

RESEARCH ARTICLE

Learning the sequence of influenza A genome assembly during viral replication using point process models and fluorescence in situ hybridization

Timothy D. Majarian^{1,2[‡]}, Robert F. Murphy^{1,2,3,4*}, Seema S. Lakdawala^{5*}

1 Computational Biology Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **2** Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **3** Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **4** Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **5** Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

[‡] Current address: Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America

* murphy@cmu.edu (RFM); Lakdawala@pitt.edu (SSL)



OPEN ACCESS

Citation: Majarian TD, Murphy RF, Lakdawala SS (2019) Learning the sequence of influenza A genome assembly during viral replication using point process models and fluorescence in situ hybridization. *PLoS Comput Biol* 15(1): e1006199. <https://doi.org/10.1371/journal.pcbi.1006199>

Editor: Rob J. De Boer, Utrecht University, NETHERLANDS

Received: May 13, 2018

Accepted: November 20, 2018

Published: January 28, 2019

Copyright: © 2019 Majarian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. All source code and results of calculations are available at <http://murphylab.cbd.cmu.edu/software>.

Funding: This study was supported in part by National Institutes of Health grants AI108600-01 (SL), GM090033 (RFM), and GM103712 and by National Science Foundation grant MCB1616492 (RFM). The funders had no role in study design,

Abstract

Within influenza virus infected cells, viral genomic RNA are selectively packed into progeny virions, which predominantly contain a single copy of 8 viral RNA segments. Intersegmental RNA-RNA interactions are thought to mediate selective packaging of each viral ribonucleoprotein complex (vRNP). Clear evidence of a specific interaction network culminating in the full genomic set has yet to be identified. Using multi-color fluorescence *in situ* hybridization to visualize four vRNP segments within a single cell, we developed image-based models of vRNP-vRNP spatial dependence. These models were used to construct likely sequences of vRNP associations resulting in the full genomic set. Our results support the notion that selective packaging occurs during cytoplasmic transport and identifies the formation of multiple distinct vRNP sub-complexes that likely form as intermediate steps toward full genomic inclusion into a progeny virion. The methods employed demonstrate a statistically driven, model based approach applicable to other interaction and assembly problems.

Author summary

Influenza virus consists of eight viral ribonucleoproteins (vRNPs) that are assembled by infected cells to produce new virions. The process by which all eight vRNPs are assembled is not yet understood. We therefore used images from a previous study in which up to four vRNPs had been visualized in the same cell to construct spatial point process models that measure how well the subcellular distribution of one vRNP can be predicted from one or more other vRNPs. We used the likelihood of these models as an estimate of the extent of association between vRNPs and thereby constructed likely sequences of vRNP assembly that would produce full virions. Our work identifies the formation of multiple

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

distinct vRNP sub-complexes that likely form as intermediate steps toward production of a virion. The results may be of use in designing strategies to interfere with virus assembly. We also anticipate that the approach may be useful for studying other assembly processes, especially for complexes with modest affinities and more components than can be visualized simultaneously.

Introduction

Influenza A virus, part of the *orthomyxoviridae* family, has a segmented genome of eight distinct viral RNA segments coding at least 11 major proteins and several auxiliary peptides. Most notable of the 11 viral proteins are hemagglutinin (HA) and neuraminidase (NA), the canonical *H* and *N* in influenza strain designation. A segmented genome offers potential evolutionary advantages during viral replication, in the form of reassortment. The exchange of genomic material between two distinct viral strains in a co-infected cell often causes a shift within the genome, greatly increasing viral genetic diversity. In only the last century, influenza pandemics of 1957 (H2N2), 1968 (H3N2), and 2009 (H1N1) have emerged through reassortment of at least two viruses [1]. Segmented genomes also come with an inherent mechanistic challenge: ensuring that progeny receive a complete set of genomic segments. Random packaging is possible but comes at the cost of producing an overwhelming majority of progeny that are not viable [2]. Evidence suggests that there exists an active mechanism of selective packaging ensuring progeny viability through efficient and accurate genomic packaging [3, 4].

Within a virion, the viral genome is organized into individual viral ribonucleoprotein (vRNP) complexes, composed of the viral RNA, virally encoded nucleoprotein (NP), and a heterotrimeric polymerase complex made up of PB1, PB2 and PA. The vRNP structure is known classically as a helical panhandle, where the RNA is wrapped around NP with RNA bases exposed and the 3' and 5' ends associated with the polymerase complex [3]. We have recently demonstrated that the structure of vRNP is more complex than the classically depicted “beads-on-a-string” schematic. The NP and viral RNA associate in a non-uniform manner, where regions of the RNA are unbound by NP and capable of forming complex structures [5]. Macro-organization of vRNPs within the viral capsid show tight packaging, organized in a ‘7 +1’ orientation, a single vRNP center shaft surrounded by seven others (4–6).

Given the need to package one copy of all eight vRNP segments, there is strong evidence for the existence of a selective packaging mechanism prior to budding [3, 6–9]. Intersegmental RNA-RNA interactions have been proposed to mediate selective packaging of all eight influenza vRNP segments. The 5' and 3' regions of each segment have been implicated to be essential for efficient packaging and may be critical for the proposed RNA-RNA interactions [3, 10–13]. Previous studies using *in vitro* transcribed viral RNA have successfully demonstrated RNA-RNA binding. These studies also show that a set of eight reproducible interactions can suffice to form a full genome complex [14–16]. However, these studies were performed on viral RNA in the absence of NP which would alter the regions of viral RNA accessible for the proposed RNA-RNA interactions. Recent *in vivo* experiments have also pointed towards RNA-RNA interaction as the driving mechanism for selective packaging [17]. While there is consensus of its existence, a clear network of segment interactions leading to a full genome complex has yet to be elucidated.

We and others have previously used fluorescence *in situ* hybridization (FISH) studies to visualize the intracellular localization of multiple vRNP segments during a productive viral infection [18, 19]. These studies, performed eight hours post infection to capture an initial infection cycle and reduce complication of cytoplasmic vRNP from a subsequent infection,

have shown that distinct vRNP segments colocalize within the cytoplasm. Based on our previous studies, we proposed a model whereby vRNP segments are exported from the nucleus, the site of vRNP synthesis, as subcomplexes that undergo further assembly en route to the plasma membrane through dynamic fusion or colocalization events [19]. We also performed a series of binary colocalization comparisons, which would postulate a simple linear vRNP interaction network. However, no such network was identified, suggesting the presence of a more complex vRNP interaction network, including higher order subcomplexes as intermediate steps [19].

The lack of a method to image all eight vRNP segments within a single infected cell has limited our ability to examine the precise spatial relationship between segments. In this study, we developed a model based approach, rooted in point process theory, to quantify vRNP spatial dependence as a metric for RNA-RNA interaction using multi-color FISH images. A spatial point process model is a statistical model that captures the probability that an event will happen at any position within a given geometry. The probability can depend on the spatial distributions of other events or objects, often referred to as covariates or factors. In this context, an event is the observance of fluorescent signal (corresponding to one or more vRNPs) at a given location. We have previously used point process modeling to construct generative models reflecting the spatial dependencies between various punctate cell organelles and other cellular structures, such as the cell and nuclear membranes, microtubules and the endoplasmic reticulum [20, 21]. Related analytical methods have been used to analyze clustering of molecules reflecting spatial dependencies between copies of the same structure [22, 23]. Here we use point process models to present a statistically rigorous analysis of intracellular influenza A genome assembly dynamics from a spatial perspective during a productive viral infection.

For an imaged cell, the locations of each vRNP segment can be represented as “realizations” of an underlying probability density specific to that vRNP, calculated over the cytoplasm, and dependent on the spatial proximity to various other structures within the cell. That is, at any given position within the cell, there exists a probability of observing a single segment, of a particular identity, that depends on the positions of cellular organelles and other vRNP segments. By modeling the locational densities of each vRNP, we can therefore learn a dependency network from which likely vRNP interactions are implied.

FISH imaging yields many distinct observations of individual cells, each with multiple vRNP patterns observed. These point patterns, when viewed as realizations of point processes, allow for statistical learning of spatial dependencies over many replicates for each vRNP segment. By defining a set of covariates as the minimum distances between a given segment and other segments observed within the same cell, we produce a set of models that describe the spatial relationship between distinct segments. The extent to which two (or more) vRNPs are dependent upon each other (e.g., likely to be found at nearby positions) is reflected in the likelihood that images of those vRNPs would have been produced by a learned model: high likelihoods signify that the positions of a given vRNP can be predicted well from the others in that model. Model likelihoods can then be seen as a metric of vRNP interaction, with higher likelihoods indicating more probable vRNP association. These likelihoods can then be used to construct the most probable sequence of interactions to form the full genome with methods borrowed from evolutionary tree construction [24].

Results

Imaging influenza virus infected cells

To investigate the spatial dependence between vRNP segments, we utilized images from a previous study of infected cells stained for different combinations of vRNP segments using four-color FISH at eight hours post infection [19]. The data set included 14 different probe

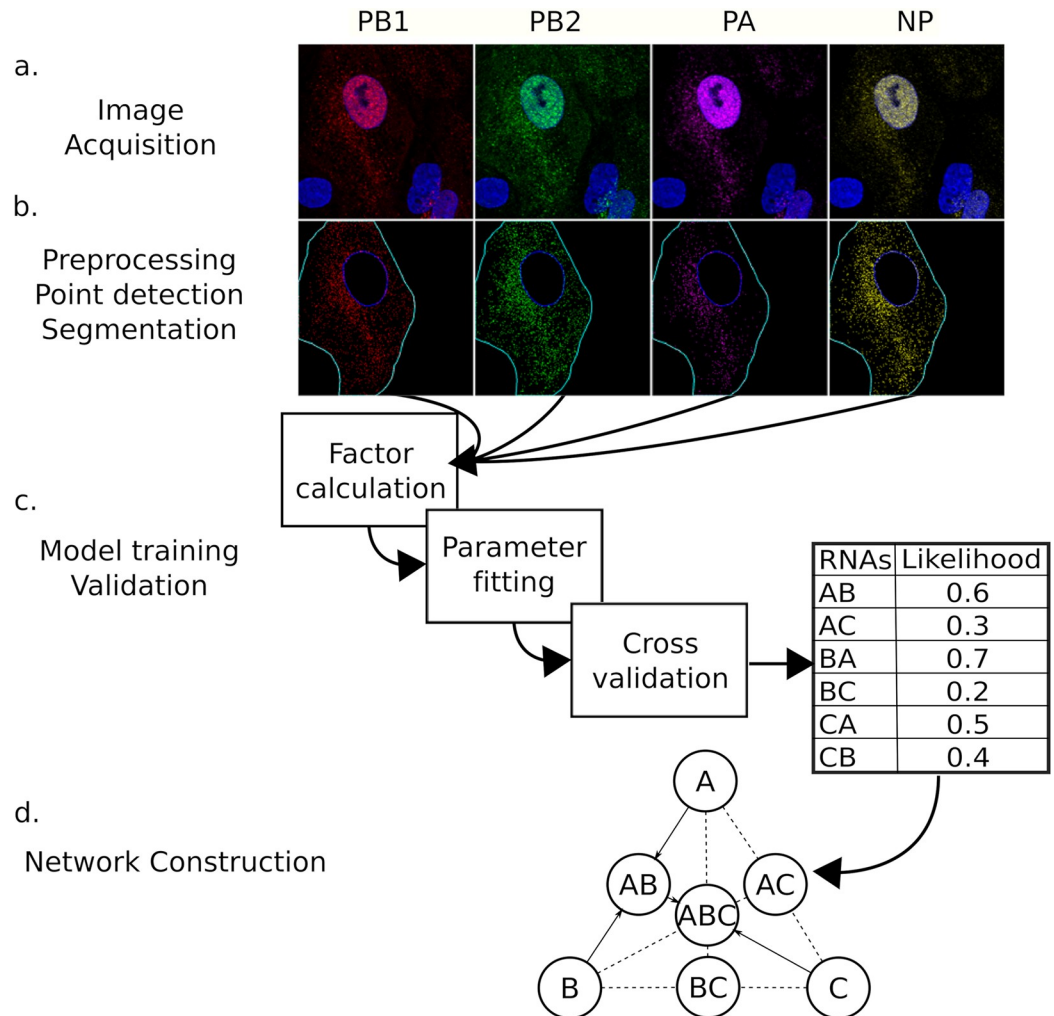


Fig 1. Illustration of the analysis pipeline. (a) Unprocessed fluorescence channels for cells FISH tagged for four vRNA segments and stained with DAPI. (b) Preprocessed image channels after filtering, nuclear and cell membrane segmentation, and point identification. The 3D images in (a) and (b) are displayed as maximum value projections. (c) Factors are calculated for each channel in each image, model parameters are fit, and a likelihood generated. (d) Dynamic programming is used to generate a most likely set of interactions yielding a full genome. Edges are possible interactions, solid lines indicate edges that are part of the most likely set of interactions. In this toy example, the most likely set of interactions is A and B binding to form AB, and then C binding to AB.

<https://doi.org/10.1371/journal.pcbi.1006199.g001>

combinations that covered all pairwise vRNP associations and 32 out of 56 possible triple vRNP combinations (S1 and S2 Tables). Prior to initiating pattern analysis, each image was processed to remove background, isolate individual cells, and find punctate vRNP segments. Fig 1A and 1B presents representative raw and segmented images with detected points.

Assessing randomness of the spatial distributions of vRNP

We began by determining the extent to which the individual vRNPs could be considered randomly distributed throughout the cell, since our search for spatial dependencies between different vRNPs would be illogical if their positions were all random. We defined the null hypothesis as complete spatial randomness, as would be exhibited by a homogenous Poisson process, and estimated a p-value for this hypothesis using Monte Carlo methods. Briefly, for a

Table 1. Monte Carlo p-values and variances for the hypothesis that the spatial distribution of each vRNA segment is distributed by a homogeneous Poisson process.

	Mean p-value	Variance
PB2	0.0034	0.00011
PB1	0.00524	0.00034
PA	0.00467	0.0002
HA	0.00671	0.00015
NP	0.01185	0.00036
NA	0.00249	0.00003
M	0.00005	0
NS	0.00306	0.00007

<https://doi.org/10.1371/journal.pcbi.1006199.t001>

given vRNP in a particular cell, we started by segmenting the cytoplasm and counting the number of points observed within. We then simulated (generated) many random point configurations, of the same number of points, within the segmented cytoplasm. We calculated a test statistic for each random configuration (see [Methods](#)) and used these to assign a p-value to the test statistic for the observed pattern for that cell; these were averaged for each vRNP. As shown in [Table 1](#), every vRNP segment showed significant deviation (at the $p < 0.05$ level), with low variance, from a spatially random distribution. This indicates the presence of a spatial trend and some dependency on cellular structures for their location.

Modeling cell and nuclear membrane dependence

Given that the localizations of vRNPs are not random, we sought to determine if some relationship to the cell and nuclear membranes could explain the spatial trend of vRNP segments. Three covariates were defined for each position in the cell (minimum distance to the cell membrane, minimum distance to the nuclear membrane, and the minimum ratio of cell to nuclear distance for each point) and models were created for each vRNP using each of the covariates separately. We estimated the accuracy of each model by cross validated likelihood calculation (see [Methods](#)), where values closer to 0 are more likely. It is important to note that, because actual likelihoods were calculated, they can be compared between models of different complexity.

Dependency on the cell membrane produced the lowest likelihood (worst) models relative to all others ([Table 2](#)) for all vRNP segments. Congregation of vRNP to the apical cell membrane is a potential skewing factor, since there will be a high concentration of all segments at this cellular location. The many observed points within close proximity to the membrane potentially caused low fitted probability of observing other points closer to the nucleus. Cell

Table 2. Cross validated negative log-likelihoods for models dependent on distance to different reference locations.

	Cell Model	Nuclear Model	Cell+Nuclear Model
PB2	2.5547	0.4153	0.9158
PB1	2.3908	0.408	1.2574
PA	2.4228	0.4643	1.2352
HA	2.443	2.3928	4.396
NP	2.4885	2.4584	4.5447
NA	2.5019	2.4372	2.1754
M	2.4871	2.1294	3.0799
NS	2.3906	0.6666	1.4624

<https://doi.org/10.1371/journal.pcbi.1006199.t002>

Table 3. Cross validated negative log-likelihoods for pairwise interaction models. The row label indicates the independent (secondary) variable, the vRNP upon which a model predicting the column label (the dependent or primary variable) was constructed.

	PB2	PB1	PA	HA	NP	NA	M	NS
PB2		0.3705	0.2694	0.282	0.3146	0.5605	0.2146	0.3926
PB1	0.2398		0.2018	0.2044	0.1186	0.4015	0.1701	0.2905
PA	0.2276	0.2217		0.2284	0.1502	0.3222	0.2816	0.2315
HA	0.3334	0.1266	0.1838		0.2787	0.232	0.2623	0.2845
NP	0.3438	0.1805	0.1942	0.2895		0.2529	0.275	0.3411
NA	0.3191	0.324	0.3093	0.2947	0.2742		0.2747	0.2982
M	0.3287	0.2236	0.2832	0.2803	0.2975	0.2897		0.2848
NS	0.3427	0.2888	0.2292	0.2838	0.325	0.3436	0.2841	

<https://doi.org/10.1371/journal.pcbi.1006199.t003>

shape differences combined with lower resolution in the z-dimension may have also influenced these models.

For most vRNPs, models incorporating nuclear distance showed higher likelihoods than those using cell membrane distance (Table 2). Some vRNP segments, such as PB2, PB1, PA and NS, showed greater dependency on nuclear distance than the other segments. The ratio of cell membrane to nuclear distance generally produced models better than cell membrane alone, but they were worse than nuclear alone in all cases except NA (Table 2). Thus nuclear distance provides the most cogent model for dependence of vRNP localization on a cellular structure.

Inter-segment dependence in pairs, triplets, and quadruplets

Every pair of vRNP segments was observed in at least three cells, allowing for modeling of vRNP-vRNP dependence among all segment pairs. For each pair of vRNP segments, two models were trained, switching the dependent (primary) and independent (secondary) patterns. Since the minimum inter-pattern distance measure is not commutative, the two models need not be the same. More simply, a given vRNP location may depend on its proximity to another vRNP, but the converse is not necessarily true. This can be observed when one vRNP segment is found in two distinct locations, and a second vRNP segment is found in only one of those locations. The second vRNP segment would be observed to have a high likelihood for being predicted from the first, but the first would not have a high likelihood of being predicted from the second.

In all cases, pairwise inter-pattern distance models (Table 3) for a given vRNP had higher likelihoods than those produced for that vRNP with only the cell and/or nuclear membrane factors (Table 2). Some pairs showed higher likelihoods than others, such as PB1-NP, PB1-M, PA-NP, HA-PB1, and NP-PB1. These pairings also showed relatively low variance in cross-validated likelihood estimates (S3 Table). Many of the highest likelihood vRNP pairs are observed between segments encoding the viral polymerase components (PB2, PB1, PA and NP). Given the dependence between these proteins both structurally and functionally [25–27], it is possible that these segments evolved a dependence upon each other to ensure their joint packaging in a virion (a possibility that would require extensive additional work to examine).

Extending the pairwise models, nuclear distance was added as another covariate. Minor improvements in model likelihood were seen in some models (Table 4) with others showing decreased likelihood. Overall there was an increase in the variance of our estimates (S4 Table). Overfitting to nuclear distance may be the cause of this increased variance. Fig 2 highlights the pairwise models in Table 4 by illustrating vRNP pairs that are below 0.23 log likelihood. Overall, vRNPs showed the greatest dependence on PA and PB1, with the most connections

Table 4. Cross validated negative log-likelihoods for pairwise interaction models that also include nuclear distance. The row label indicates the independent (secondary) variable, the vRNP upon which a model predicting the column label (the dependent or primary variable) was constructed.

	PB2	PB1	PA	HA	NP	NA	M	NS
PB2		0.3817	0.2694	0.2833	0.3145	0.5643	0.2166	0.3977
PB1	0.2427		0.2023	0.2191	0.1215	0.4188	0.1714	0.3097
PA	0.2278	0.2211		0.661	0.1504	0.3238	0.4967	0.5378
HA	0.336	0.13	0.9138		1.944	1.7048	1.2324	0.5447
NP	0.3084	0.1746	0.1949	0.2095		1.6062	1.6495	0.3523
NA	0.3166	0.3317	0.3114	1.7801	1.4877		1.0989	0.3
M	0.33	0.2246	0.5374	1.2838	1.7105	1.1884		0.4153
NS	0.3454	0.2816	0.6259	0.4561	1.7998	0.3466	0.4368	

<https://doi.org/10.1371/journal.pcbi.1006199.t004>

between these two segments and the others. The least dependence was observed for NA and NS vRNP segments.

Thirty-two of 56 possible triplet vRNP segment complexes (referred to as triplets) were present within the dataset analyzed and could therefore be modeled within the same cell. This process allows for modeling of complex vRNP relations containing more than two vRNP segments. Three models were trained for each observed triplet, one for each vRNP depending on the other two. The highest likelihood model was chosen. Model likelihoods tended to be higher than that of either pair model for a given vRNP (Table 5). We observed that the most likely triplet models for each vRNP included four of the five most likely vRNP pairings observed in pair likelihood models (Table 3), (namely PB1-NP, PA-NP, HA-PB1, and NP-PB1). However, not all of the most likely pairwise associations (e.g., PB1-M) were observed in the most likely triplet models, highlighting that models based solely on pairwise comparisons might not accurately represent larger order vRNP complexes. Model likelihoods for all triplets observed can be found in S5 Table.

A small subset of all possible four vRNP segment complexes (referred to as quadruplets) were contained within the dataset and models for predicting one of the four from the other three were also constructed. Again, considerable overlap is seen between the most likely quadruplet models and both triplet and pair models (Table 6), such as PB2-PB1-PA and PB1-PA. The consistency in vRNP composition between observed likely quadruplet, triples, and pairs further validates the spatial dependence of vRNP segments and the presence of vRNP subcomplexes as influenza A assembly intermediates. Model likelihoods for all quadruplets observed can be found in S6 Table.

One of our major goals was to attempt to extrapolate higher order relationships from the pair, triple and quadruplet models. To project how accurately this might be done, we compared our observed triple and quadruple models with predicted models calculated from any of the four pair models (vRNP pairs (Table 3), vRNP pairs + nucleus (Table 4), vRNP pairs + cellular membrane, and vRNP pairs +nuclear/cell membrane ratio) (Fig 3A). We also compared observed quadruplet likelihoods to those predicted using both the pair and triple models, similar to the pair models any of the 4 models including the various cellular features were included (Fig 3B). Most of our observed triples and quadruplets could be predicted with modest accuracy from the pair models but the overall r^2 value was only 0.302. This was due to inaccuracy in predicting the triple and quadruple clusters containing the PB2-NA pair with either NS, PB1 or both. These combinations were predicted to occur much more often than was observed in the experimental data sets. The pairwise likelihood of NA with PB2 is > -0.6 , which is consistent with the observed likelihood and may be the driving force behind the score compared to the other pairwise interactions. Analysis of the predicted triples and quadruples

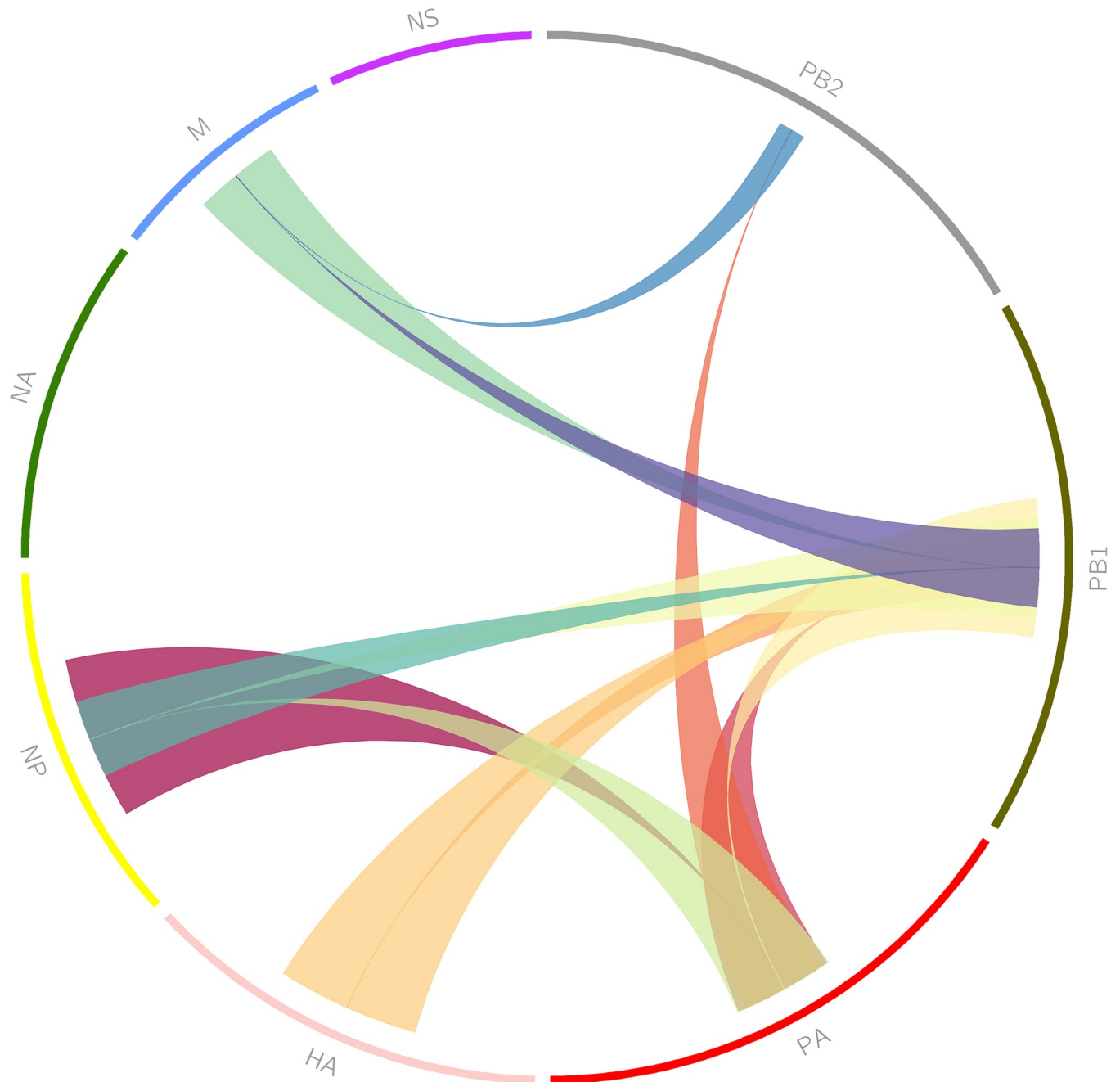


Fig 2. Pairwise interactions learned from MDCK cell images. Each vRNA segment is arranged around the perimeter. Bands connect vRNA segments that interact with high likelihood. The position of the base of each band indicates the independent (secondary) variable and the larger the width of the base the more likely the interaction. Only interactions whose log likelihoods are below 0.23 from Table 4 are shown.

<https://doi.org/10.1371/journal.pcbi.1006199.g002>

after removing the PB2-NA containing clusters resulted in a r^2 value increases to 0.642 (we cannot justify this removal on a statistical basis but rather on the observation that the combinations with the poorest predictions all involved the same pair). Predicting quadruples from pairs and triples resulted in a r^2 value of 0.450, and with removal of the same quadruple it rose to 0.714. Overall, these data support the notion that pairwise point process models can be used to predict higher order complexes with moderate accuracy, and that prediction accuracy increases if higher order models are used.

Table 5. Cross validated negative log-likelihoods for the top two triple models with each vRNA segment as the primary pattern. Pairwise model negative log-likelihoods for the members of each model are shown in the right columns.

RNA	CV Likelihood	PB2	PB1	PA	HA	NP	NA	M	NS
PB2									
PB2 M PB1	0.114		0.382					0.217	
PB2 HA M	0.162				0.283			0.217	
PB1									
PB1 PA PB2	0.062	0.243		0.202					
PB1 NP PB2	0.071	0.243				0.122			
PA									
PA PB2 PB1	0.113	0.228	0.222						
PA PB2 NP	0.121	0.228				0.15			
HA									
HA M PB1	0.059		0.13					1.232	
HA PB1 NS	0.111		0.13						0.545
NP									
NP PB1 PB2	0.141	0.344	0.181						
NP PB2 PA	0.145	0.344		0.195					
NA									
NA HA NP	0.225				1.78	1.488			
NA HA NS	0.246				1.78				0.3
M									
M PB1 NS	0.201		0.225						0.415
M HA PB1	0.201		0.225		1.284				
NS									
NS M PB1	0.169		0.289					0.437	
NS HA PB1	0.184		0.289		0.456				

<https://doi.org/10.1371/journal.pcbi.1006199.t005>

Interaction network construction

Given these results, we sought to create a model of the set of vRNP interactions (complexes) that occur at all steps in the formation of a complete viral genome. To do this, we represented all possible interactions between vRNP segments as a weighted, directed, acyclic graph with nodes labeled by each unique, unordered set of vRNP segments. A path from all single nodes to the root represents a set of vRNP interactions resulting in a full genome complex (see [Methods](#)). For illustration, if all eight vRNP segments directly formed a complete set without forming any intermediate complexes (a highly unlikely scenario), this would be represented by weights of one for the connection between each vRNP and the complete vRNP supramolecular complex containing all eight segments and weights of zero for all other complexes. Such a scenario would result in a tree with the root connected to each of the leaves.

In our analysis, we estimated the weight for each possible intermediate vRNP complex assembly subcomplex to contain at least two and up to eight segments based upon either the observed or predicted likelihoods (see [Methods](#)). Dynamic programming, an efficient method for recursively finding the highest scoring path to a given node, was then used to find the most probable path within the graph using the highest likelihood trained model for a given nodes from those with or without cell and/or nuclear features ([Fig 4A](#)). The resultant tree produces two distinct clusters HA, M, NA, NS (hereafter referred to as C1) and PB1, PB2, NP (C2) that merge with PA as a final step towards forming the full genome. This network suggests three steps to form a complete set of all eight vRNP segments: 1) formation of C1, 2) formation of C2 and 3) addition of PA to C1 and C2 to form the final product.

Table 6. Cross validated negative log-likelihoods for the top two quadruple models with each vRNA segment as the primary pattern. Pairwise model negative log-likelihoods for the members of each model are shown in the right columns.

RNA	CV Likelihood	PB2	PB1	PA	HA	NP	NA	M	NS
PB2									
PB2 NS PB1 NA	0.571		0.382				0.564		0.398
PB2 NP PA PB1	0.258		0.382	0.269		0.315			
PB1									
PB1 NS NA PB2	0.347	0.243					0.419		0.31
PB1 PA M PB2	0.16	0.243		0.202				0.171	
PA									
PA NS M HA	0.655			0.438	0.661			0.497	0.538
PA NA M NS	0.272			0.438			0.324	0.497	0.538
HA									
HA PA M NS	0.869				0.914			1.232	0.545
HA NS PB2 M	0.317	0.336						1.232	0.545
NP									
NP HA M NA	0.233				2.095		1.606	1.65	
NP PB2 PA PB1	0.139	0.344	0.181	0.195					
NA									
NA PA NS M	0.272			0.311				1.099	0.3
NA PB1 PB2 NS	0.268	0.319	0.332						0.3
M									
M NS HA PA	0.7			0.537	1.284				0.415
M PA NS NA	0.336			0.537			1.188		0.415
NS									
NS HA PA M	0.705			0.626	0.456			0.437	
NS PB2 M HA	0.340	0.343			0.288			0.289	

<https://doi.org/10.1371/journal.pcbi.1006199.t006>

To assess whether a particular vRNP localization dataset was driving this network, we performed a series of network analyses with data excluded. First, the contribution of observed quadruplets and triplets was assessed by excluding observed quadruplets (Fig 4B) and without observed quadruplets and triplets (Fig 4C). Interestingly, exclusion of all observed quadruplets did not drastically alter the network, with both C1 and C2 still being present, although formation of C1 required two steps. However, networks built only with observed pairs (excluding triplets and quadruplets) resulted in a quite different network lacking larger subcomplexes (Fig 4C), reflecting the fact that pairs were only able to make approximate estimates of higher order interactions.

Given that the dynamic programming approach yields only a single most likely tree, we sought to assess the robustness of the inferred tree to potential inaccuracy in the estimated dependencies. We therefore assessed the stability of the tree by adding various amounts of random noise to the estimated likelihoods (see Methods). Surprisingly, the addition of noise of up to a quarter of the mean log-likelihood had no effect on the network found, yielding the same network (Fig 4A). As noise levels increased to half of the mean log-likelihood, one other tree was generated (Fig 4F). This demonstrates the stability limit of the construction but notably this revised tree still contains clusters that resemble the original. If the original tree was based on biased likelihoods from imaging artifacts, the addition of noise would have resulted in an altered tree network at lower noise levels and with higher frequency, but since this did not occur, we are confident in the accuracy of our original likelihood estimations.

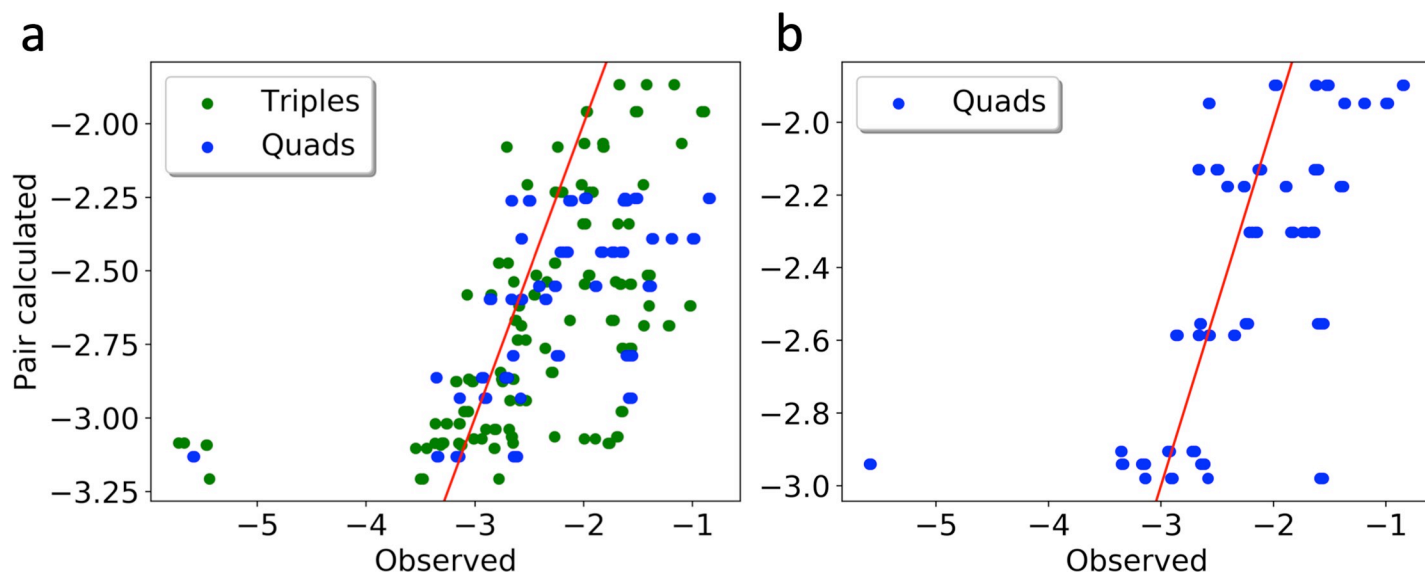


Fig 3. Comparison of composite and observed model likelihoods for triplet and quadruplet models. To ensure that the recursion in our dynamic programming scheme accurately predicts model likelihoods, we compared triplet and quadruplet observed likelihoods to those generated from only pairwise models (that include all four possible models: vRNP pairs alone, vRNP pairs + nucleus, vRNP pairs + cell membrane, vRNP pairs + nuclear/cell membrane ratio) (a). We also compared observed quadruplet likelihoods to those generated from pairwise and triplet models (using all 4 possible models) (b). Ideally, all points would fall on the red line, matching observed and predicted values perfectly. The cluster in the lower left corner represent interactions that are predicted to occur with high likelihood but in reality, do not. All points within this cluster contain the same four vRNA segments (PB1, PB2, NA, NS). Exclusion of this cluster from our analysis increases the association between predicted and observed triplets and quadruplets.

<https://doi.org/10.1371/journal.pcbi.1006199.g003>

Finally, to exclude the possibility that datasets directly measuring C1 and C2 (reactions L and B, respectively) were biasing the network towards inclusion of those sets of vRNPs, we performed a network analysis of a dataset where the measured likelihoods of these two clusters was excluded. Removal of data derived from images of the sets of vRNPs in C1 and C2 (Fig 4E) yielded networks that resemble Fig 4A, indicating that even in the absence of data from these clusters, similar clusters will form during assembly. Removal of each cluster alone was recapitulated by lower order interactions. Removal of C1 [HA, M, NA, NS] data resulted in the same tree presented in Fig 4B, where HA, M, NA, NS form in two steps from a triplet and a single vRNP segment. Exclusion of C2 [NP, PB1, PB2] result in a tree similar to Fig 4D, where C2 cluster forms from a pair and a single segment (Fig 4D). Removal of data was also combined with the addition of noise to further test stability. Predominantly, the same networks were generated with and without noise. As noise levels increased in the exclusion constructions, a tree with HA, M, NS, PB1 in place of C1 and NP, PA, PB2 in place of C2, where PB1 and PA were not observed in the C1 or C2 original clusters respectively (Fig 4F). Together, these constructed networks suggest a primary interaction scheme that can be further experimentally tested and the importance of triplet and quadruplet observations in predicting higher order vRNP complexes. Other, very low frequency trees that were generated in the presence of high levels of noise are located in the reproducible research archive.

Discussion

Influenza A vRNP segments selectively assemble within an infected cell to produce progeny virions containing one copy of all eight segments. The mechanism driving selective assembly is still largely unknown, but RNA-RNA interaction between the segments has been proposed [14–16]. In this study, we have modeled the *in vivo* spatial dependencies of influenza A vRNP

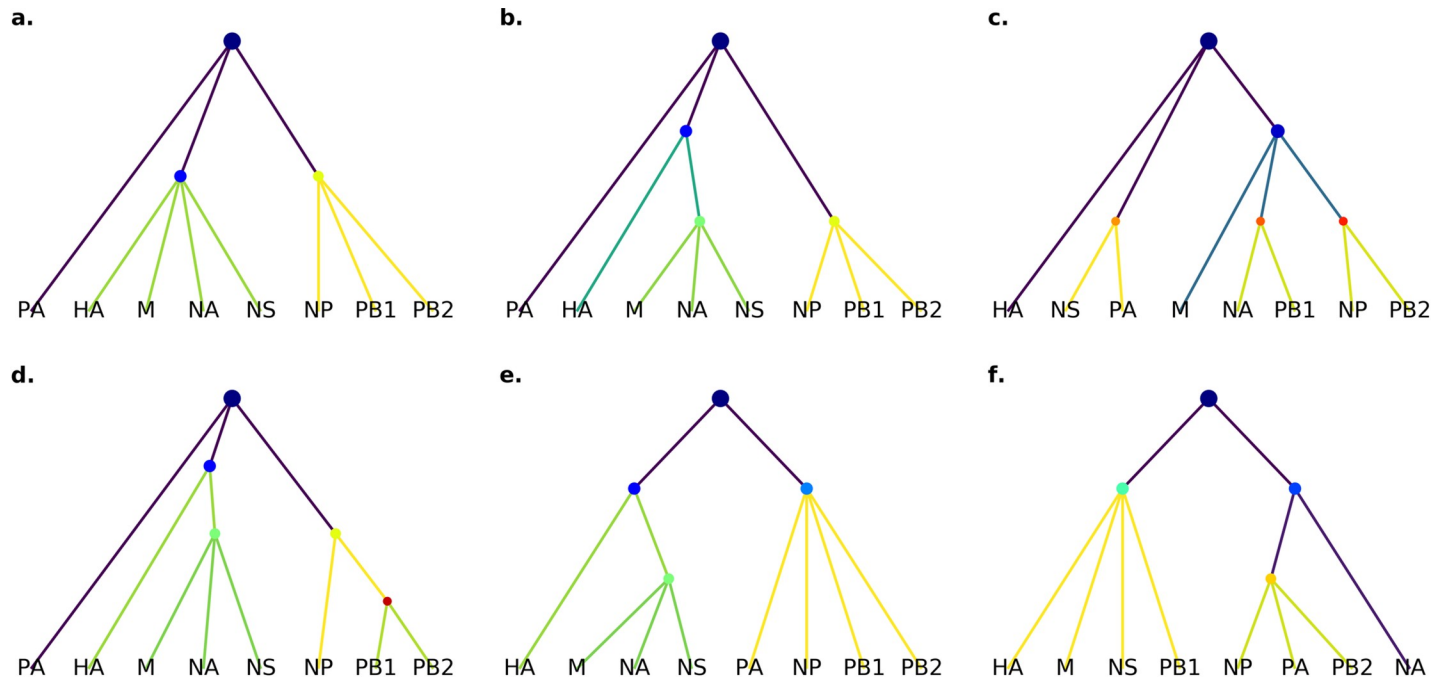


Fig 4. Constructed interaction networks. Using dynamic programming, we constructed the single most likely set of interactions yielding a full genome. First, we used all model likelihoods with and without nuclear and cell features (a) to build a master tree. Hotter edge colors correspond to more likely interactions. Since the images contain only a subset of the possible triplets and quadruplet, we assessed how much this affected the master tree by removing all quadruplets (b) or both quadruplets and triplets (c). To test the stability of the master tree, trees were constructed with random noise added to each observed model likelihood before performing the dynamic programming. Multiple noise levels were tested up to 1/2 of the mean model likelihood with the noise level being the standard deviation of a Gaussian centered at 0. For each level of noise, network construction was performed 100 times. Noise up to 1/4 mean likelihood always resulted in (a). Increasing noise levels yielded (a) on 40% to 60% of constructions. Given the two observed clusters in the master tree, we next asked whether the composite likelihood function could recapture their interactions without actually observing them. Removing all models involving HA, M, NS, NA (C1) and NP, PB1, PB2 (C2) yielded (d). Adding noise up to 1/4 mean likelihood and removing both clusters also generated (d). Removal with higher levels of noise generated (e) for 27% of constructions. In all stability analyses with noise greater than 1/4 mean likelihood, (f) was generated with frequency between 5% and 15%.

<https://doi.org/10.1371/journal.pcbi.1006199.g004>

segments using multi-color fluorescent *in situ* images to generate possible vRNP-vRNP interaction networks and propose a new perspective on genome packaging during viral replication. By utilizing a rigorous statistical framework, we have extended the efforts of previous groups and present a novel method for construction of vRNP interaction networks based on their precise spatial information within actively infected cells. Using spatial proximity as a proxy for physical interaction, point process modeling in this study suggests clear spatial dependencies between certain vRNP segments, and confirms our previous observations of subcomplex formation during cytoplasmic transport (18).

Previous studies using *in vitro* transcribed RNA suggest multiple interactions between vRNP segments are expected [3, 6, 14, 16, 28]. Similarly, the modeled likelihoods show potentially multiple interactions for each vRNP segment. Electron tomography studies have revealed a conserved ‘7+1’ supramolecular structure within the viral interior, where a center shaft is surrounded by seven remaining segments [3, 6, 28], demonstrating an ordered process in genome assembly.

Our results suggest two candidates for the center shaft or ‘master segment’: PA and PB1. From the six constructed networks, PA is the most variable in that it is least often paired early in assembly but, with the addition of noise and exclusion of data, may interact with both primary observed clusters. This would be expected if a segment was evenly dependent on most others, a possible signature of a core segment. PB1 also behaves similarly but to a lesser

degree. Both PB1 and PA show the highest average pairwise model likelihood over all other segments.

A few vRNP pairs were seen often in the most likely pairwise, triplet, and quadruplet models and in the final constructed network. PB1-PB2 and HA-M, show commutative relationships, appearing in one of the top most likely triplet and quadruplet models for each vRNP within the pair. Some non-commutative dependencies were also seen, most notably in NP depending on PB2 in both triplet and quadruplet models with little PB2-NP dependence.

By using dynamic programming, we were able to generate the most likely interaction network based on all models. While this network was surprising robust to the addition of noise to the model likelihoods, subtle variations in the vRNP interaction network were observed with the exclusion of higher order datasets, suggesting that there may exist multiple interaction pathways all acting at once in a single cell. Some of the variation in our results may also be due to the presence of some lower specificity vRNA-vRNA interactions also occurring during genome assembly. This would explain some of the highly likely pairwise interactions observed that were not included in the final interaction network, such as PB1 and HA. In addition, vRNP networks could also change over the course of a viral infection when packaging fidelity may be compromised [29]. Our current data set only considers one time point, eight hours post infection, which represents the time of initial virion release from infected cells, and would miss alternate networks present at later time points. Therefore, the analysis presented here may only capture a snapshot of a dynamic assembly process that could change throughout an infection. A system with multiple possible interaction networks could be potentially advantageous in both ensuring packaging of a full genome in viral particles and in future reassortment events. For example, reassortment of one vRNA segment during a coinfection may decrease the binding efficacy necessary for a single interaction network but a second network may serve to rescue viral viability while also increasing genetic diversity.

With this, we also point out the consistent presence of the PB2, PB1 and NP cluster. Since these three proteins function together, with PA, to form the viral polymerase, which promote vRNA replication and transcription, this cluster is potentially important. Viruses capable of packaging complementary polymerase segments into progeny virions will have a fitness advantage compared to viruses with polymerase mismatches. This phenotype has been observed in experimental reassortment experiments between 2009 H1N1 pandemic and seasonal H3N2 viruses [25, 27].

More generally, the computational pipeline presented may function as a useful tool in elucidating spatial dependency over a wide range of biological phenomena observable through microscopy. As we have demonstrated, point process models can capture key aspects of point distributions while retaining a basis in probability. Network construction synthesizes the results derived through modeling into a cogent, most-likely set of spatial interactions or dependencies.

Future work in spatial modeling of influenza A packaging should incorporate a temporal dimension through live cell imaging. Point process models have been adapted for spatiotemporal modeling [30]. Incorporation of intracellular markers important for influenza A vRNP transport may increase the accuracy of models. Influenza A vRNP transport from the nucleus, the site of vRNP synthesis, to the plasma membrane is a complex process utilizing a variety of host proteins [31]. Rab11A-containing vesicles are thought to be the primary mode of transport although there is evidence that a Rab11A-independent mechanism exists [32–34]. A recent study has implicated the ER to mediating transport of vRNP segments through anchored Rab11A proteins [35]. Novel methods in fluorescent multiplexing [36] present an exciting opportunity for observing many cellular structures which will provide a holistic image the spatial dependence of vRNP segments upon subcellular structures and to each other.

Methods

Multi-color fluorescent in situ hybridization images

This study used previously published multi-color FISH images from [19] that were generated at the National Institutes of Health. Briefly, MDCK cells were infected with recombinant WSN/1933 H1N1 for 8 hours and then fixed and stained with FISH probes, obtained from Biosearch technologies, against four distinct vRNP segments. DAPI was included as well to label DNA. Multi-color FISH samples were imaged on an Leica SP5 white light laser to ensure spectral separation of the five colors (Dapi, Alexa 488, Quasar 570, Cal Fluor Red 590, Quasar 670). The specificity of the probes and spectral separation between fluorophores was previously confirmed [19].

Image preprocessing, cell segmentation and point detection

All image preprocessing was performed in MATLAB, v. 2015a. Prior to point detection and segmentation, image noise was removed by convolution with a Gaussian mask. Each image channel was denoised individually. As described in [19], each image was captured with 0.17 μm z-step size spanning the entire cell volume, defined for each cell using both the nuclear and FISH staining to define the apical cell membrane. Each image was captured with a pixel size of $\sim 50 \times 50 \times 168$ nm.

Individual vRNP segments were identified by finding connected areas of signal within the denoised image. The center of each object was taken as the maximum intensity pixel within, yielding a set of point coordinates. Note that due to the thickness of the z sections, the apparent distance between points may be an underestimate of the true distance, but this effect is expected to average out when considering many points.

Since there was no fluorescent tag for cell membrane components, we estimated the cell boundary using the vRNP images. We assumed that fluorescent signal would be denser within the cell cytoplasm than in the surrounding area with cell-cell junctions subtly defined by ‘valleys’, curves of low signal density whose normal vectors point towards increasing density. The gradient of point density over the image was then used to segment individual cells from both the background and each other by the mean-shift algorithm [37]. Hand segmentation was also used to ensure accurate cell membrane segmentation. In most cases, the convex hull of all points within an identified cell region then defined the cell membrane. The nucleus was segmented through simple thresholding and smoothing.

Point processes and patterns

Within each cell, the vRNP segments form a point pattern, $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, where n is the number of observed points and x_i is the 3-dimensional coordinate vector for point i . The point pattern is defined over a bounded region, W , the segmented cell cytoplasm. \mathbf{x} is then viewed as a “realization” (an output) of some random point process \mathbf{X} , the generating distribution for all patterns of the particular vRNP identity. The process \mathbf{X} then represents the culmination of biological factors that determines vRNP segment location, eg. nuclear export, directed transport over the microtubule network, inhibition by other organelles, etc.

To model \mathbf{X} , we first define its locational density, f , a function over the cytoplasm, where $f(\mathbf{u})$ is the probability of observing a point at position \mathbf{u} . The simplest model for this density is the Homogenous Poisson, characterized by complete randomness over space, or uniform probability:

$$f(\mathbf{x}|n) = \frac{1}{Z} \lambda^n \quad (1)$$

where λ determines point density, and Z a normalizing constant. In the point process literature, λ is referred to as *intensity*, but we refer to it as point density to avoid confusion with fluorescence intensity.

Characterizing spatial randomness

Hypothesis testing for spatial randomness, comparing an observed pattern to an expectation under a Poisson assumption, was used as an initial motivation for further modeling. Ripley's K-function describes the number of neighboring points within a given distance (r) of each observed point in a pattern, a measure of clustering or inhibition of points within space. Under a Poisson assumption,

$$K_{poi}(r) = \pi r^2 \tag{2}$$

The expected K-function value can then be used to measure the difference between a given pattern and that of a Homogenous Poisson with an equal number of points. To assess this for observed point patterns, we used the test statistic

$$T = \int_0^{r_{max}} (\hat{K}(r) - K_{poi}(r))^2 dr \tag{3}$$

where $\hat{K}(r)$ is the estimated K-function for a given pattern and r a radius defining the point neighborhood. To provide a background distribution for this statistic for a given cell geometry, we generated 100 samples from a Homogenous Poisson process defined over its cell cytoplasm and calculated the test statistic. We then obtained a p-value for the hypothesis that an observed distribution was drawn from a homogenous Poisson process:

$$p = \frac{1 + \sum_{i=1}^m I(t_{obs} \geq t_i)}{m + 1} \tag{4}$$

where m is the number of samples drawn. The p-values were averaged over all cells for a given vRNP.

Modeling spatial dependence on cellular structures

An Inhomogenous Poisson Process, in its simplest form, is characterized by a spatially dependent locational density. Per point "factors" (e.g., distance to cell membrane) are used to determine the probability of a point occurring at a given location. As above, let \mathbf{X} be some point process with realization \mathbf{x} , and define $s(\mathbf{x})$, an $n \times k$ matrix where each $s_{i,j}$ is some factor j that can be calculated for each point x_i . The Inhomogenous Poisson point density function is:

$$\lambda(x_i; \theta) = (\exp(\theta^T s(x_i)))$$

where θ is the k-dimensional parameter vector. The likelihood of the model given point data is then defined by:

$$l(\mathbf{x}; \theta) = \left(\prod_{i=1}^n \lambda(x_i) \right) \exp \left(- \int_w \lambda(u) du \right)$$

To quantify vRNP dependence on cellular structures, we defined the first factor (f_1) as the distance to the nearest point on the nuclear membrane, the distance to the nearest point on the cell membrane, or the ratio of cell to nuclear distance.

Modeling spatial dependence between vRNPs

The second factor, minimum inter-pattern point distance, was used to explore whether there is vRNP-vRNP interaction between vRNPs of different types. For each observed vRNP segment of type l , and some other vRNP point pattern of type j , the minimum inter-pattern point distance is:

$$f_2(x_i^l | l \neq t) = \min_j d(x_i^l, x_j^t)$$

For each vRNP type, we use a ‘one depends on all’ schema. That is, each model with the same set of vRNPs is not equivalent but exclusively represents a single vRNP pattern (the primary pattern) depending on others (the secondary patterns). Models were constructed for every unique pair, triplet, and quadruplet of vRNP types.

For each model, all instances of its unique components were gathered, all images with tags for every vRNP in the model. Parameters of the model were then fit for each instance using the maximum pseudolikelihood [38]. The basic idea is that a given set of parameters can be used to calculate a value that is proportional to the likelihood (which is referred to as a pseudolikelihood) of observing a point at a given position, and the parameters can then be adjusted to give the highest total pseudolikelihood for available observations. Hold one out cross validation was used to assess the quality of fit over all instances yielding cross-validated pseudolikelihood estimated parameters and variances. These fitted parameters were then used to convert the pseudolikelihood to a true model likelihood by estimating the proportionality constant as previously described (19).

There is some variation in the number of images in which different vRNPs were visualized (S1 Table). This ranged from 4 for PB1 to 9 for M (the others were present 6 or 7 times). As pointed out by a reviewer, this difference raises the possibility that a vRNP that was imaged less frequently might be found to be underrepresented in the extent to which other segments were found to depend upon it. As seen in Table 3, this did not turn out to be the case, as PB1 and M showed similar average likelihoods and more vRNPs showed strong dependence on PB1 than on the others. Similarly, PB1 was not underrepresented in high scoring triples (Table 5).

Interaction network assembly

We formulate the task of finding the most probable set of vRNA interactions yielding a full genome as a graph problem. First, let the set of nodes in the graph be every possible combination of vRNAs ranging in size from a single vRNA to all 8 segments. Directed edges in the graph connect subset to superset nodes. We weight each edge of the graph by the likelihood of the most likely model (chosen from models with only vRNP dependencies, and those with additional dependencies on either nuclear distance, cell membrane distance, or cell to nuclear distance ratio) that results in the superset and contains the subset as a primary or secondary segment (Fig 1). Edges that represent models that have not been observed remain unweighted.

From this graph, a set of interactions yielding the full genome is a tree with leaves as the vRNA singletons and root as the full 8-mer of vRNA segments. In these trees, we require each node to have at least two incoming edges, barring the leaf nodes. The set of incoming edges to a given node must emanate from nodes whose labels, when combined, exactly equal the given node. A set of incoming edges essentially ‘combines’ nodes, directly corresponding to the physical event of vRNP segments binding. The most likely set of interactions is also the most highly weighted tree. The problem of finding the most probable interaction network then reduces to finding the highest weighted tree in the graph. This problem can easily be solved

through dynamic programming within a paradigm often used in evolutionary tree construction [24].

The likelihoods of models involving multiple vRNP segments can be taken as a measure of spatial dependence of primary segments on secondary segments. vRNP segments that truly interact are expected to be highly dependent either unidirectionally or bidirectionally. The cross validated model likelihoods can be used to weight groups of edges in the graph. Since the graph also has edges that were not observed in actual images (eg: any complex of more than 4 vRNAs), composite or inferred likelihoods were generated through recursion in a dynamic programming scheme. For each unobserved complex, we took the most likely set of observed models that could generate the complex.

For a given unobserved complex $N = \{N_1, N_2, N_3, \dots\}$,

$$L(N) = \max(L(N_1|N_2, N_3, \dots), L(N_2|N_1, N_3, \dots), \dots)$$

Then for some likelihood in the above, $L(N_1|N_2)$ where $N_1 = (n^1_1, \dots, n^1_k)$ and $N_2 = (n^2_1, \dots, n^2_l)$, $|N_1| = k$, $|N_2| = l$,

$$L((N_1|N_2)) = \sum_{i=1}^k \text{average}_{N_j^2 \in \text{all unique sets of } N_2} (L(n^1_i|N_j^2))$$

Interaction network stability analysis

To assess the stability of the generated interaction network, two procedures of perturbation were performed: introduction of noise and removing certain high likelihood models. We simulated noise as a normal random distribution centered at 0 with varying standard deviations, termed the noise levels. Under increasing noise levels, each observed model likelihood was amended with a noise value drawn from this distribution prior to interaction network construction. For each noise level, we simulated 100 trials and tallied the proportion of output trees that contained each possible edge. Trees that align with the non-noisy assembly signal stability while highly variable trees signal instability.

Reproducible research archive

All images, derived data and source code are available at <http://murphylab.cbd.cmu.edu/software>.

Supporting information

S1 Table. FISH imaging probe sets. The leftmost column is the experiment label with the right 4 columns showing which vRNA segments were tagged. Each experiment also used DAPI staining.

(CSV)

S2 Table. Observed triplets within the reaction set. 32 of 56 possible triplets were observed.

(CSV)

S3 Table. Cross validated log-variances for pairwise interaction models shown in Table 3.

(CSV)

S4 Table. Variance in the cross validated negative log-likelihoods for pairwise interaction models with cell-to-nuclear distance ratio shown in Table 4. The row label indicates the independent variable, the vRNP upon which a model predicting the column label (dependent

variable) was constructed.
(CSV)

S5 Table. Cross validated log-likelihoods for all observed triplets.
(CSV)

S6 Table. Cross validated log-likelihoods for all observed quadruplets.
(CSV)

Author Contributions

Conceptualization: Robert F. Murphy, Seema S. Lakdawala.

Funding acquisition: Robert F. Murphy, Seema S. Lakdawala.

Methodology: Timothy D. Majarian.

Software: Timothy D. Majarian.

Supervision: Robert F. Murphy, Seema S. Lakdawala.

Writing – original draft: Timothy D. Majarian, Robert F. Murphy, Seema S. Lakdawala.

Writing – review & editing: Timothy D. Majarian, Robert F. Murphy, Seema S. Lakdawala.

References

1. Morens DM, Taubenberger JK. Pandemic influenza: certain uncertainties. *Rev Med Virol.* 2011; 21(5):262–84. Epub 2011/06/28. <https://doi.org/10.1002/rmv.689> PMID: 21706672.
2. Sidorenko Y, Reichl U. Structured model of influenza virus replication in MDCK cells. *Biotechnol Bioeng.* 2004; 88(1):1–14. <https://doi.org/10.1002/bit.20096> PMID: 15384040.
3. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. *J Gen Virol.* 2010; 91(Pt 2):313–28. Epub 2009/12/04. <https://doi.org/10.1099/vir.0.017608-0> PMID: 19955561.
4. Nakatsu S, Sagara H, Sakai-Tagawa Y, Sugaya N, Noda T, Kawaoka Y. Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *MBio.* 2016; 7(5). <https://doi.org/10.1128/mBio.01248-16> PMID: 27601575; PubMed Central PMCID: PMC5013298.
5. Lee N, Le Sage V, Nanni AV, Snyder DJ, Cooper VS, Lakdawala SS. Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* 2017; 45(15):8968–77. <https://doi.org/10.1093/nar/gkx584> PMID: 28911100.
6. Noda T, Sagara H, Yen A, Takada A, Kida H, Cheng RH, et al. Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature.* 2006; 439(7075):490–2. Epub 2006/01/27. <https://doi.org/10.1038/nature04378> PMID: 16437116.
7. Chou YY, Vafabakhsh R, Doganay S, Gao Q, Ha T, Palese P. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proc Natl Acad Sci U S A.* 2012; 109(23):9101–6. Epub 2012/05/02. <https://doi.org/10.1073/pnas.1206069109> PMID: 22547828; PubMed Central PMCID: PMC3384215.
8. Noda T, Kawaoka Y. Packaging of influenza virus genome: robustness of selection. *Proc Natl Acad Sci U S A.* 2012; 109(23):8797–8. Epub 2012/05/23. <https://doi.org/10.1073/pnas.1206736109> PMID: 22615368; PubMed Central PMCID: PMC3384147.
9. Isel C, Munier S, Naffakh N. Experimental Approaches to Study Genome Packaging of Influenza A Viruses. *Viruses.* 2016; 8(8). <https://doi.org/10.3390/v8080218> PMID: 27517951; PubMed Central PMCID: PMC5013298.
10. Fujii Y, Goto H, Watanabe T, Yoshida T, Kawaoka Y. Selective incorporation of influenza virus RNA segments into virions. *Proc Natl Acad Sci U S A.* 2003; 100(4):2002–7. <https://doi.org/10.1073/pnas.0437772100> PMID: 12574509; PubMed Central PMCID: PMC149948.
11. Marsh GA, Rabadan R, Levine AJ, Palese P. Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. *J Virol.* 2008; 82(5):2295–304. <https://doi.org/10.1128/JVI.02267-07> PMID: 18094182; PubMed Central PMCID: PMC2258914.

12. Essere B, Yver M, Gavazzi C, Terrier O, Isel C, Fournier E, et al. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc Natl Acad Sci U S A*. 2013; 110(40): E3840–8. <https://doi.org/10.1073/pnas.1308649110> PMID: 24043788; PubMed Central PMCID: PMC3791739.
13. Gerber M, Isel C, Moules V, Marquet R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol*. 2014; 22(8):446–55. <https://doi.org/10.1016/j.tim.2014.04.001> PMID: 24798745.
14. Fournier E, Moules V, Essere B, Paillart JC, Sirbat JD, Cavalier A, et al. Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine*. 2012; 30(51):7359–67. Epub 2012/10/16. <https://doi.org/10.1016/j.vaccine.2012.09.079> PMID: 23063835.
15. Fournier E, Moules V, Essere B, Paillart JC, Sirbat JD, Isel C, et al. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res*. 2012; 40(5):2197–209. Epub 2011/11/15. <https://doi.org/10.1093/nar/gkr985> PMID: 22075989; PubMed Central PMCID: PMC3300030.
16. Gavazzi C, Isel C, Fournier E, Moules V, Cavalier A, Thomas D, et al. An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res*. 2013; 41(2):1241–54. Epub 2012/12/12. <https://doi.org/10.1093/nar/gks1181> PMID: 23221636; PubMed Central PMCID: PMC3553942.
17. Gilbertson B, Zheng T, Gerber M, Printz-Schweigert A, Ong C, Marquet R, et al. Influenza NA and PB1 Gene Segments Interact during the Formation of Viral Progeny: Localization of the Binding Region within the PB1 Gene. *Viruses*. 2016; 8(8). <https://doi.org/10.3390/v8080238> PMID: 27556479; PubMed Central PMCID: PMC4997600.
18. Chou YY, Heaton NS, Gao Q, Palese P, Singer R, Lionnet T. Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog*. 2013; 9(5):e1003358. Epub 2013/05/15. <https://doi.org/10.1371/journal.ppat.1003358> PMID: 23671419; PubMed Central PMCID: PMC3649991.
19. Lakdawala SS, Wu Y, Wawrzusin P, Kabat J, Broadbent AJ, Lamirande EW, et al. Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog*. 2014; 10(3):e1003971. <https://doi.org/10.1371/journal.ppat.1003971> PMID: 24603687; PubMed Central PMCID: PMC3946384.
20. Johnson GR, Li J, Shariff A, Rohde GK, Murphy RF. Automated Learning of Subcellular Variation among Punctate Protein Patterns and a Generative Model of Their Relation to Microtubules. *PLoS Comput Biol*. 2015; 11(12):e1004614. <https://doi.org/10.1371/journal.pcbi.1004614> PMID: 26624011; PubMed Central PMCID: PMC4704559.
21. Li Y, Majarian TD, Naik AW, Johnson GR, Murphy RF. Point process models for localization and interdependence of punctate cellular structures. *Cytometry A*. 2016; 89(7):633–43. <https://doi.org/10.1002/cyto.a.22873> PMID: 27327612.
22. Griffie J, Shlomovich L, Williamson DJ, Shannon M, Aaron J, Khuon S, et al. 3D Bayesian cluster analysis of super-resolution data reveals LAT recruitment to the T cell synapse. *Sci Rep*. 2017; 7(1):4077. <https://doi.org/10.1038/s41598-017-04450-w> PMID: 28642595; PubMed Central PMCID: PMC5481387.
23. Yunta ML, Lagache T, Santi-Rocca J, Bastin P, Olivo-Marin JC, editors. A statistical analysis of spatial clustering along cell filaments using Ripley's K function. 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); 2014 April 29 2014-May 2 2014.
24. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22(22):4673–80. PMID: 7984417
25. Li C, Hatta M, Watanabe S, Neumann G, Kawaoka Y. Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *J Virol*. 2008; 82(23):11880–8. <https://doi.org/10.1128/JVI.01445-08> PMID: 18815312; PubMed Central PMCID: PMC2583690.
26. Te Velthuis AJ, Fodor E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat Rev Microbiol*. 2016; 14(8):479–93. <https://doi.org/10.1038/nrmicro.2016.87> PMID: 27396566; PubMed Central PMCID: PMC4966622.
27. Phipps KL, Marshall N, Tao H, Danzy S, Onuoha N, Steel J, et al. Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny. *J Virol*. 2017; 91(17). <https://doi.org/10.1128/JVI.00830-17> PMID: 28637755; PubMed Central PMCID: PMC5553182.
28. Noda T, Sugita Y, Aoyama K, Hirase A, Kawakami E, Miyazawa A, et al. Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat Commun*. 2012; 3:639. Epub 2012/01/26. <https://doi.org/10.1038/ncomms1647> PMID: 22273677; PubMed Central PMCID: PMC3272569.

29. Brooke CB. Population Diversity and Collective Interactions during Influenza Virus Infection. *J Virol.* 2017; 91(22). Epub 2017/09/01. <https://doi.org/10.1128/JVI.01164-17> PMID: 28855247; PubMed Central PMCID: PMC5660503.
30. Illian JB, Soerbye S, Rue H & Hendrichsen D. Using INLA to fit a complex point process model with temporally varying effects—a case study *Journal of Environmental Statistics.* 2012; 3(7).
31. Lakdawala SS, Fodor E, Subbarao K. Moving On Out: Transport and Packaging of Influenza Viral RNA into Virions. *Annu Rev Virol.* 2016; 3(1):411–27. <https://doi.org/10.1146/annurev-virology-110615-042345> PMID: 27741407.
32. Bruce EA, Digard P, Stuart AD. The Rab11 pathway is required for influenza A virus budding and filament formation. *J Virol.* 2010; 84(12):5848–59. Epub 2010/04/02. <https://doi.org/10.1128/JVI.00307-10> PMID: 20357086; PubMed Central PMCID: PMC2876627.
33. Einfeld AJ, Kawakami E, Watanabe T, Neumann G, Kawaoka Y. RAB11A is essential for transport of the influenza virus genome to the plasma membrane. *J Virol.* 2011; 85(13):6117–26. Epub 2011/04/29. <https://doi.org/10.1128/JVI.00378-11> PMID: 21525351; PubMed Central PMCID: PMC3126513.
34. Nturibi E, Bhagwat AR, Coburn S, Myerburg MM, Lakdawala SS. Intracellular Colocalization of Influenza Viral RNA and Rab11A is Dependent upon Microtubule Filaments. *J Virol.* 2017; 91(19). <https://doi.org/10.1128/JVI.01179-17> PMID: 28724771.
35. de Castro Martin IF, Fournier G, Sachse M, Pizarro-Cerda J, Risco C, Naffakh N. Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles. *Nat Commun.* 2017; 8(1):1396. <https://doi.org/10.1038/s41467-017-01557-6> PMID: 29123131; PubMed Central PMCID: PMC5680169.
36. Lin JR, Fallahi-Sichani M, Sorger PK. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat Commun.* 2015; 6:8390. <https://doi.org/10.1038/ncomms9390> PMID: 26399630; PubMed Central PMCID: PMC4587398.
37. Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence.* 2002; 24(5):603–19.
38. Baddeley A, Turner R. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics.* 2000; 42(3):283–322.