

A 'Local-Global' model for Seasonal Diseases: Influenza Subtypes

Analysis Case Study

Keywords: Influenza epidemiology; epidemics; pandemics

Running title: Local-global model for the transmission of influenza

Authors: Gal Almogy, Flurensics, Tel Aviv, Israel

email: gal.almogy@gmail.com

Abstract

Influenza epidemics in temperate regions display dynamics that are characterized by pronounced seasonal peaks during the winter. The general lack of influenza cases during the off-season may result from the virus physically disappearing at the end of the season, in which case it must be imported annually. Alternatively, it may result from persistent asymptomatic carriers or unnoticed local transmission chains that develop into local epidemics as conditions become conducive. Here I attempt to understand these differing explanations by analyzing the global distribution of the four major subtypes that comprise influenza over a period of 18 years based on FluNet data, the surveillance network and database compiled by the WHO, and the NCBI influenza data resource, a repository of relevant genetic information.

Examining the annual proportion of each subtype, I find considerable variations in subtype annual proportions between the regions. Moreover, I find that seasonal influenza subtypes can remain confined to specific temperate regions, without showing measurable global presence. These results indicate that although largely undetected during the off-season, influenza is likely to persist locally, and imply a ‘local-global’ model where annual influenza epidemics are a mixture of local strains undergoing reactivation together with an influx of global variants.

Introduction

Influenza virus infections are among the leading causes of winter morbidity and mortality, resulting in an estimated 40,000 deaths annually in the United States alone and up to ~500K globally^{1,2}. Influenza epidemics in temperate climates exhibit distinctive seasonal dynamics, characterized by a peak in disease incidence over the winter months, but only sporadic disease cases are detected during the rest of the year³⁻⁹. Off-season epidemics have been observed in temperate regions, but these are not common and mostly confined to narrow geographic settings, e.g. correctional facilities¹⁰, military bases¹¹, cruise-ships¹² and long-term care facilities¹³.

The low disease incidence during the off-season may indicate that influenza in temperate regions physically goes extinct at the end of each winter season, and that consequently influenza must be imported annually via ‘carriers’ travelling from source regions. Phylogenetic studies focused on the long-term spatial-temporal dynamics of influenza support this type of model (termed here a ‘strictly global model’), either with the semi-tropical southeast-Asia region as a perennial source^{14,15}, or as has been suggested more recently, influenza lineages are dynamically maintained through time and space via complex migration networks, essentially persisting in no population for extended periods^{3,16}.

However, increased influenza surveillance following the 2009 H1N1pdm pandemic revealed higher than previously assumed off-season incidence rates^{17,18}. For example, active transmission chains of the H3N2 influenza subtype were detected in rural areas of New York, USA (latitude: 40N) from March to June 2012, at which point¹⁹.

Moreover, persistent ‘local influenza lineages’ have been identified in subtropical

Africa²⁰, suggesting that at least in some regions long-term influenza persistence is possible.

Here I postulate that if annual epidemics in temperate regions *absolutely* require the introduction of strains from global sources (i.e. a ‘strictly global model’), then the annual subtype proportions in temperate/sink regions will tend to be similar to, and correlate well with the proportions in the source region(s). Conversely, if local persistence is not sustainable or does not occur, then if a subtype is absent from a source region during a season, it should not be detected during that season in any of the temperate (sink) regions.

I studied the annual distribution pattern of the influenza subtypes most common during the period from 1999 to 2017 (classified by their HA (H) and NA (N) genes⁸): H3N2, H1N1, H1N1pn and influenza type B, FluB, using data compiled by the World Health Organization's FluNet²¹. The data comprises over 2M data-points representing systematic influenza surveillance reports from 33 countries spanning four large regions: Asia, Europe, North America, and ‘southern hemisphere’ countries from and South America, South Africa and Oceania (figure 1 and table 1).

I found that the annual influenza subtype distributions in different geographical regions were poorly correlated, and homogeneous in only 4 out of the 18 seasons examined. Notably, two of these homogeneous seasons coincided with the emergence of novel influenza strains capable of rapid global spread, H3N2 Fujian (2003-04)²² and H1N1pn^{23,24} (2009-10), and in 2000-01 the dominant subtype in US and Europe was, atypically, of an H1N1 strain (A/New Caledonia/20/99 strain²⁵. Furthermore, I

find that the H1N1 subtype was detected almost exclusively in the US and Europe from 2002 to 2005, providing positive evidence in support of longer-term local persistence^{26,27}.

Finally, I add perspective by repeating the analysis using data derived from the NCBI influenza database²⁸, which is the primary resource used in corresponding phylogenetic studies. Importantly, I did not find evidence of neither a global pattern nor of local persistence in this dataset.

These results suggest that a dynamic mixture of global and local sources shape the subtype-composition of annual influenza epidemics, and specifically, that the influenza virus can persist locally in temperate regions. Furthermore, a comparison of data from the two main influenza resources, FluNet and NCBI, showed that although lacking sequence information, the wealth of systematically collected surveillance data (FluNet) could shed light on questions that usually are addressed via phylogenetic analysis.

Results

Expected subtype proportions in strictly global/local models

I consider two hypothetical models that can explain influenza long-term dynamics:

The first is a strictly global model, where all annual influenza epidemics (of any subtypes) are introduced into temperate regions via external carriers, either from one particular source or from multiple sources. The second is a strictly local model, where the source for all influenza epidemics is assumed to be local.

The annual proportions of each influenza subtype are expected to differ considerably between these models. In a strictly local model, the subtype proportions in different regions should be entirely independent and poorly correlated. However, pandemics, as well as numerous molecular studies^{2,7,29,30} prove that this is almost certainly not generally the case and I shall not examine this possibility further.

In a strictly global model influenza dynamics in temperate regions are expected to critically depend on the volume of human transport between sinks and source(s): lower volumes of transport could lead to increasingly asynchronous epidemics, or even entire seasons where influenza is absent from some or even all the temperate regions (sinks). Conversely, higher volumes of transport, and by extension a non-limiting influx of influenza carriers are expected to ensure that epidemics of any subtype occurring in a source(s) will be continuously introduced to temperate regions. In which case, epidemics in sink regions could initiate whenever the local conditions become favorable, and hence epidemics in such regions, which also share the same

seasonality pattern, are expected to begin at roughly the same time, and comprise similar subtype proportions, mimicking the contemporary proportions at the source.

I compared the timing of influenza epidemics across the four regions (figure 2) by performing a cross-correlation analysis of influenza incidence. Results indicate virtually no time lag between influenza epidemics in Asia and the other northern hemisphere regions (figure 2a-c). In the context of a global model, this implies that the limiting step in influenza epidemics is the season, i.e., ‘influenza seeds’ are constantly introduced to the regions examined, but epidemics occur only when environmental conditions permit effective virus transmission, a conclusion consistent with results from previous studies ^{3,5,31}.

The southern hemisphere incidence data lags behind the other regions by approximately 5 weeks (figure 2d-e), after taking into account an arbitrary 26-weeks shift between the winter season in the northern and southern hemispheres. However, it is unlikely that transport rate is the limiting factor even in this case, as the literature shows. For example, it was found ¹⁴ that a majority of the various influenza lineages circulating in the New York winter can be detected at the same time in the summer of the very distant Christchurch, New Zealand.

Influenza subtypes annual distribution by region

In a global model, given non-limiting transport rates, subtypes should be represented homogeneously in sink and source regions, leading to roughly similar annual subtype proportions in connected regions, and consequently subtype proportions should correlate between these regions.

Such a uniform pattern of annual subtype proportions however was rarely evident in the data. The annual proportions of each subtype are represented by color (blue = 0, red = 1) in figure 3a (rows = regions, columns = seasons). If regions were homogeneous in terms of subtype proportions, vertical columns in figure 3a should be uniform in color. Thus, figure 3a shows the annual proportions vary considerably between the regions since they do not show a vertical (seasonal) uniform subtype distribution pattern.

Consistently, the associated annual correlation values are low for all subtypes examined (figure 3b). The mean global correlation scores for H1N1, H3N2 and Flu type B were 0.3, 0.38 and 0.33, respectively. The H1N1pn strain showed only slightly higher levels of correlation (mean = 0.6), despite the strain's recent pandemic origins

23

Globally homogeneous periods

Within the general lack of a homogeneous pattern in the annual subtype proportions (figure 3a), several seasons showed a visually distinct, vertical distribution pattern characterized by one subtype present at unusually elevated proportions in all regions (most notably, H1N1pdm during 09-10). To investigate this observation further, I define a subtype as 'Highly Represented' (HR) if it is detected at significantly higher proportion than its annual average in a region (for definition and statistical criteria determining when a subtype is highly represented see Eq. 2). If a subtype is highly represented (HR) in all regions during a season, I define it as Highly Represented Globally (HR_G, Eq. 3 in Methods).

During the period examined four seasons, 00-01, 03-04, 09-10, and 15-16 were identified as HR_G, for subtypes H1N1, H3N2, H1N1pn and FluB respectively (seasons indicated in figure 4). Two of these seasons (03-04 and 09-10) are known to be associated with the emergence of highly transmissible, antigenically novel influenza variants capable of rapid global spread^{14,22,27,32,33}, and the 00-01 season showed an unusually high level of H1N1 activity in many regions, notably in Europe and the US^{25,26,34}.

Note the probability of four HR_G seasons arising by chance (i.e. in the absence of global dynamics) is $p < 0.01$, strongly suggesting most if not all these HR_G seasons were the result of a particular strain reaching rapid global dominance.

Limited local dominance

Intuitively, influenza variants can become highly represented in a region (HR in Eqn.2) by gaining a competitive advantage, e.g. through mutations and genomic re-assortment events^{1,14,15,35-39}. In the context of a strictly global model, the bulk of such virus evolution occurs at the source regions, whereas strains in sink regions are close relatives of contemporary strains at the sources^{16,35}. Hence, it is expected that strains that become HR emerge in a source region i (HR_i), and from the source spread to connected sink regions. If the source is sufficiently well connected then these unusually fit strains are expected to spread globally and become HR_G (Eqn.3).

However, in the majority of seasons examined, at least one region was classified as HR (figure 4) for a subtype, yet the same subtype was not HR in other regions. This

observation is consistent with the low correlation in subtype proportions, but is difficult to explain within a strictly global context, whether the model assumes a single source region, or multiple source networks.

Local persistence

Another prediction that follows from assuming a strictly global model is that since influenza epidemics in temperate regions completely die out annually, then any subtype detected in any temperate region must also be detected in at least one source region.

However, a careful analysis of the country incidence data revealed a counterexample where the H1N1 subtype was detected in temperate regions over several seasons without an ostensible source elsewhere. H1N1 was prevalent in Asia (most notably Japan) between 1999 and 2001, 65, 50, and 40% of all cases in Asia, respectively (figure 5, orange bars). However, during the following 3 seasons (02-04) it was almost absent from Asia (~0, ~0 and 5%) yet appeared in substantial numbers in Europe and the US during 01-02 (e.g. France, Italy and the US), 02-03 (e.g. Italy and France) and 03-04 (all over Europe).

Note that the H1N1 vaccine component remained unchanged (A/New Caledonia/20/99(H1N1)) between 2001 and 2005³⁴, consistent with little genetic evolution of the virus during the period, which would be expected given geographically limited, local persistence^{40,41}.

Subtype distribution in the context of influenza phylogenetic studies

Our results point to dynamics in which globally circulating strains can have a dramatic impact on regional epidemics, at times overshadowing any local contributions, and even replacing existing subtypes altogether, e.g. the annual H1N1 subtype was replaced by H1N1pdm over the 2009-10 season ²³.

However, while phylogenetic studies point to a general lack of local influenza persistence ^{3,16,18}, our results suggest there are both local and global sources for annual epidemics, i.e. persistent local sources contribute measurably to the proportions of influenza subtypes in annual epidemics. I explored the possibility that data collection methodology can be the source for this discrepancy by repeating the analysis using the NCBI influenza database ²⁸.

The NCBI database contains date and locations, as well as the genetic information, which is of great utility in phylogenetic analyses. However, while the data is more detailed than that supplied by FluNet (which does not contain sequence information), it is limited in the total number of samples available for the period examined, and may be less standard in sample submission methodology. Note in context that the FluNet data analyzed here comprises >2M data points, compared to ~50K data points in the NCBI data (compare table 1 (FluNet) and table 2 (NCBI)).

As with the FluNet data, annual subtype proportions were overall heterogeneous and showed only little correlation between annual proportions across regions (figure 6a, b). (Strong correlations would be indicated by yellow or red shading in figure 6b.)

Remarkably however, in the NCBI data no season was HR_G for the proportions of any subtype (figure 6c). Rather, results showed a horizontal (regional) pattern and a pronounced paucity of FluB genetic data in most countries, suggesting a potential bias in the sampling methodology, see e.g. the supposedly unusually high proportion of H3N2 in Europe between 2002 and 2005 (compare with figure 3a and 4 for FluNet data). Furthermore, there was no evidence in the data to support the localized subtype persistence pattern that was found analyzing the FluNet data (figure 5). That is, there were no 2 consecutive seasons in which samples of any subtype were detected mainly/only in temperate regions, and specifically not H1N1 in Europe.

Discussion

Annual influenza epidemics in temperate regions may result from synchronized activation of local viral reservoirs when conditions become conducive, i.e. a 'local model'. Alternatively, in a strictly global model influenza dies out at the end of each season and annual epidemics must originate from global sources, either one specific region where the virus can persist year-round, or via continuous circulation.

In the period between 1999 and 2016 influenza epidemics comprised four prominent subtypes. Treating each subtype as a distinct component of the overall influenza epidemic, I examined the per-region, annual proportion of each subtype out of the total influenza incidence. If the sources of influenza epidemics in temperate regions were *invariably* external, then the following pattern would be evident:

- The 'annual subtype proportions' should be homogeneous among the seeded regions, and correlate with the subtype proportions at the sources
- If during a season a subtype is highly-represented (HR_i) in any region i , it should be similarly highly-represented in all connected regions (HR_G if there is a single source)
- During a season, a subtype can be present in a temperate region only if also present in a source, i.e., there is no 'local persistence' in temperate regions.

These predictions do not generally bear out. First, the global distribution of the annual subtype proportions showed a large degree of heterogeneity (figure 3a), reflected also in low overall correlation between subtype proportions in the different regions (figure 3b). Second, only four of the examined seasons showed unusually elevated levels of a single subtype globally (figure 4), and importantly, two of these seasons are known

for the emergence of novel/pandemic strains (03-04 and 09-10), and while the first ‘global season’ I find (00-01, HR_G for H1N1) is not considered a ‘pandemic season’, unusually high levels of H1N1 incidence were noted in various regions, e.g. the US and Europe. However, these four ‘global pattern’ seasons were the exception and there was otherwise little association between a subtype being highly represented in one region and its proportion in other regions, despite the advantage these HR subtypes appear to possess in at least one region. (Note that the absence of HR in a region/year simply means that no subtype was detected at an unusually high frequency, relative to the average frequency of the subtype in a region.)

Finally, I found that the H1N1 subtype was confined to countries in temperate regions during a 3-year period (figure 5, seasons 2002-04), e.g. France and Italy, while showing minimal signs of activity in Asia, and only little genetic evolution during this entire period, consistent with dynamics expected given ‘persistence’ in geographically limited settings. Interestingly, while detected in large numbers in both the 01-02 and 03-04 seasons in Italy and France (and at lower frequency in 02-03), it was only irregularly detected in other European countries, e.g. UK mainly during 01-02 (see also figure S5), suggesting a high degree of localization even within the region where the virus persisted.

Taken together, these results indicate that local persistence in temperate regions does indeed occur, which implies a ‘local-global’ model, where the annual composition of influenza epidemics is a mixture of local strains undergoing reactivation together with an influx of globally circulating variants. When the competitive advantage of global variants is sufficiently large, the contribution of local sources may vanish, as indicated

by the correspondence between the emergence of antigenically advanced variants and years identified here as ‘globally homogeneous’ (HR_G). However, in the absence of such advanced variants, the presence of local strains has a non-negligible effect, which leads to regional variations in annual influenza subtype compositions.

Examining subtype incidence data derived from the NCBI database (rather than FluNet) had an important impact on the results; specifically, I did not find any ‘global seasons’ (HR_G, compare figures 4 and 6c) and there was no evidence of the local persistence pattern found in the FluNet data. It is of note that FluNet was designed purposely for broadening the understanding of global patterns in influenza spread, and provides a large set of systematically collected case samples for statistical analysis (Table 1), containing > 2 orders of magnitude more samples than in the NCBI database.

The contrast between the FluNet- and NCBI-based results may provide a clue as to why studies of influenza sequence information from roughly the same time period conclude the virus does not survive the off-season, and must be imported annually from source regions. It is quite surprising that so little consideration had been given hitherto to the relative proportions of subtypes or lineages in public health related issues, such as vaccine strain selection. For example, significant efforts have been directed at identifying a link between the genetic composition of variants and their ability to evade neutralizing antibodies^{42–44}, which are important but not sufficient parameters in determining the pandemic potential of a variant. In contrast, I am not familiar with studies that aim to identify significant deviations in the relative

proportions of subtypes, a computationally simpler and intuitively more direct measure of a strain's pandemic potential.

I note the time-series data from the southern hemisphere was arbitrarily set to lag behind that of the northern hemisphere, however, reversing this decision, i.e. having the southern hemisphere season precede the northern hemisphere, had no effect on the results (not shown). Also, the use of alternative methods for defining seasons, e.g. detecting peaks of influenza activity in the data, per country, region and globally did not affect the observed dynamics (not shown).

The type of model I propose in which global sources promote influenza diversity yet are not the only factor explaining seasonality, is consistent with several general observations. For example, it is well documented that influenza strains may be detected all year-round in non-human hosts, not only in Avian hosts, where the virus is maintained via a complex intra-species/reservoir systems^{30,45}, but also in non-migrating Equine and Swine populations where no natural reservoir had been identified⁴⁶⁻⁵¹. Interestingly, it was recently found⁴⁰ that while difficult to detect through conventional surveillance, swine influenza lineages can persist in surprisingly small host population (~500 animals).

The results presented here emphasize the importance of using multiple data sources and applying diverse analytic approaches to complex problems, and taking into consideration the knowledge of the genetic composition, as well as the relative abundance of influenza variants in the selection of seasonal flu vaccines.

Materials and Methods

FluNet data

Human influenza virus epidemics in recent years are comprised of four distinct subtypes: H1N1, H3N2, H1N1pn and influenza type B (FluB). The incidence of these subtypes per country is quantified on a massive scale as part of the FluNet international project coordinated by the World Health Organization, WHO²¹. The numbers of samples per country, from week 40 of 1999 to week 8 of 2017 are presented in table 1.

NCBI data

I retrieved HA sequence data from the NCBI influenza virus database²⁸ for the H1N1, H3N2 and H1pdm subtypes, and NA sequence data for FluB in the period from September 1999 to February 2017 (including partial sequences). Using the header information I divided the data into countries/seasons as with the FluNet data. Note this data was used only to examine subtype proportions per country/season, and without reference to the sequence data itself. The numbers of samples per country are presented in table 2.

Regions and seasons

Countries were divided into four regions based on general geographic location: Asia (Japan, China, South Korea, Thailand, Singapore, and Russia), Europe (Norway, Sweden, Denmark, Netherlands, Poland, Germany, France, UK, Finland, Switzerland, Czech, Romania, Hungary, Austria, Iceland, Ireland, Italy, Israel, Spain), North America (USA, Canada) and 'southern hemisphere' countries (Peru, Argentina, Chile,

Australia and New Zealand, and South Africa), referred to as 'south' (figure 1). Data (Tables 1 and 2) was added for each of the countries in these specific regions.

The influenza season was defined according to calendar months: the winter-season in the northern hemisphere was approximated as week 40 to week 20 in the following year and weeks 14-47 for the south, representing a 26 weeks shift forward (or backward, S3).

Subtype proportions

For year A, the per season, proportion $f_{A\alpha i}$ for each region i , i.e. the number of positive events of a given subtype α , over the overall influenza events detected, was calculated using the following:

$$Eqn.1 f_{A\alpha i} = \frac{NA_{\alpha i}}{\sum_{\alpha} NA_{\alpha i}}$$

where $NA_{\alpha i}$ is the number of influenza subtype α cases (H1N1, H3N2, FluB and H1N1pn) in region i , in season A and the sum of these is the total cases of influenza reported per region, per season.

Examining each subtype α in region i over all the seasons, I can calculate the mean and the standard deviation $\sigma_{\alpha i}$, of $f_{A\alpha i}$. I use these to define a subtype α as highly represented (HR) in season A , region i , if the subtype's frequency is greater than the mean (per region and subtype) by at least one standard deviation:

$$Eqn.2 HR_{A\alpha i} = 1 \Leftrightarrow \frac{\sum_{A=1}^{n_A} f_{A\alpha i}}{N_A} + \sigma_{\alpha i} < f_{A\alpha i}$$

for any region i in season A . Otherwise $HR_{Aai} = 0$.

A season is defined as homogeneous for a subtype α , if all regions are highly represented for the same subtype, i.e.:

$$\text{Eqn.3} \quad HR_G_{Aai} = 1 \Leftrightarrow [HR_{Aai} = 1, \forall_i]$$

$$HR_G_{Aai} = 0, \text{ otherwise.}$$

Correlations

Let $\rho(x,y)$ be the standard Pearson's correlation coefficient between two variables $x(t)$ and $y(t)$. Then the lagged cross-correlation between the variables is defined as $\rho(x(t),y(t-\tau))$, where τ is the time-lag. To quantify the epidemic timing (figure 2b) between regions I calculated the cross-correlation, over a time lag of 20 weeks, between the incidences of the total influenza (of all subtypes), per region, for the period 1999-2009, i.e., before the 2009 pandemic. To quantify the correlation between the annual subtype proportions in each region, f_{Aai} , I calculated the Pearson's correlation coefficient r with zero time-lag, over the entire period (18 seasons) for all subtypes except the H1N1pn subtype, for which I limited the calculation to begin from 2009, when it was first identified (figure 3b).

Acknowledgments

I would like to thank Lewi Stone, Andrei Bernevig, Giora Almogy and Eli Engelberg for their critical reading and helpful suggestions.

Competing financial interests

The authors declare no competing financial interests.

References

1. Cannell, J. J. *et al.* On the epidemiology of influenza. *Virol J* **5**, 422X--5 (2008).
2. Cox, N. J. & Subbarao, K. Global epidemiology of influenza: past and present. *Annu. Rev. Med.* **51**, 407–421 (2000).
3. Tamerius, J. *et al.* Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environ. Health Perspect.* **119**, 439 (2011).
4. Fisman, D. Seasonality of viral infections: mechanisms and unknowns. *Clin. Microbiol. Infect.* **18**, 946–954 (2012).
5. Cox, N. Influenza seasonality: timing and formulation of vaccines. *Bull. World Health Organ.* **92**, 311 (2014).
6. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (80-.).* **312**, 447–451 (2006).
7. Barreca, A. I. & Shimshack, J. P. Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *Am. J. Epidemiol.* **176**, S114--S122 (2012).
8. Steinhauer, D. A. & Skehel, J. J. Genetics of influenza viruses. *Annu. Rev. Genet.* **36**, 305–332 (2002).
9. Zinder, D. *et al.* Seasonality in the migration and establishment of H3N2 Influenza lineages with epidemic growth and decline. *BMC Evol. Biol.* **14**, 272 (2014).
10. Young, L. C. *et al.* Summer outbreak of respiratory disease in an Australian prison due to an influenza A/Fujian/411/2002 (H3N2)-like virus. *Epidemiol. Infect.* **133**, 107–112 (2005).
11. Cosby, M. T. *et al.* Outbreak of H3N2 influenza at a US military base in Djibouti during the H1N1 pandemic of 2009. *PLoS One* **8**, (2013).
12. Fernandes, E. G. *et al.* Influenza B Outbreak on a Cruise Ship off the São Paulo Coast, Brazil. *J. Travel Med.* **21**, 298–303 (2014).
13. Louie, J. K. *et al.* A summer outbreak of human metapneumovirus infection in a long-term-care facility. *J. Infect. Dis.* **196**, 705–708 (2007).
14. Russell, C. A. *et al.* The global circulation of seasonal influenza A (H3N2) viruses. *Science (80-.).* **320**, 340–346 (2008).

15. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008).
16. Bahl, J. *et al.* Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proc. Natl. Acad. Sci.* **108**, 19359–19364 (2011).
17. Cummings, D. A. T. & Lessler, J. Infectious disease dynamics. *Infect. Dis. Epidemiol.* 131 (2014).
18. Viboud, C., Nelson, M. I., Tan, Y. & Holmes, E. C. Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120199 (2013).
19. Ghedin, E. *et al.* Unseasonal transmission of H3N2 influenza A virus during the swine-origin H1N1 pandemic. *J. Virol.* **84**, 5715–5718 (2010).
20. Nelson, M. I. *et al.* Multi-year persistence of two pandemic A/H1N1 influenza virus lineages in West Africa. *J. Infect. Dis.* jiu047 (2014).
21. Flahault, A. *et al.* FluNet as a tool for global monitoring of influenza on the Web. *Jama* **280**, 1330–1332 (1998).
22. Finkelman, B. S. *et al.* Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One* **2**, e1296 (2007).
23. Girard, M. P., Tam, J. S., Assossou, O. M. & Kieny, M. P. The 2009 A (H1N1) influenza virus pandemic: A review. *Vaccine* **28**, 4895–4902 (2010).
24. Gog, J. R. *et al.* Spatial transmission of 2009 pandemic influenza in the US. *PLoS Comput. Biol.* **10**, e1003635 (2014).
25. (CDC), C. for D. C. and P. & Anonymous. Influenza activity--United States, 2001-02 season. *MMWR. Morb. Mortal. Wkly. Rep.* **50**, 1084–1086 (2001).
26. Paget, J., Marquet, R., Meijer, A. & van der Velden, K. Influenza activity in Europe during eight seasons (1999--2007): an evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe. *BMC Infect. Dis.* **7**, 141 (2007).
27. Centers for Disease Control and Prevention (CDC). Update: influenza activity--United States and worldwide, 2003-04 season, and composition of the 2004-05 influenza vaccine. *MMWR Morb Mortal Wkly Rep* **53**,

- 547–552 (2004).
28. Bao, Y. *et al.* The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* **82**, 596–601 (2008).
 29. Holmes, E. C. *et al.* Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300 (2005).
 30. Brown, V. L. *et al.* Neutrality, Cross-Immunity and Subtype Dominance in Avian Influenza Viruses. *PLoS One* **9**, e88817 (2014).
 31. Bloom-Feshbach, K. *et al.* Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PLoS One* **8**, e54445 (2013).
 32. Ward, K. A., Armstrong, P., McAnulty, J. M., Iwasenko, J. M. & Dwyer, D. E. Outbreaks of pandemic (H1N1) 2009 and seasonal influenza A (H3N2) on cruise ship. *Emerg. Infect. Dis.* **16**, 1731–1737 (2010).
 33. Barr, I. G. *et al.* An influenza A (H3) reassortant was epidemic in Australia and New Zealand in 2003. *J. Med. Virol.* **76**, 391–397 (2005).
 34. Bragstad, K., Nielsen, L. P. & Fomsgaard, A. The evolution of human influenza A viruses from 1999 to 2006 - a complete genome study. *Virology* **5**, 40 (2008).
 35. Bedford, T. *et al.* Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217–220 (2015).
 36. Barr, I. G. *et al.* Epidemiological, antigenic and genetic characteristics of seasonal influenza A (H1N1), A (H3N2) and B influenza viruses: basis for the WHO recommendation on the composition of influenza vaccines for use in the 2009--2010 Northern Hemisphere season. *Vaccine* **28**, 1156–1167 (2010).
 37. Brown, I. H. The epidemiology and evolution of influenza viruses in pigs. *Vet. Microbiol.* **74**, 29–46 (2000).
 38. Nelson, M. I. *et al.* Stochastic processes are key determinants of short-term evolution in influenza A virus. *PLoS Pathog.* **2**, e125 (2006).
 39. Nelson, M. I. & Holmes, E. C. The evolution of epidemic influenza. *Nat. Rev. Genet.* **8**, 196–205 (2007).
 40. Pitzer, V. E. *et al.* High turnover drives prolonged persistence of influenza in managed pig herds. *J. R. Soc. Interface* **13**, 519–522 (2016).

41. Bedford, T., Cobey, S., Beerli, P. & Pascual, M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathog.* **6**, e1000918 (2010).
42. Hale, B. G., Albrecht, R. A. & Garcia-Sastre, A. Innate immune evasion strategies of influenza viruses. *Futur. Microbiol* **5**, 23–41 (2010).
43. Quiñones-Parra, S. M. *et al.* A Role of Influenza Virus Exposure History in Determining Pandemic Susceptibility and CD8⁺ T Cell Responses. *J. Virol.* **90**, 6936–6947 (2016).
44. Grant, E. J. *et al.* T-cell immunity to influenza A viruses. *Crit. Rev. Immunol.* **34**, 15–39 (2014).
45. Salomon, R. & Webster, R. G. The Influenza Virus Enigma. *Cell* **136**, 402–410 (2009).
46. Spokes, P. J. *et al.* Equine influenza. *Vet. Rec.* **125**, 257–282 (1993).
47. Gibbs, E. P. J. & Anderson, T. C. Equine and canine influenza: a review of current events. *Anim. Heal. Res. Rev.* **11**, 43–51 (2010).
48. Gilkerson, J. R. Equine influenza in Australia: A clinical overview. *Aust. Vet. J.* **89**, 11–13 (2011).
49. Thacker, E. & Janke, B. Swine Influenza Virus: Zoonotic Potential and Vaccination Strategies for the Control of Avian and Swine Influenzas. *J. Infect. Dis.* **197**, S19–S24 (2008).
50. Reynolds, J. J. H., Torremorell, M. & Craft, M. E. Mathematical modeling of influenza a virus dynamics within swine farms and the effects of vaccination. *PLoS One* **9**, (2014).
51. Poljak, Z., Carman, S. & Mcewen, B. Assessment of seasonality of influenza in swine using field submissions to a diagnostic laboratory in Ontario between 2007 and 2012. *Influenza Other Respi. Viruses* **8**, 482–492 (2014).

Table 1: **FluNet incidence data.** The per country major influenza subtype numbers over the entire analyzed period (1999-2016).

Country	Total	H1N1	H3N2	H1pdm	FluB
1.Norway	49244	454	6077	20167	22546
2.Sweden	38841	80	4826	19891	14044
3.Denmark	13254	133	2900	4732	5489
4.Netherlan ds	14161	28	4719	6279	3135
5.Poland	10074	91	181	6838	2964
6.Germany	22479	2142	10393	4856	5088
7.France	72380	390	10543	32406	29041
8.UK	45814	1322	13825	13514	17153
9.Finland	3597	284	1323	1016	974
10.Switzerla nd	12609	234	2307	3746	6322
11.Czech R.	4420	237	1078	1959	1146
12.Romania	12832	593	2114	8273	1852

13.Hungary	4610	20	1070	2664	856
14.Austria	6851	64	2222	1861	2704
15.Iceland	1960	89	863	445	563
16.Ireland	15604	95	4558	8029	2922
17.Italy	28234	412	7429	14656	5737
18.Israel	14651	387	2210	9499	2555
19.Spain	46318	546	13212	20872	11688
20.Russia	105330	2984	24822	59547	17977
21.Japan	144268	19097	51741	45504	27926
22.Korea	37771	5217	13179	9010	10365
23.China	508954	16854	200159	141383	150558
24.Singapore	16383	278	4735	8028	3342
25.Thailand	11333	1311	3658	2856	3508
26.Argentina	30190	723	6307	17769	5391
27.Chile	23257	825	8568	9569	4295

28.Peru	6613	400	1490	2922	1801
29.Australia	28513	475	12684	9057	6297
30.New Zealand	13310	455	4214	3158	5483
31.South Africa	12599	893	4468	4084	3154
32.Canada	155919	339	35420	64443	55717
33.USA	518901	19711	173343	169300	156547
Sum	2031274	77163	636638	728333	589140

Table 2: **NCBI incidence data.** The per country major influenza subtype numbers over the entire analyzed period (1999-2016).

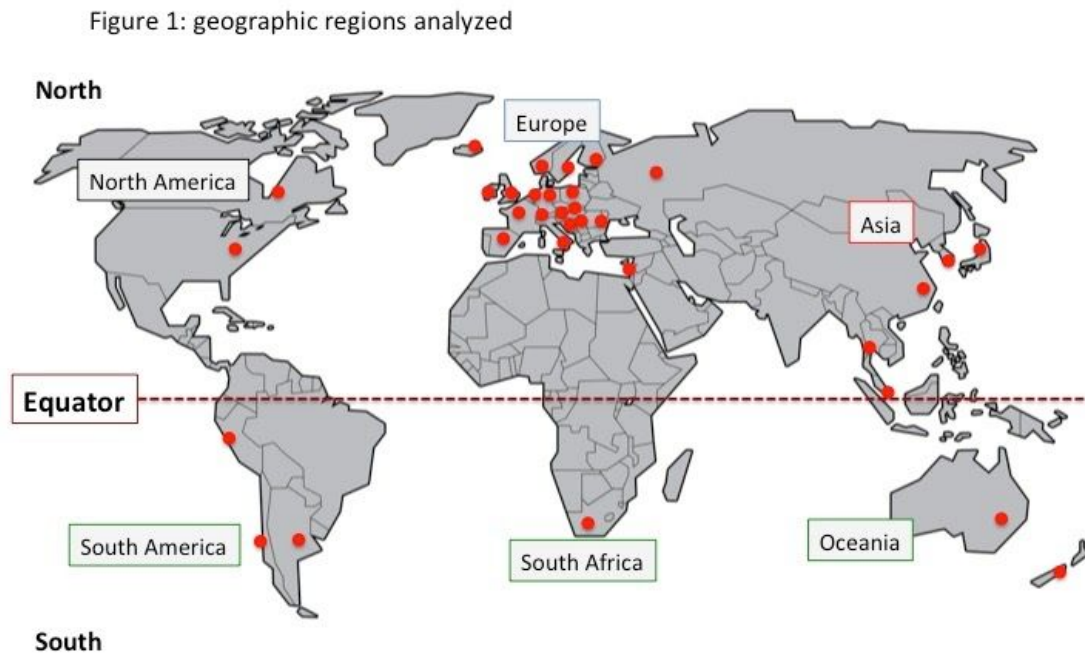
Country	Total	H1N1	H3N2	H1pdm	FluB
1.Norway	267	43	87	108	29
2.Sweden	90	0	65	25	0
3.Denmark	341	0	222	117	2
4.Netherlan ds	137	1	64	72	0
5.Poland	25	0	1	24	0
6.Germany	313	0	120	189	4
7.France	559	4	294	235	26
8.UK	1445	14	260	1115	56
9.Finland	603	0	123	401	79
10.Switzerla nd	23	0	11	8	4
11.Czech R.	0	0	0	0	0
12.Romania	95	4	15	59	17

13.Hungary	22	1	11	10	0
14.Austria	183	0	112	65	6
15.Iceland	11	0	8	3	0
16.Ireland	24	0	4	20	0
17.Italy	1223	13	272	546	392
18.Israel	105	0	23	11	71
19.Spain	706	0	282	398	26
20.Russia	513	1	127	340	45
21.Japan	4081	74	1648	1659	700
22.Korea	896	33	429	318	33
23.China	4846	84	1978	1404	1380
24.Singapore	1930	24	616	989	301
25.Thailand	1547	27	501	642	377
26.Argentina	225	1	93	99	32
27.Chile	254	2	98	129	25

28.Peru	863	6	482	102	273
29.Australia	2166	39	634	287	1206
30.New Zealand	1398	298	569	29	502
31.South Africa	183	7	43	105	28
32.Canada	2337	0	1184	794	359
33.USA	24392	762	10207	7526	5897
Sum	51803	1438	20583	17829	11870

Figures

Figure 1: Countries represented in the dataset

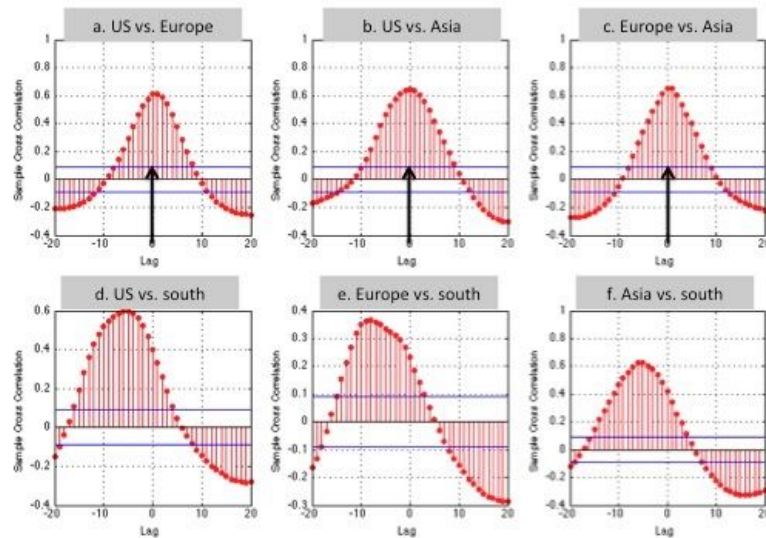


Source: Wikimedia commons (free to use)

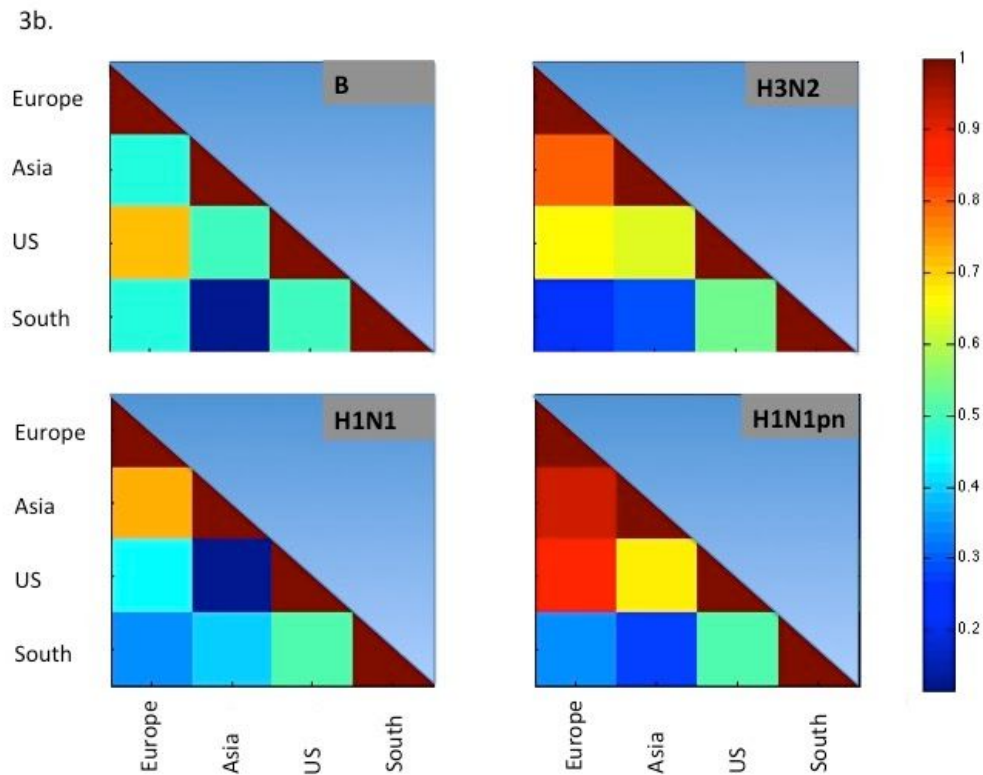
Legend: The map shows the geographic context and region of the countries (red circles) appearing in table 1 and 2. South America, South Africa and Oceania are grouped in the text as 'South'. Image was adapted from a map generated by Wikimedia commons.

Figure 2: Global timing of influenza epidemics

Figure 2: global synchronization of Influenza seasons



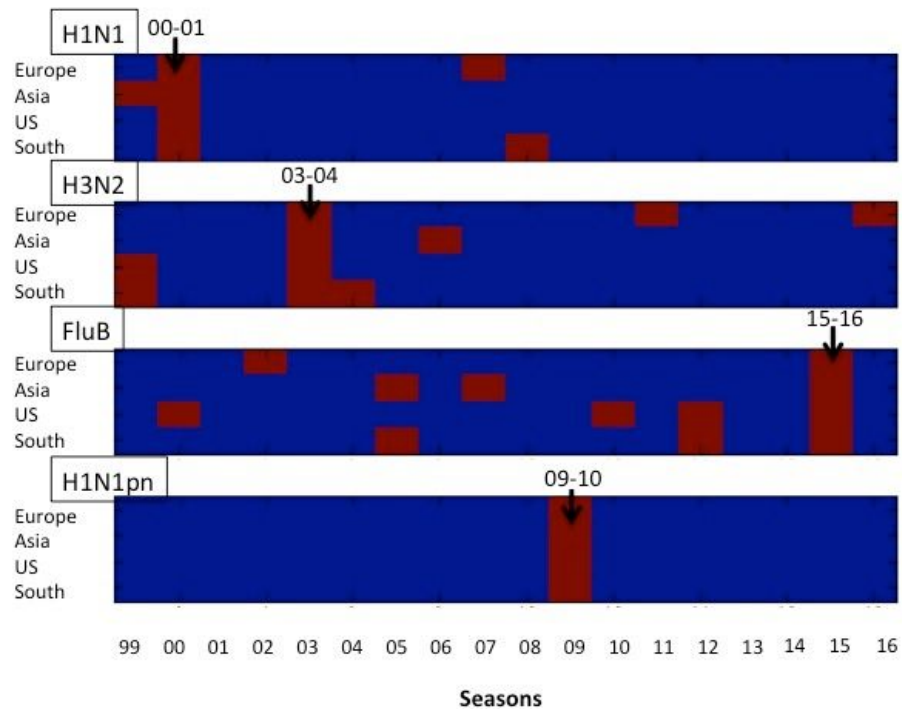
Legend: Influenza (total) cross correlation scores (y-axis) for every pair of defined regions in northern hemisphere (a-c), and versus the 'South' (d-f). Peak at 'zero shift' (a-c) indicates the synchronized epidemics. The shift observed for the south vs. US/Europe/Asia (d-f) is in addition to the 26-week seasonal arbitrary adjustment added to compensate for hemisphere seasons.



Legend: 3a. Influenza subtype relative frequency per region, per season ($f_{A_{it}}$), (Eqn.1, see Methods). The per-region sum over all subtypes is equal to one by definition. Blue squares indicate zero frequency and red indicates a relative frequency of one, i.e. all the annual isolates belong to one subtype (see also color bar). Note the first column represents the 99-00 season and the last column represents the 16-17 season (data ending at Feb 2017). 3b. Pairwise cross correlation scores for per-season frequency in all regions with blue indicating zero and red one (see color bar). Diagonal (red) represents self-correlation.

Figure 4: Highly-represented periods

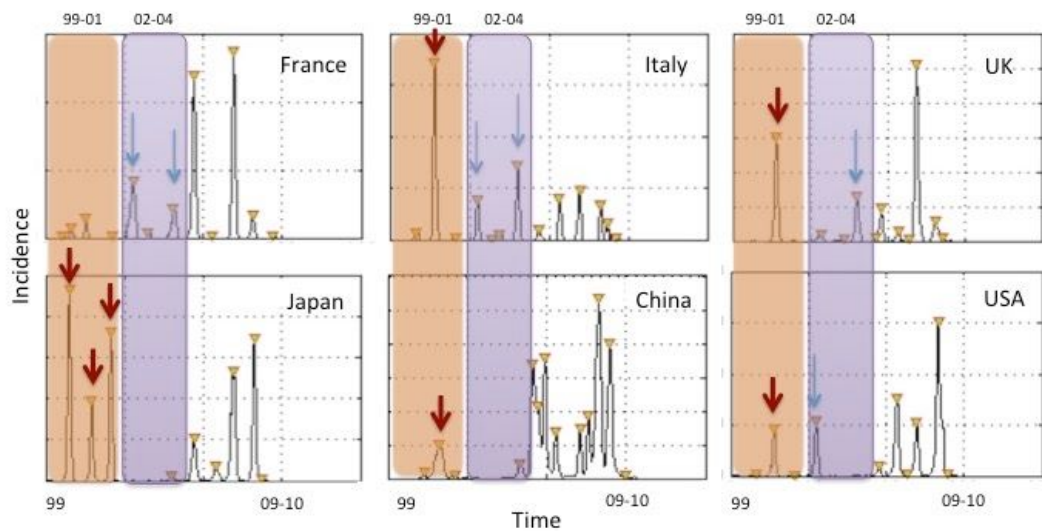
Figure 4: Homogeneous periods – 4 regions (33 countries total), 18 seasons



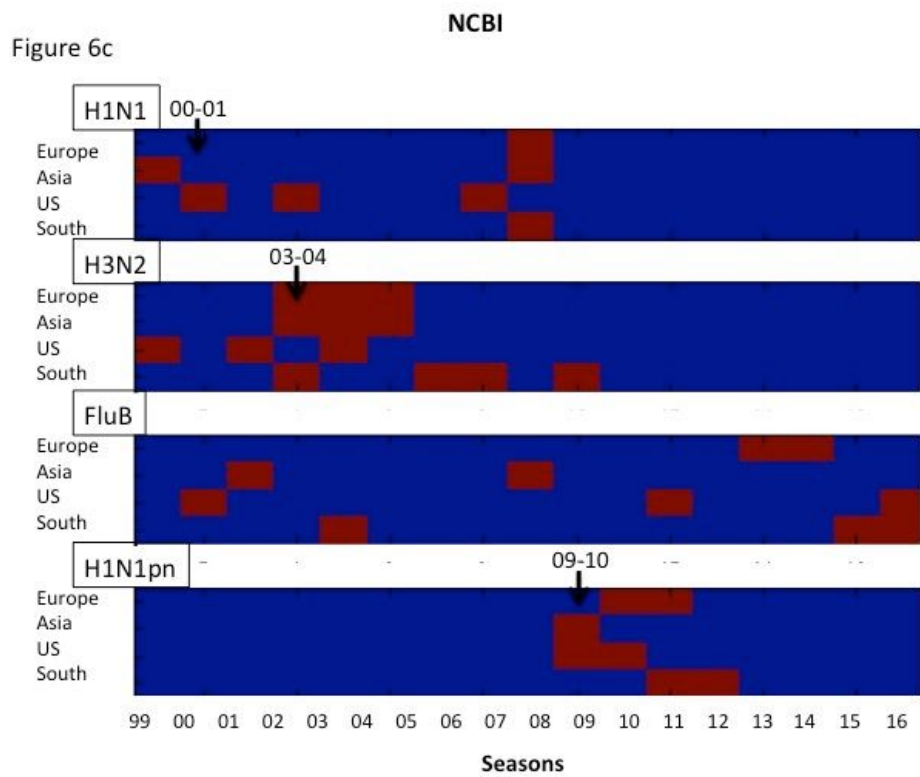
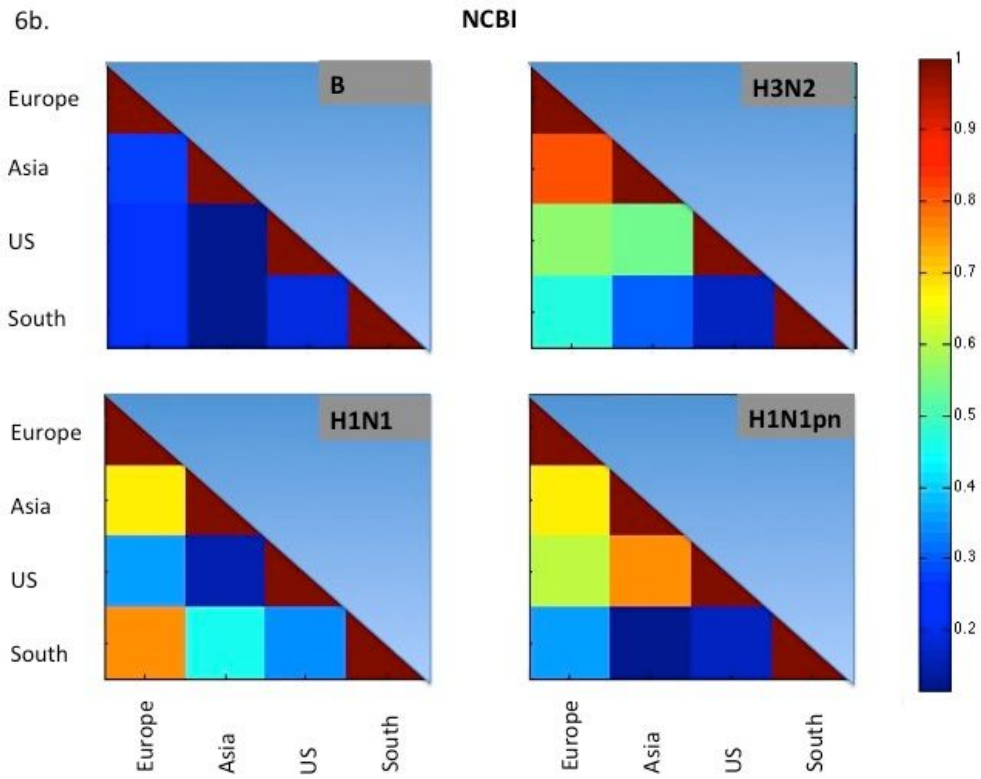
Legend: Highly represented seasons ($HR = 1$), per each of the influenza subtypes, and per region are marked in red, and 'regular' seasons ($HR \sim 1$) are in blue. Globally highly represented seasons ($HR_G = 1$) are indicated by the arrows at 00-01, 03-04, 09-10 and 15-16. Note a region may be highly represented during a season for more than a single subtype (Eqn.2)

Figure 5: H1N1 incidence data

Figure 5: H1N1 incidence data



Legend: Comparison of raw H1N1 per week (x-axis) incidence data (y-axis) from 6 countries. Orange coloring represents the 1999-01 period, purple represents 2002-04 and arrows indicate per country peak periods. Inverted triangles represent periods where incidence is above background noise and these are entirely absent in both China and Japan during 2002-03, but not in European countries or in the US. Europe was the only region showing continuous H1N1 incidence throughout 2002-04.



Legend: 6a. Global proportions of the four subtypes, across regions (rows) and seasons (columns) from 99-00 to 2016-17. Blue squares indicate zero frequency and red indicates a relative frequency of one 6b. Pairwise cross-correlation scores (r) for per-season frequency in all regions with blue indicating $r=0$ and red $r=1$ (see color bar). Diagonal (red) represents self-correlation 6c. Highly represented seasons ($HR = 1$), per each of the influenza subtypes, and per region are marked in red, and 'regular' seasons ($HR \sim 1$) are in blue. Note that none of the seasons showed a globally highly represented (HR_G) pattern, for any of the subtypes.