Genomics-informed Surveillance of RNA Viruses

Matthew Scotch, PhD MPH Associate Professor, Arizona State University Assistant Director, Biodesign Center for Environmental Health Engineering



Public Health Surveillance

- Ongoing, systematic collection of data of individual health events for monitoring population-level trends
 – WHO, WHAT, WHEN, WHERE, WHY
- States have list of reportable diseases

Source: Thacker SB, Qualters JR, Lee LM. *Public Health Surveillance in the United States: Evolution and Challenges*. Morbidity and Mortality Weekly Report (MMWR). July 27, 2012. 61(03);3-9.





Gaps in Public Health Surveillance

- Recent outbreaks: Ebola in Africa; Zika in Americas ullet
- Traditional surveillance is no longer sufficient
- Recent understanding of need for enhancement of: ullet
 - Molecular diagnostics
 - Data sharing
- Pathogen genome sequencing





Phylogeography



<u>Geographay</u> is additional dimension

- RNA Viruses
- Virus DNA sequences



Reducing Virus Spread

- Viruses don't spread in a vacuum!
- What can health agencies do to curb spread?
- Factors can promote/reduce spread:
 - Climate
 - Population density (incl. humans, animals, vectors)
 - Global Travel
 - Genetic (incl. vaccine accuracy)





Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2

Philippe Lemey¹*, Andrew Rambaut^{2,3}, Trevor Bedford², Nuno Faria¹, Filip Bielejec¹, Guy Baele¹, Colin A. Russell^{4,5}, Derek J. Smith^{4,5,6}, Oliver G. Pybus⁷, Dirk Brockmann^{8,9,10}, Marc A. Suchard^{11,12}







Magee D, Suchard MA, Scotch M (2017) Bayesian phylogeography of influenza A/H3N2 for the 2014-15 season in the United States using three frameworks of ancestral state reconstruction. PLOS Computational Biology 13(2): e1005389. https://doi.org/10.1371/journal.pcbi.1005389 https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005389

Human Movement Model (Local)



ZooPhy

- Public Health, Agriculture, Wildlife agencies
- Researchers
- https://zodo.asu.edu/zoophy/
- On-campus influenza surveillance









Please send any questions or concerns to zoophylab@gmail.com



Number of **Records** in the ZooPhy Database

Number of normalized Locations in the ZooPhy Database



Number of normalized Hosts in the ZooPhy Database





Search	Resul

esults Run

71/71 Influenza A records selected 71 complete records selected				 ZooPhy Jobs must contain 5 to 1000 Record 9 distinct locations select 			
Export	Import F	ASTA					Random Sample
ALL 🔽	\$ID	\$ Genes	¢ Date	\$ Host	Country	Length	
	CY055059	HA	17-Jul-2009	human; gender m; age 27y	Australia	1734	
	CY055067	HA	17-Jul-2009	human; gender f; age 2y	Australia	1734	
	CY055075	HA	18-Jul-2009	human; gender f; age 30m	Australia	1734	
	CY055526	HA	16-Jul-2009	human; gender f; age 59y	Australia	1721	
	CY055534	HA	18-Jul-2009	human; gender m; age 8m	Australia	1734	
	CY055542	HA	20-Jul-2009	human; gender m; age 36y	Australia	1734	
	CY055550	HA	20-Jul-2009	human; gender m; age 8m	Australia	1734	
	CY055558	HA	20-Jul-2009	human; gender m; age 42y	Australia	1734	
	CY055565	HA	20-Jul-2009	human; gender m; age 36y	Australia	1734	
•	CY055573	HA	20-Jul-2009	human; gender f; age 29y	Australia	1734	
	CY055581	HA	20-Jul-2009	human; gender m; age 6y	Australia	1734	
<	CY055589	HA	20-Jul-2009	human; gender f; age 2y	Australia	1734	

Record Details: CY055067

Date: 17-Jul-2009	PubMed ID: n/a
Taxon: 708521	Strain: A/Australia/4/2009
Isolate: Unknown	Host: human; gender F; age 2Y
Location: merrylands,new south wales,AU	Genes: HA
Virus: Influenza A virus (A/Australia/4/20 viruses; ssRNA negative-strand viruses; Influenzavirus A.	009(H1N1)) Viruses; ssRNA Orthomyxoviridae;
Definition: Influenza A virus (A/Australia complete sequence.	/4/2009(H1N1)) segment 4,

View Genbank Record



About Zoophy



1	Id.Date.State	
2	HQ712184,27Aug2009,Queensland	jtataggttatcatgcgaa
3	HQ712183,28Aug2009,Queensland	accttctagaagacaagc
4	HM754655,24Jul2009,Western Australia	tgctggctggatcctggga
5	HM624087,17Aug2009,Western Australia	ittcagacaatggaacgtg
6	HM624086,13Aug2009,Western Australia	ittgaaaggtttgagatat
7	HM624085,07Aug2009,Western Australia	tcatgctggagcaaaaagc
8	GQ258462,01May2009,South Australia	ctacattaatgataaagg
9	GQ243761,21May2009,Victoria	tctatcagaatgcagatg
10	GQ243757,29May2009,Northern Territory	acccaaagwgagggatcaa
11	GQ243755,26May2009,Western Australia	caactogaaatctagtggt
12	GQ243751,19May2009,Victoria	cagtccacgattgcaata
13	GQ160611,07May2009,Queensland	atcacaattagaaaatat
14	CY092438,24Nov2009,New South Wales	tcaatctagaggggggggg
15	CY092193,24Jul2009,New South Wales	tatcaccatcaaaatnanc
16	CY092185,21Jul2009,New South Wales	raaantaaattetattatt
17	CY055940,02Sep2009,New South Wales	
18	CY055932,02Sep2009,New South Wales	sata a a a a a a a c t t a a a c
19	CY055924,26Aug2009,New South Wales	
20	CY055916,20Aug2009,New South Wales	
21	CY055908,12Aug2009,New South Wales	
22	CY055900,10Aug2009,New South Wales	
23	CY055892,06Aug2009,New South Wales	JUGCICIAALGGGUCICIA
24	CY055884,05Aug2009,New South Wales	
25	CY055876,04Aug2009,New South Wales	
26	CY055868,03Aug2009,New South Wales	********
27	CY055860,03Aug2009,New South Wales	
28	CY055852,03Aug2009,New South Wales	Jaccilclagaagacaage
29	CY055844,03Aug2009,New South Wales	
30	CY055836,02Aug2009,New South Wales	Jttcagacaatggaacgtg
31	CY055828,02Aug2009,New South Wales	
32	CY055020,01AU92009,New South Wales	
33	CY055904 201ul2009,New South Wales	ctacattaatgataaagg
34 25	CY055706 16Aug2000 New South Wales	ctctatcagaatgcagatg
20	CV055788 201012000 New South Wales	icccaaagtgagggatcaa
30 27	CV055780 281u12009,New South Wales	заастддааатстадтддт
20	$(Y055772 \ 271u12009)$ New South Wales	cagtccacgattgcaata
30	CY055764 271u12009 New South Wales	jatcacaattggaaaatgt
<u> </u>	CY055756 271u12009 New South Wales	ttcaatctagaggcctatt
40 //1	CY055748 261u12009 New South Wales	tatcaccatcaaaatgagc
42	CY055740,261u12009,New South Wales	caaagtaaattctgttatt
43	CY055732.25Jul2009.New South Wales	tagagaatttaaataaaaa
44	CY055724,24Jul2009,New South Wales	<pre>atgaaagaactttggact</pre>
45	CY055716.24Jul2009.New South Wales	caaggaaattggaaacggc
46	CY055708.24Jul2009.New South Wales	<pre>stgactacccaaaatactc</pre>
47	CY055700,24Jul2009.New South Wales	taccagattttggcgatct
48	CY055692.24Jul2009.New South Wales	jtgctctaatgggtctcta
49	CY055684,23Jul2009,New South Wales	

~/Downloads/ZooPhyData(1).csv 🗸

Search	Results	Run							
		ø	Successfully Star	ted the Zool	Phy Job: AU-pdm09. The follow	wing 1 incompl	ete records were e	xcluded from this job: GQ2	258462
Run ZooPhy Job Settings									
					Selected	Records: 71			
Email: Job Name (optional):				al):					
				mscotch@	Pasu.edu	AU-pdm	AU-pdm09		
						Ontions			
				lleo Dofault		Options	load Custom Brod	listors O	
					Default Predictors	Generate Template		ate	
					Advanced	Parame	Parameters		
		S	Substution Model	0	Site Heterogeneity 😧	Clock Mo	del 😧	Tree Prior 😧	
			HKY	•	Gamma 🔽 🤨 Invariant 🗌 😧	Strict	•	Skygrid	•
		c	Chain Length 😧		Sample Rate 0				
			5000000	٢	5000 ©				
							≢ Advanced O	Options Start ZooPhy	





Current date: 2009/11/23









Limitations of Discrete Phylogeography

- Sampling location fixed to a single or small number of locations
- No sampling error



Rabies virus strain 4102Ct08 glycoprotein (G) gene, partial cds; and G-L intergenic spacer, partial sequence

GenBank: GU123662.1

EASTA Creek	
FASTA Grapt	
Features Sequ	
LOCUS DEFINITION	$\frac{GU123662}{Rabies virus strain}$
ACCESSION VERSION KEYWORDS	GU123662 GU123662.1 GI:2624804 / VVUIILLY DLQQLL
SOURCE ORGANISM	Rabies virus <u>Rabies virus</u> Viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Lyssavirus.
REFERENCE AUTHORS	1 (bases 1 to 1305) Carnieli,P. Jr., Fahl Wde,O., Brandao,P.E., Oliveira Rde,N., Macedo,C.I., Durymanova,E. and Castilho,J.G.
TITLE	Comparative analysis of rabies virus isolates from Brazilian canids and bats based on the G gene and G-L intergenic region
JOURNAL PUBMED	Arch. Virol. 155 (6), 941-948 (2010) 20401500
REFERENCE AUTHORS	2 (bases 1 to 1305) Carnieli,P. Jr., Fahl,W.O., Castilho,J.G., Oliveira,R.N., Macedo,C.I., Brandao,P.E., Carrieri,M.L. and Kotait,I.
TITLE	Genetic typification of rabies virus isolated from their main hosts in Brazil
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1305)
AUTHORS	Carnieli,P. Jr., Fahl,W.O., Castilho,J.G., Oliveira,R.N., Macedo,C.I., Brandao,P.E., Carrieri,M.L. and Kotait,I.
TITLE	Direct Submission
JOURNAL	Submitted (19-OCT-2009) Virglogy, Pasteur Institute, Av. Paulista,
	393, Sao Paulo, Sao Paulo Ø1311-000, Brazil
FEATURES	Location/Qualifiers
source	
	/organism="Rables Virus" (mal turne-"union a DNN"
	/mol_type="v2fal CRNA"
	/bost="Cordony thous"
	/db_vref=/tavon:11292"
	/country="Brazil"
	/collection_date="2008"
gene	<1840 ////////////////////////////////////
CDS	<1846
000	/gene="G"
	/codon start=1
	/product="glycoprotein"
	/protein_id="ACY68993.1"
	/db_xref="GI:262480440"
	/translation="LKLCGVLGLRLVDGTWIAMQTSDEIKWCPPDQLVNLHDFRSDEI
	FULVUERT.VKKRERCT.DAT.FST.VTTKSVSFRRT.SHT.RKT.VDGFGKAVTTENKTI.MFAD

Public Health Importance

- Sparse or incomplete geospatial data
 - Difficult to analyze local outbreaks
 - Inaccurate conclusions about spread



Goals of this work

- Relax strong assumption of no sampling error
- Analytically integrate over the sampling error to evaluate the likelihood of the spatial process
- Evaluate the use of sampling error for virus phylogeography

Source: Scotch M, Tahsin T, Weissenbacher D, O'Connor K, Magge A, Vaiente M, Suchard MA, Gonzalez-Hernandez G. *Incorporating sampling uncertainty in the geospatial assignment of taxa for virus phylogeography.* Virus Evol. 2019 Feb 28;5(1):vey043.







Volume 5, Issue 1 January 2019

Article Contents

Abstract

Incorporating sampling uncertainty in the geospatial assignment of taxa for virus phylogeography 👌

Matthew Scotch ⊠, Tasnia Tahsin, Davy Weissenbacher, Karen O'Connor, Arjun Magge, Matteo Vaiente, Marc A Suchard, Graciela Gonzalez-Hernandez

Virus Evolution, Volume 5, Issue 1, January 2019, vey043, https://doi.org/10.1093 /ve/vey043

Published: 28 February 2019

▶ PDF ■ Split View 66 Cite

🔎 Permissions 💿 < Share 🔻



Email alerts

New issue alert In progress issue alert Advance article alerts

Article activity alert

Bioinformatics, 34(9), 2018, 1606–1608 doi: 10.1093/bioinformatics/btx799 Advance Access Publication Date: 12 December 2017 Applications Note

OXFORD

Data and text mining

GeoBoost: accelerating research involving the geospatial metadata of virus GenBank records

Tasnia Tahsin^{1,*}, Davy Weissenbacher¹, Karen O'Connor², Arjun Magge¹, Matthew Scotch^{1,3} and Graciela Gonzalez-Hernandez²

¹Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ 85259, USA, ²Institute of Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA and ³Biodesign Center for Environmental Health Engineering, Arizona State University, Tempe, AZ 85281, USA

*To whom correspondence should be addressed. Associate Editor: Jonathan Wren

Received on August 4, 2017; revised on December 4, 2017; editorial decision on December 6, 2017; accepted on December 11, 2017



GeoBoost (ZoDo) Pipeline



Getting GeoBoost

- <u>https://tinyurl.com/geoboost</u>
- Next command-line version will integrate NN
- GUI will be available as part of phylogeography portal



Use for phylogeography

- Set of candidate locations for each taxa
 P (L|R) = 1.0
- Analytically integrate over observation process
- BEAST 1.10

Source: Suchard, M.A., et al., Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol, 2018. 4(1)



Integration into **BEAST**











Select Input Type	GenBank Ids PubMed/PubMedCentral Ids Raw Text	
5297743, 268971	108	
Extract	2 records retrieved.	
PubMed	Locations in PubMed	COLOMBIA
15297743	Genetic and phylogenetic analysis of glycoprotein of rabies virus isolated from several species in Brazil.	+ AMAZON BASIN
26897108	Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. BACKGROUND: The incidence of microcephaly in Brazil in 2015 was 20 times higher than in previous years. Epidemiological data suggest that microcephaly cases in Brazil might be associated with the	BRAZIL BOLIVIA
	introduction of Zika virus. We aimed to detect and sequence the Zika virus genome in amniotic fluid samples of two pregnant women in <mark>Brazil</mark> whose fetuses were diagnosed with microcephaly.	Sest Known Location Possible Locations
	METHODS: In this case study, amniotic fluid samples from two pregnant women from the state of Paraíba in Brazil whose fetuses had been diagnosed with microcephaly were obtained, on the recommendation of the Brazilian health authorities, by ultrasound-guided transabdominal amniocenteries at 29 weeks' activities.	
	Phylogenetic reconstruction and investigation of recombination events were done by comparing the Brazilian Zika virus genome with sequences from other Zika strains and from flaviviruses that occur in similar regions in Brazil.	
	After sequencing of the complete genome of the Brazilian Zika virus isolated from patient 1, phylogenetic analyses showed that the virus shares 97-100% of its genomic identity with lineages isolated during an outbreak in French Polynesia in 2013, and that in both envelope and NS5	
	genomic regions, it clustered with sequences from North and South America, southeast Asia, and the Pacific. INTERPRETATION: These findings strengthen the putative association between Zika virus and	
	cases of microcephaly in neonates in Brazil. FUNDING: Consellho Nacional de Desenvolvimento e Pesquisa (CNPq), Fundação de Amparo a	

Web Portal

- Will soon add GeoBoost to ZooPhy web portal
- Enable user to run phylogeography with uncertainty
- Additional features:
 - Log Analyzer/Combiner, Skygrid/plots, infectivity (R₀).



Conclusion

- Informatics has great potential to improve public health surveillance
- This include approaches to leverage sequence-informed surveillance
- Phylogeography is a science that can use sequences to help surveillance



Team

- Graciela Gonzalez, PhD (Penn, Multiple PI)
- Marc Suchard, MD, PhD (UCLA, Co-investigator)
- Arjun Magge (ASU, PhD Student)
- Matteo Vaiente (ASU, PhD Student)
- Karansher Bhangal (ASU, MS Student)
- Prashanth Gonarkar (ASU, MS Student)



Thank You

- Funding:
 - NIH R01LM012080 and R01AI117011
 - PLuS Alliance PA18A02 Accelerator Award

