \frac{\sum_j (j_n)^2}{G_n^2} \text{ for component size } j_n, \text{ network size } G_n \quad (3)$$

Thus $F(V, G)$ is the expected number of strains an outbreak can reach, through known strains: it is the expected component size of an outbreak at a random strain. Minimizing this value will reduce the number of known strains an outbreak will reach, and necessitate evolutionary detours around vaccinated regions of genotype space were the virus to connect to other known components.

GA-Evolved Vaccination Strategies

Here we implement a near-canonical GA. Each solution, or vaccination strategy, exist as vector V, whose length equals the number of vaccination strains. V contains the indices of the nodes (strains) to be vaccinated, with values from 1 to network size N.

For a given network, a population of P random solutions is initialized. For up to N_{gen} repetitions, the population is

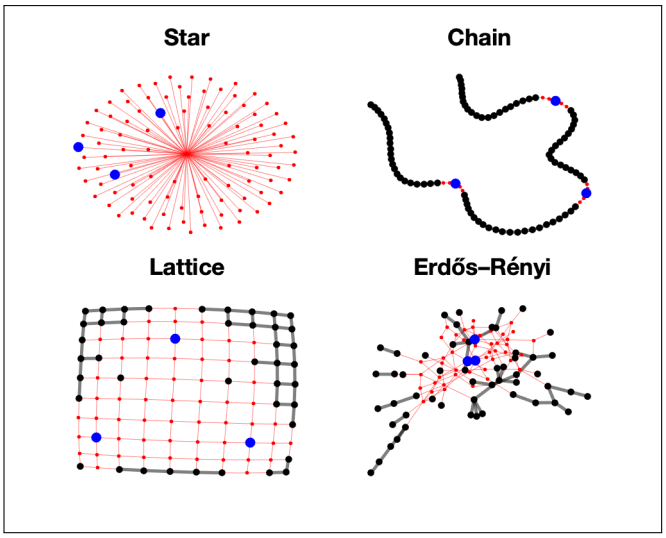

Figure 1: Representative vaccination strategy solutions for toy networks. Blue nodes represent strains included in the vaccination. Red nodes indicate nodes that are below the critical threshold for an outbreak. Black nodes are above that threshold.

evolved through parent selection based on fitness, crossover, and mutation. Parents are selected through tournament selection with tournament size T_n. Parents are then recombined via single-point crossover with probability P_c. Indices within each solution are then mutated to a random value from 1 to N according to probability P_m. The best solution at each time step is noted, with the GA exiting before N_{gen} reps if the absolute minimum fitness $F(V, G) = 0$ is found.

Experimental Design

Our investigation is three-part: (i) evolving solutions on toy networks, to understand the effects of network structure on solutions, (ii) evolving solution on real-world genotype networks, and (iii) evaluating decay of fitness on a growing network.

In the first part, we constructed the toy networks consisting of a star, lattice, and chain network of size $N = 100$, as well as an Erdős Rényi random network of size $N = 100$, existing as the giant component of a $G(N, p) = G(110, 0.025)$ graph (Figure 1). For 20 repetitions, we ran a GA on each toy network according to the parameters in Table 1. The GA exited when a perfect solution was found ($F(V, G) = 0$) or upon reaching N_{gen} generations. The GA solutions were compared to a distribution of 10^3 random solutions.

In the second part we evolved solutions on a series of real-world influenza A H3N2 genotype networks. These networks were constructed from amino acid sequences of HA observed globally January 2000 through May 2019, sourced from the Influenza Research Database [19], in which sequences are represented as nodes and edges exist between sequences differing at one amino acid — indicating a plausible mutation pathway. The real-world networks represent 9 components selected from this network to give a distribution of network sizes from $N = 20$ to $N = 1430$. For both 3 and 4 vaccination strategies, the GA was run 20 times for each network, for 3 values of transcendence (δ =[1,2,3]) and the parameters found in Table 2. The GA solutions were compared to a distribution of 10^3 random solutions.

In the third part we evaluated changes in fitness as a network grows beyond the time at which a solution was evolved. This simulates vaccination strategies evolved on present strains prior to the emergence of novel strains, at which point fitness may be reduced as the genotype network has grown. Solutions were evolved on a subset of a genotype network of size $N = 791$. The first half of the network to appear ($N = 384$, approximated to the nearest day at which 50% of nodes exist) is used to evolve solutions according to Table 3 across 20 reps. Note that fitness calculations were given knowledge of the full network for accurate genetic distance values. Fitness values were then found for these solutions on the network after 3, 6, and 12 months, as well as for a distribution of 10^3 random networks.

Statistical Analysis

To examine the fitness differences between random solutions and GA-derived solutions across the different transcendence values and network sizes, and to analyze the number of function calls required by the different parameter sets, we conducted a series of ANOVAs for each section of our three-part experimental design. For the toy networks, influenza networks, and growing influenza network, we structured our model to examine fitness by group (GA-evolved or random) and the main and interaction effects between network size and transcendence values. To examine the effect of the transcendence value, network size, and their interaction on the number of function calls for the same data sets, we employed an additional three models. In the third part of the study, we examined how random and GA-derived solutions change in fitness over time as the network grows by modeling fitness as a function of group (GA-evolved or random), days after vaccine selection, and their interaction effect. Additionally, we show how the exponential scaling in the number of function calls increases for the size of the network and the number of nodes vaccinated. All analyses were conducted in the R statistical programming language [13].

3 RESULTS

The GA was consistently able to derive useful solutions for a combination of different network structures, network sizes and transcendence values (Figures 1-3). For the toy networks,

Table 1: GA parameters

GA Parameter	Symbol	Toy Nets	Real Nets	Temporal Net
Population size	P	300	300	200
Genome length (# vaccine strains)	V	3	[3,4]	4
Mutation rate	P_m	1/V	1/V	1/V
Crossover probability	P_c	0.2	0.2	0.2
Max generations	N_{gen}	50	50	20
Tournament size	T_n	2	2	2
Network size (# strains)	N	100	20-1430	384 $\longrightarrow$ 791
Epidemic threshold	R_0^{crit}	0.5	0.5	0.5
Transcendence	δ	1	[1,2,3]	1

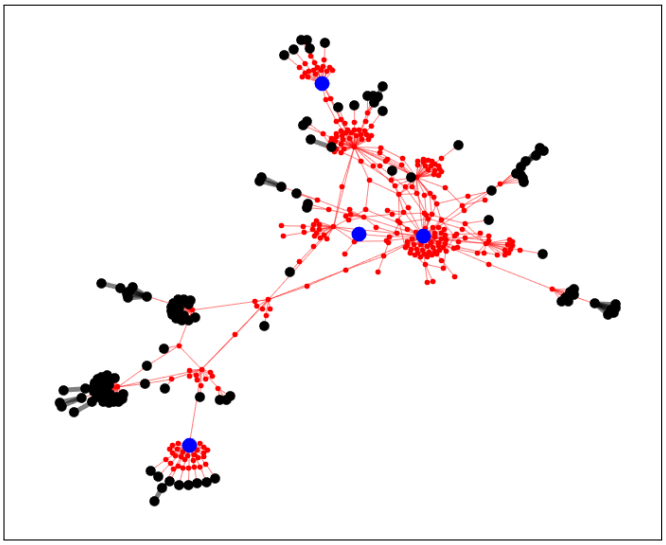

Figure 2: Representative vaccination strategy for a moderately sized real flu genotype network ($N = 400$). Blue nodes represent strains included in the vaccination strategy. Red nodes indicate nodes that are below the critical threshold for an outbreak. Black nodes are above that threshold.

vaccination strategies selected by the GA showed in a manageable setting how the algorithm took advantage of simple structures to minimize super-critical nodes (Figure 1). There was a difference in the number of function calls required to find a solution between network types, but that was driven by the star network which only needed an average of 200 function evaluations to find a perfect solution (Figure 3A). We found in our experimental runs that the GA solutions performed significantly better than the random solutions in terms of fitness ($p < 0.00001$). On average, different networks structures performed differently depending on the transcendence value used ($p < 0.00001$) (Figure 3B).

For real networks evolved for both 3 and 4 vaccinations, useful strategies were discovered by the GA. A representative example is shown in Figure 2. There was a significant difference in the number of function evaluations depending on the network size and transcendence value with of the smaller networks requiring fewer calls (p<0.00001) (Figure 3C and 3E). In terms of fitness, on both 3 and 4 vaccination strategies, the GA performed significantly better than the random solutions (p<0.00001) (Figure 3D and 3F). Again, on average, different networks sizes performed differently depending on the transcendence value used with large networks with low transcendence performing the worst ($p < 0.00001$).

We found that when random and GA solutions were evolved on a portion of a large example genotype network and the network was allowed to grow, the GA solutions performed significantly better than random ones (p<0.00001) (Figure 4). However, both solutions slowly worsened through time as the network grew (p<0.00001). No interaction was observed between time and how the solutions were derived suggesting that both solutions decayed at a similar rate ($p = 0.955$).

The GA was able to find successful solutions on the real networks with linear scaling in the number of function evaluations required to find a workable solution. Figure 5 shows the size of the search space for 1, 2, 3 and 4 vaccine strategies on a log_{10} scale. The GA search effort for 4 vaccines, shown as black points, falls in-between the search space of 1 vaccine and 2 vaccines. For the largest real network (N=1430), to search the entire search space for 4 vaccination strategies, $1.74 * 10^{11}$ function evaluations would be required. At the 0.02 seconds it takes for one evaluation, it would take 107.8 years to search the entire network. The GA only performed $3.8 * 10^3$ evaluations on average and found near-perfect ($F = 0$) solutions for transcendence values of $\delta = [2, 3]$.

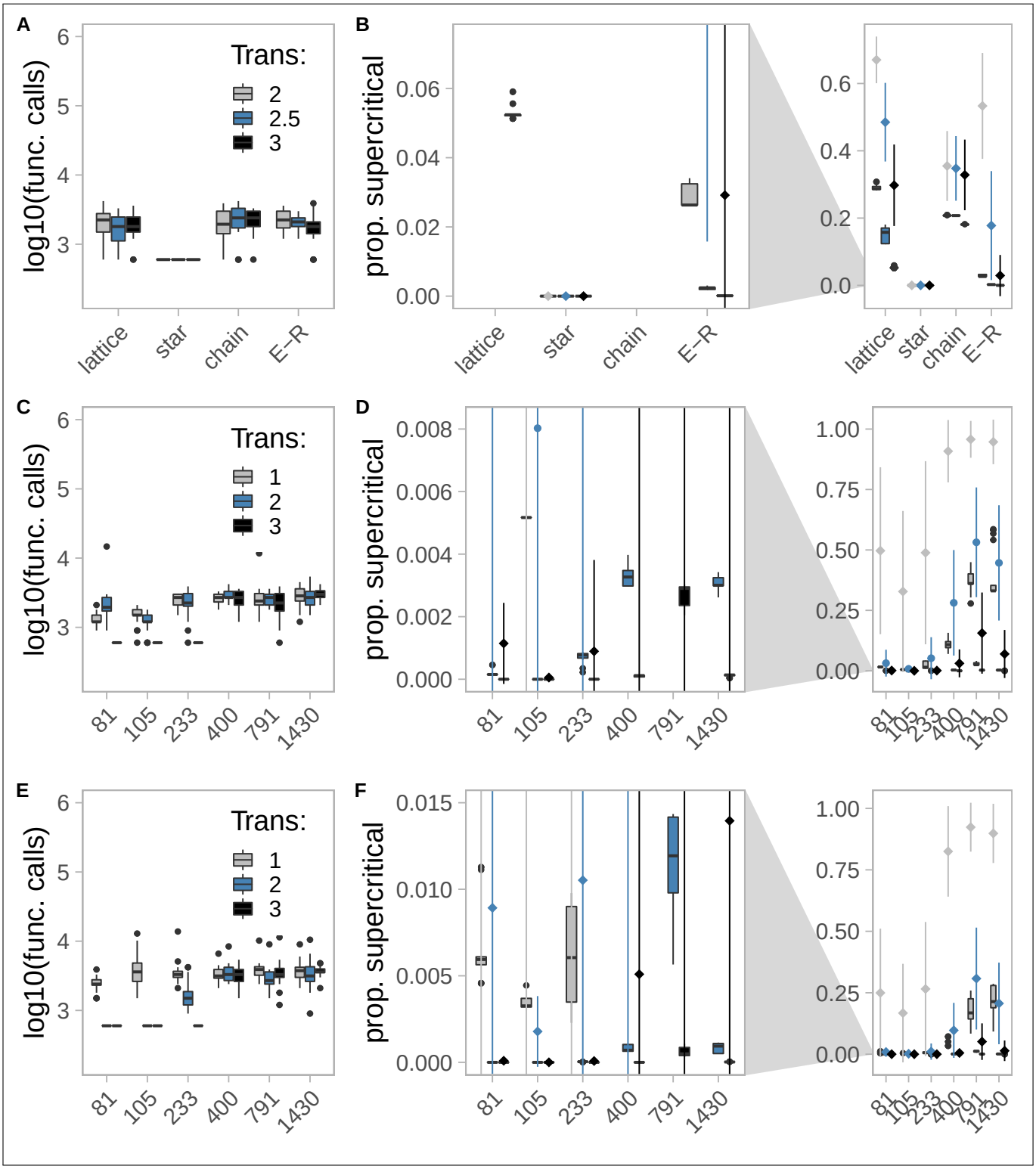

Figure 3: (Left most column) The number of function calls (computational effort) on a log scale for each network and transcendence value. (Right most column) The proportion of super-critical nodes to total nodes (fitness) by network for three transcendence values with random solutions with mean shown as diamonds ± 1 standard deviation. Colors grey, blue and black refer to small, medium, and high transcendence values respectively. The middle column is an expanded panel that shows the variance in the smallest distributions of solutions. A and B refer to the toy networks (lattice, star, chain and Erdős-Rényi) utilizing a vaccination strategy of three vaccines. C and D refer to the real networks from size 81 to 1430 with a vaccine strategy of three and E and F refer to a strategy of four on real networks of size 81 to 1430.

4 DISCUSSION

Network Structure and Strategies

Network structure heavily influences both optimal fitness and location of vaccination strains within a genotype network. Fitness measured by the ability of the vaccination strategy to fragment the network into small components allowed for minimization of expected outbreak size (by strain access) in the known genotype space. Thus solutions are rewarded for their ability to not only remove nodes from the network, but to fragment the remaining components. This is seen in the toy networks of Figure 1. The chain is broken into 4 nearly if not exactly evenly sized components, minimizing the mean expected outbreak size.

A comparison of the star and the chain indicate the effects of network diameter. Networks of small diameter allow more nodes to fall within the radius of sub-critical influence for a vaccine strain. Although the star and chain are of the same number of nodes, the star's small diameter allows many (if not all) vaccination strategies to provide complete coverage of the known genotype space, indicating that no outbreak would occur. These star-like hubs are found in the influenza networks, whose degree correlates with duplicate samples of a sequence (i.e. greater incidence). Hubs may indicate a particularly virulent or novel strain, yet one whose vaccine would cover a large number of strains, and thus be a target for vaccines. Indeed, hubs were important building blocks for solutions to the influenza networks. However, reducing total super-critical strains is only one way to reduce outbreak size. Fragmentation of the network into smaller components reduces mean outbreak size, and in larger networks, may only be achievable to significant effect through cooperation between vaccine strains.

The lattice and Erdős-Rényi random graphs demonstrate the cooperation between vaccines, in which only through their combined effect: (i) do some nodes become sub-critical, and (ii) may the network be split at multi-node bridges between large components. The former effect models the case where immunity to multiple strains have a multiplicative or additive effect, or immunity not determined by just the nearest strain. This is an assumption of the model, that the influence from multiple vaccines has a multiplicative effect on immunity, which may be more optimistic than what would be found in transcendence in a real-world application (in comparison to using the maximum immunity, or another interaction between them). The latter effect of cooperation, vaccination at multi-node bridges, implements what may be an important control mechanism on genotype space: blocking evolutionary routes between large or virulent regions of potential protein structures. Deep mutational scanning, which predicts protein stability, could evaluate the effectiveness of targeting these evolutionary bridges by indicating the presence or absence of other pathways between large regions of genotype space [12].

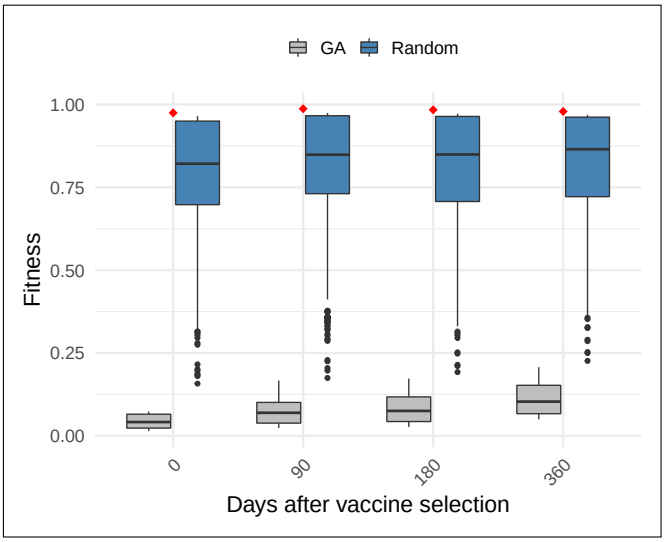

Figure 4: Distribution of the fitnesses from GA-evolved and random vaccination strategies on a growing *Influenza A* (H3N2) genotype network, 0, 90, 180, and 360 days after the vaccination strategy was selected.

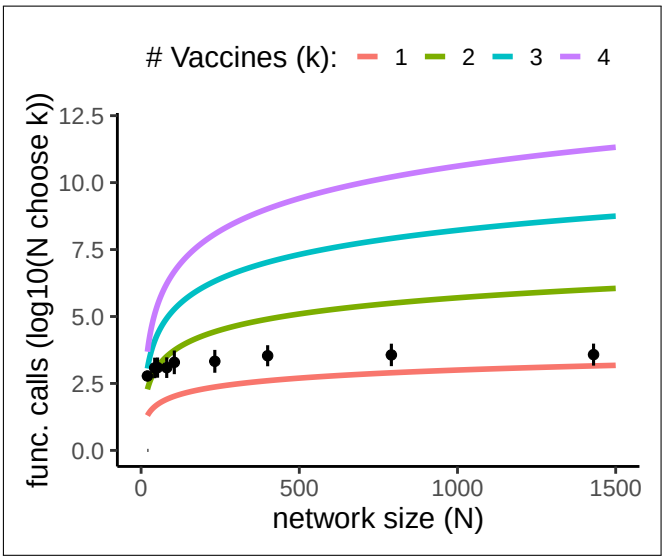

Figure 5: Theoretical number of fitness evaluations, $\binom{N}{k}$, for four different numbers of vaccines, k, shown within the size range of real genotype networks, N, (20 to 1430) shown on a log base 10 scale. The mean number of fitness evaluations for 4 vaccines (k=4) required to reach an optimal solution for each real network size are superimposed in black ± 1 standard error.

Cooperation between vaccine strain placement is crucial to fragment the network. For instance, a one-strain vaccine strategy on the lattice is optimized with an internal and central placement, while two or more strains must be placed such as to split the network in half, depending on the level of immunity transcendence. In Figure 1 we see the vaccine strains placed to not only make many nodes sub-critical but to isolate the peripheral regions, reducing mean component size. The Erdős-Rényi network shows a similar strategy: dense central regions remove numerous nodes, while optimal placement fragments the peripheral regions as much as possible.

Real-World Vaccination Strategies

Vaccination strategies on influenza networks exhibited the same behaviors seen on toy networks. Both 3- and 4-strain vaccination strategies frequently included hubs, while not exclusively using these high degree nodes to fragment the networks. Figure 2 shows a 4-strain strategy on an influenza A (H3N2) HA network of size $N = 400$, that included 3 hubs, while also utilizing a low-degree node to separate the lower-left region component from the upper-right. For transcending immunity levels of $\delta = [1, 2, 3]$ and 3 to 4 vaccine strains, GA-evolved solutions consistently performed better than random solutions ($p < 0.00001$). This demonstrates the superiority of the GA for multi-strain vaccine implementation.

The function calls of the GA scaled well with both network size and number of vaccination strains, in addition to tolerating variation in transcendence of immunity (Figure 3, leftmost column). This is in contrast to the $O\binom{n}{k}$ time complexity of exhaustively searching solutions, visualized in Figure 4. This indicates that a simple GA implementation can sufficiently find low-fitness solutions for large search spaces.

Evolved Strategies Tolerate Network Growth

GA-evolved vaccine strategies suffered no excess fitness losses relative to random strategies on a growing network. This contradicted our suspicion that random strategies could be more resilient to evolved strategies as novel strains emerged in the genotype network, if their location became more optimal as the network grew. Instead, we see no such advantage in random solutions, as even the best random solutions worsened in time (Figure 4). The insignificance of the strategy-by-date interaction (p=0.955) indicates no reduced fitness decay in random strategies. Combined with initial superiority, evolved solutions retain the best fitness values with modest increases for 12 months (Figure 4) and beyond. Fitness evaluations beyond 12 months post-solution evolution are not considered, since few strains in the initial portion of the network are likely to be prevalent (thus relevant for vaccine consideration).

Random solutions that improved in time were rare, and it is unlikely to find a random solution with both fitness comparable to GA-evolved solutions and improvements as the network grows. If a random solution were to be found that became better than GA-evolved solutions as the network grew, there would be no justification for its implementation given the unknown future of the structure network. GA-evolved solutions remain superior for coverage of future outbreaks.

Future Directions

The fitness function assumes immunity transcends as a logarithmic function of genetic distance between HA sequences of strains, which could be refined by: (i) a more data-driven selection of the transcendence function via HA inhibition assays, such as the experiments that have been conducted on the avian *Influenza A* H5N1 [14] subtype, and (ii) more closely approximating of how multiple acquired immunities combine to affect other strains (*e.g.* multiplicative or additive effects, if not more complex).

More information could be added to the network structure through weighting the edges by the similarity of the amino acid substitutions between nodes, by using an approach similar to BLOSUM [8]. This could be used to update the transcending immunity between genetically similar strains. Due to local optima observed within the fitness landscape, this GA approach could also be improved by implementing an algorithm that promotes diversity and decreases premature convergence, such as an Age-Layered Population Structure (ALPS) [10], to increase the likelihood that the global optima is found, as well as reducing the need for multiple restarts of the GA.

5 CONCLUSION

Here we identified the features of GA-evolved vaccination strategies on genotype networks and demonstrated their success in reducing expected outbreak size by number of strains. Our approach consistently identified efficacious solutions on a variety of different network structures, sizes, and transcendence values.

The location of vaccination strains within the network greatly influences the overall fitness of the vaccination strategy. A simple GA identifies these optimal vaccine strain selection strategies with considerably less effort than would a brute force search. The GA-evolved solutions were observed to be robust to network growth, resulting from mutations leading to novel strain emergence in real-world viral genotype networks. GA solutions consistently lead to better strategies than random search, across network size, number of vaccine strains, and parameter settings.

We call for investigations that address the following: (i) identification of the viable regions of genotype space, such as through deep mutational scanning, to allow for evolution of vaccination strategies that include future strains, (ii) refinement of the relationship between genetic similarity of viral strains and the transcendence of immunity, to better inform vaccine coverage, and (iii) evolutionary strategies of vaccine implementation that account for forecasting of active regions of genotype space.

ACKNOWLEDGMENTS

Special thanks to Dr. Maggie Eppstein for her instruction and guidance through the Evolutionary Computation course. We acknowledge Laurent Hébert-Dufresne (LHD) for his contributions to work leading to this project — to be listed as an author in submission outside of CSYS 352. BJMW and LHD acknowledge support from the National Institutes of Health 1P20 GM125498-01 Centers of Biomedical Research Excellence Award.

REFERENCES

[1] 2018. Selecting Viruses for the Seasonal Influenza Vaccine. https://www.cdc.gov/flu/prevent/vaccine-selection.htm{#}summary

[2] Kendra J Alfson, Gabriella Worwa, Ricardo Carrion, and Anthony Griffiths. 2016. Spontaneous Mutation Frequency Notes. *Journal of Virology* 90, 5 (2016), 2345–2355. https://doi.org/10.1128/JVI.02701-15.Editor

[3] Edward A. Belongia, Burney A. Kieke, James G. Donahue, Laura A. Coleman, Stephanie A. Irving, Jennifer K. Meece, Mary Vandermause, Stephen Lindstrom, Paul Gargiullo, and David K. Shay. 2011. Influenza vaccine effectiveness in Wisconsin during the 2007-08 season: Comparison of interim and final results. *Vaccine* 29, 38 (2011), 6558–6563. https://doi.org/10.1016/j.vaccine.2011.07.002

[4] F. Carrat and A. Flahault. 2007. Influenza vaccine: The challenge of antigenic drift. *Vaccine* 25, 39-40 (2007), 6852–6862. https://doi.org/10.1016/j.vaccine.2007.07.027

[5] Brendan Flannery, Rebecca J Garten Kondor, Jessie R Chung, Manjusha Gaglani, Michael Reis, Richard K Zimmerman, Mary Patricia Nowalk, Michael L Jackson, Lisa A Jackson, Arnold S Monto, Emily T Martin, Edward A Belongia, Huong Q McLean, Sara S Kim, Lenee Blanton, Krista Kniss, Alicia P Budd, Lynnette Brammer, Thomas J Stark, John R Barnes, David E Wentworth, Alicia M Fry, and Manish Patel. 2019. Spread of antigenically drifted influenza A(H3N2) viruses and vaccine effectiveness in the United States during the 2018-2019 season. *The Journal of infectious diseases* 30329 (2019), 1–8. https://doi.org/10.1093/infdis/jiz543

[6] Marie R. Griffin, Arnold S. Monto, Edward A. Belongia, John J. Treanor, Qingxia Chen, Jufu Chen, H. Keipp Talbot, Suzanne E. Ohmit, Laura A. Coleman, Gerry Lofthus, Joshua G. Petrie, Jennifer K. Meece, Caroline Breese Hall, John V. Williams, Paul Gargiullo, La Shondra Berman, and David K. Shay. 2011. Effectiveness of non-adjuvanted pandemic influenza a vaccines for preventing pandemic influenza acute respiratory illness visits in 4 U.S. communities. *PLoS ONE* 6, 8 (2011), 4–10. https://doi.org/10.1371/journal.pone.0023085

[7] Yi Guan, Dhanasekaran Vijaykrishna, Justin Bahl, Huachen Zhu, Jia Wang, and Gavin J.D. Smith. 2010. The emergence of pandemic influenza viruses. *Protein and Cell* 1, 1 (2010), 9–13. https://doi.org/10.1007/s13238-010-0008-z

[8] S. Henikoff and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* 89, 22 (1992), 10915–10919. https://doi.org/10.1073/pnas.89.22.10915

[9] Scott E. Hensley. 2014. Challenges of selecting seasonal influenza vaccine strains for humans with diverse pre-exposure histories. *Current Opinion in Virology* 8 (2014), 85–89. https://doi.org/10.1016/j.coviro.2014.07.007

[10] Gregory S. Hornby. 2006. ALPS: The age-layered population structure for reducing the problem of premature convergence. *GECCO 2006 - Genetic and Evolutionary Computation Conference* 1 (2006), 815–822.

[11] Michael L. Jackson, Jessie R. Chung, Lisa A. Jackson, C. Hallie Phillips, Joyce Benoit, Arnold S. Monto, Emily T. Martin, Edward A. Belongia, Huong Q. McLean, Manjusha Gaglani, Kempapura Murthy, Richard Zimmerman, Mary P. Nowalk, Alicia M. Fry, and Brendan Flannery. 2017. Influenza vaccine effectiveness in the United States during the 2015–2016 season. *New England Journal of Medicine* 377, 6 (2017), 534–543. https://doi.org/10.1056/NEJMoa1700153

[12] Juhye M. Lee, John Huddleston, Michael B. Doud, Kathryn A. Hooper, Nicholas C. Wu, Trevor Bedford, and Jesse D. Bloom. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences of the United States of America* 115, 35 (2018), E8276–E8285. https://doi.org/10.1073/pnas.1806133115

[13] R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[14] Thomas Rowe, Robert A. Abernathy, Jean Hu-Primmer, William W. Thompson, Xiuhua Lu, Wilina Lim, Keiji Fukuda, Nancy J. Cox, and Jacqueline M. Katz. 1999. Detection of antibody to avian influenza a (H5N1) virus in human serum by using a combination of serologic assays. *Journal of Clinical Microbiology* 37, 4 (1999), 937–943.

[15] Charles Rupprecht, Ivan Kuzmin, and Francois Meslin. 2017. Lyssaviruses and rabies: Current conundrums, concerns, contradictions and controversies. *F1000Research* 6, 0 (2017), 1–22. https://doi.org/10.12688/f1000research.10416.1

[16] R. Sanjuan, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw. 2010. Viral Mutation Rates. *Journal of Virology* 84, 19 (2010), 9733–9748. https://doi.org/10.1128/jvi.00694-10

[17] D. Steinhauer. 1987. Rapid Evolution Of RNA Viruses. *Annual Review of Microbiology* 41, 1 (1987), 409–433. https://doi.org/10.1146/annurev.micro.41.1.409

[18] John J. Treanor, H. Keipp Talbot, Suzanne E. Ohmit, Laura A. Coleman, Mark G. Thompson, Po Yung Cheng, Joshua G. Petrie, Geraldine Lofthus, Jennifer K. Meece, John V. Williams, Lashondra Berman, Caroline Breese Hall, Arnold S. Monto, Marie R. Griffin, Edward Belongia, and David K. Shay. 2012. Effectiveness of seasonal influenza vaccines in the United States during a season with circulation of all three vaccine strains. *Clinical Infectious Diseases* 55, 7 (2012), 951–959. https://doi.org/10.1093/cid/cis574

[19] Yun Zhang, Brian D. Aevermann, Tavis K. Anderson, David F. Burke, Gwenaelle Dauphin, Zhiping Gu, Sherry He, Sanjeev Kumar, Christopher N. Larsen, Alexandra J. Lee, Xiaomei Li, Catherine MacKen, Colin Mahaffey, Brett E. Pickett, Brian Reardon, Thomas Smith, Lucy Stewart, Christian Suloway, Guangyu Sun, Lei Tong, Amy L. Vincent, Bryan Walters, Sam Zaremba, Hongtao Zhao, Liwei Zhou, Christian Zmasek, Edward B. Klem, and Richard H. Scheuermann. 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Research* 45, D1 (2017), D466–D474. https://doi.org/10.1093/nar/gkw857