Trevor Bedford, PhD
Assistant Member
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

I have no financial relationships with commercial interests to disclose
My presentation does not include discussion of off-label or investigational use
Forecasting the Future of Flu

Trevor Bedford (@trvrb)
28 Aug 2016
Options IX
Chicago, IL
What would it look like to have forecast for flu?
10-Day Weather Forecast

<table>
<thead>
<tr>
<th>Thu 08/25</th>
<th>Fri 08/26</th>
<th>Sat 08/27</th>
<th>Sun 08/28</th>
<th>Mon 08/29</th>
<th>Tue 08/30</th>
<th>Wed 08/31</th>
<th>Thu 09/01</th>
<th>Fri 09/02</th>
<th>Sat 09/03</th>
</tr>
</thead>
<tbody>
<tr>
<td>80°</td>
<td>66°</td>
<td>79°</td>
<td>69°</td>
<td>82°</td>
<td>69°</td>
<td>82°</td>
<td>69°</td>
<td>81°</td>
<td>68°</td>
</tr>
</tbody>
</table>

- Partly Cloudy
- Chance of a Thunderstorm
- Clear
- Partly Cloudy
- Partly Cloudy
- Partly Cloudy
- Chance of a Thunderstorm
- Clear
- Partly Cloudy
- Chance of a Thunderstorm

Temperature [°F]

- Chance of Precip. [%]
- Chance of Snow [%]
- Pressure [in]
- Wind Speed
Chance of winning

Hillary Clinton: 80.6%
Donald Trump: 19.4%

FiveThirtyEight
Want to forecast the make up of the future flu population from the population that exists today

Real-time updates as new information rolls in
Population turnover (in H3N2) is extremely rapid
C lades emerge, die out and take over
Clades show rapid turnover
Dynamics driven by antigenic drift
D rift variants emerge and rapidly take over in the virus population

This causes the side effect of evading existing vaccine formulations
D rift necessitates vaccine updates
H3N2 vaccine updates occur every ~2 years.
Timely surveillance and rapid analysis is essential to vaccine strain selection
nextflu

Project to provide a real-time view of the evolving influenza population
Project to provide a real-time view of the evolving influenza population

All in collaboration with Richard Neher
The `nextflu` pipeline:

1. Download all recent HA sequences from GISAID
2. Filter to remove outliers
3. Subsample across time and space
4. Align sequences
5. Build tree
6. Estimate clade frequencies
7. Infer antigenic phenotypes
8. Export for visualization
Up-to-date analysis is publicly available at:

nextflu.org
Antigenic analysis
Influenza hemagglutination inhibition (HI) assay

Hemagglutination assay:

\[
\text{Without virus, red blood cell sink to bottom of well}
\]

\[
\text{With virus, cells form diffuse lattice}
\]

Hemagglutination inhibition assay:

\[
\text{Without antibodies, agglutination of virus to RBC}
\]

\[
\text{Antibodies bind viruses, preventing agglutination}
\]
HI measures cross-reactivity across viruses

Without antibodies, agglutination of virus to RBC

Antibodies bind viruses, preventing agglutination

Reacting virus from strain A-H vs sera from strain A
Data in the form of table of maximum inhibitory titers

<table>
<thead>
<tr>
<th>Viruses</th>
<th>Collection Date</th>
<th>Passage History</th>
<th>Haemagglutination inhibition titre</th>
<th>Post infection ferret sera</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A/Wis 67/05 A/Bris 10/07 A/Uru 716/07 A/HK 1985/09 A/Perth 16/09 A/Wis 15/09 A/HK 34430/09</td>
<td></td>
</tr>
<tr>
<td><strong>REFERENCE VIRUSES</strong></td>
<td></td>
<td></td>
<td>F1/06 F29/08 F26/08 F21/09 F25/09 F24/09 F4/10</td>
<td></td>
</tr>
<tr>
<td>A/Wisconsin/67/2005</td>
<td>2005-08-31</td>
<td>SpfCk3E3/E7</td>
<td>1280 1280 1280 40 &lt; &lt; 160 40</td>
<td></td>
</tr>
<tr>
<td>A/Brisbane/10/2007</td>
<td>2007-02-06</td>
<td>E2/E3</td>
<td>2560 2560 2560 80 &lt; &lt; 160 160</td>
<td></td>
</tr>
<tr>
<td>A/Uruguay/716/2007</td>
<td>2007-06-21</td>
<td>SpfCk1, E3/E3</td>
<td>640 1280 2560 &lt; &lt; &lt; 80 40</td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/1985/2009</td>
<td>2009-04-01</td>
<td>MDCK2/SIAT1</td>
<td>40 80 160 1280 640 2560 1280</td>
<td></td>
</tr>
<tr>
<td>A/Wisconsin/15/2009</td>
<td>2009-07-06</td>
<td>E2/E3</td>
<td>&lt; &lt; 40 640 640 1280 1280</td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/34430/2009</td>
<td>2009-11-22</td>
<td>MDCK2/SIAT2</td>
<td>&lt; 80 160 5120 640 1280 1280</td>
<td></td>
</tr>
<tr>
<td><strong>TEST VIRUSES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/1737/2010</td>
<td>2010-03-24</td>
<td>MDCK2/SIAT1</td>
<td>40 80 320 5120 1280 1280 1280</td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/1775/2010</td>
<td>2010-03-28</td>
<td>MDCK2/SIAT1</td>
<td>&lt; 80 160 5120 640 2560 1280</td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/1837/2010</td>
<td>2010-03-30</td>
<td>MDCK2/SIAT1</td>
<td>40 80 160 5120 640 2560 1280</td>
<td></td>
</tr>
<tr>
<td>A/Hong Kong/1888/2010</td>
<td>2010-04-19</td>
<td>MDCK2/SIAT1</td>
<td>160 320 320 5120 1280 2560 2560</td>
<td></td>
</tr>
</tbody>
</table>

NIMR 2010
Antigenic cartography compiles measurements into an interpretation.
Instead of a geometric model, we sought a phylogenetic model of HI titer data.
Identify phylogeny branches associated with drops in HI titer

<table>
<thead>
<tr>
<th>Post infection ferret sera</th>
<th>A/Bris 10/07</th>
<th>A/Uru 716/07</th>
<th>A/HK 1985/09</th>
<th>A/Perth 16/09</th>
</tr>
</thead>
<tbody>
<tr>
<td>F29/08</td>
<td>2560</td>
<td>2560</td>
<td>80</td>
<td>&lt;</td>
</tr>
<tr>
<td>F26/08</td>
<td>1280</td>
<td>2560</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>F21/09</td>
<td>80</td>
<td>160</td>
<td>1280</td>
<td>640</td>
</tr>
</tbody>
</table>

**REFERENCE VIRUSES**

- A/Brisbane/10/2007: 2560
- A/Uruguay/716/2007: 1280
- A/Hong Kong/1985/2009: 80
- A/Perth/16/2009: <

Source: Neher et al 2016
Model can be used to interpolate across tree and predict phenotype of untested viruses

Measurements

Model

Neher et al 2016
Model is highly predictive of missing titer values

**tree model**
- Regularization: 1.0/0.3/2.0 (HI/pot/avi)
- Prediction error: 0.54/0.73 (abs/rms)
Recent HI data from London WHO Collaborating Center
Up-to-date analysis at:

nextflu.org
Phylogeny

2016 Aug 26

log₂ titer distance from A/HongKong/146/2013

Color by HI distance from
Or HA1 positions...

HI color: measurements
          tree model
          substit. model

Corrections:
            serum potency
            virus avidity

Sera:
            F40/13
            F10/15
Forecasting
"The future is here, it's just not evenly distributed yet" — William Gibson
USA music industry, 2011 dollars per capita
Influenza population turnover
Vaccine strain selection timeline

Collection by WHO National Influenza Centres

Characterization by WHO Collaborating Centres

- Virus isolation
- Strain selection
- High growth reassortants
- Manufacture
- Licensure
- Packaging
- Distribution
- Vaccination

Timeline:
Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov
Seek to explain change in clade frequencies over 1 year
Fitness models can project clade frequencies

Clade frequencies $\mathbf{X}$ derive from the fitnesses $\mathbf{f}$ and frequencies $\mathbf{x}$ of constituent viruses, such that

$$\hat{X}_v(t + \Delta t) = \sum_{i:v} x_i(t) \exp(f_i \Delta t)$$

This captures clonal interference between competing lineages

Luksza and Lässig 2014
The question of forecasting becomes: how do we accurately estimate fitnesses of circulating viruses?
Fortunately, there's lots of training data and previously successful strains have had:

1. Amino acid changes at epitope sites
2. Antigenic novelty based on HI
3. Rapid phylogenetic growth
Predictor: calculate HI drop from ancestor, drifted clades have high fitness
Predictor: project frequencies forward, growing clades have high fitness
We predict fitness based on a simple formula

where the fitness $f$ of virus $i$ is estimated as

$$f_i^{\hat{}} = \beta^{HI} f_i^{HI} + \beta^{freq} f_i^{freq}$$

where $f_i^{HI}$ measures antigenic drift via HI and $f_i^{freq}$ measures clade growth/decline

We learn coefficients and validate model based on previous 15 H3N2 seasons
Clade growth rate is well predicted ($\rho = 0.66$)
Growth vs decline correct in 84% of cases
Trajectories show more detailed congruence
Trajectories show more detailed congruence
This model is similar in formulation and performance to Łuksza and Lässig
When does the forecast fail?

Emerging clades are difficult to forecast: little antigenic data and little evidence of "past performance"

Models work well for clades at >10%, but less well for clades <5%
New mutations difficult

Models can project forward from circulating strains, but cannot foresee the appearance of new mutations

Intrinsically limits the timescale of forecasting to ~1 year
Model is only as good as the data

Requires rapid shipping of samples, rapid sequencing and rapid antigenic characterization
Current projections
Issuing reports online in Feb and Sep

Seasonal influenza circulation patterns and future projections
February 18, 2016

This is not meant as a comprehensive report, but is instead intended as particular observations that we’ve made that may be of relevance. Please also note that observed patterns reflect the GiSAID database and may not be entirely representative of underlying dynamics.

A/H3N2

We expect clade 3c2.a viruses to continue to predominate over 3c3.a and 3c3.b viruses. Diversity within 3c2.a is beginning to accrue with the emergence of several clades of appreciable frequency, the most notable of which is the new 171K clade that has very recently risen to high frequency.

We base our primary analysis on a set of viruses collected between Feb 2014 and Jan 2016, comprising approximately 100 viruses per month where available and seeking to equilibrate sample counts geographically where possible. This equilibration attempts to collect equal samples from Africa, China, Europe, Japan/South Korea, North America, Oceania, South America, South Asia, Southeast Asia and West Asia. In the following analysis we collapse samples from China,
In February we stated

"Barring substantial changes in other clades, we predict the (HA1:171K, HA2:77V/155E) variant to dominate"

Let's see how we did
The (H3N2) world today
3c2.a viruses continue to predominate (except in the USA)
Within 3c2.a clades are emerging, in particular (HA1:171K, HA2:77V/155E)
The 171K clade has recently risen in frequency
We predict the 171K clade will continue to be successful (unless supplanted by a novel mutant)

- Has an amino acid change at predicted epitope site
- Has some evidence of antigenic novelty based on HI
- Shows recent rapid expansion
Looking forward
Further improvements to predictive modeling

1. Extend to other seasonal viruses
2. Forecast NA evolution
3. Integrate neutralization (FRA) assay data
4. Model effects of egg adaptation
5. Incorporate an explicit geographic model
Phylogeny of H3 with geographic history

- USA & Canada
- South America
- Europe
- India
- North China
- South China
- Japan & Korea
- Southeast Asia
- Oceania

Bedford et al. 2015
Geographic location of phylogeny trunk

Bedford et al. 2015
More generally real-time analyses may be useful for other viruses
Ebola at ebola.nextstrain.org
Zika at [nextstrain.org/zika/](http://nextstrain.org/zika/)
Major opportunity to track evolution non-human influenza viruses
Adaptation to avian flu viruses by Yukia Zhou and Justin Bahl
All tools are completely open source and we encourage other groups to get involved and push the project forward.
General purpose genomic surveillance tool
Acknowledgements

Analysis: Richard Neher, Colin Russell, Charlton Callender, Colin Megill, Andrew Rambaut, Charles Cheung, Marc Suchard, Steven Riley, Philippe Lemey, Gytis Dudas, Boris Shraiman, Marta Łuksza, Michael Lässig

GISRS: Ian Barr, Shobha Broor, Mandeep Chadha, Nancy Cox, Rod Daniels, Becky Garten, Palani Gunasekaran, Aeron Hurt, Anne Kelso, Jackie Katz, Nicola Lewis, Xiyan Li, John McCauley, Takato Odagiri, Varsha Potdar, Yelong Shu, Eugene Skepner, Masato Tashiro, Dayan Wang, Dave Wentworth, Xiyan Xu
Contact

- Website: bedford.io
- Twitter: @trvrb
- Slides: bedford.io/talks/forecasting-flu-options/