

ZooPhy: A bioinformatics pipeline for virus phylogeography and surveillance

Matthew Scotch, Arjun Magge, Matteo Vaiente

Arizona State University, Tempe, Arizona, United States

Objective

We will describe the ZooPhy system for virus phylogeography and public health surveillance [1]. ZooPhy is designed for public health personnel that do not have expertise in bioinformatics or phylogeography. We will show its functionality by performing case studies of different viruses of public health concern including influenza and rabies virus. We will also provide its URL for user feedback by ISDS delegates.

Introduction

Sequence-informed surveillance is now recognized as an important extension to the monitoring of rapidly evolving pathogens [2]. This includes phylogeography, a field that studies the geographical lineages of species including viruses [3] by using sequence data (and relevant metadata such as sampling location). This work relies on bioinformatics knowledge. For example, the user first needs to find a relevant sequence database, navigate through it, and use proper search parameters to obtain the desired data. They also must ensure that there is sufficient metadata such as collection date and sampling location. They then need to align the sequences and integrate everything into specific software for phylogeography. For example, BEAST [4] is a popular tool for discrete phylogeography. For proper use, the software requires knowledge of phylogenetics and utilization of BEAUti, its XML processing software. The user then needs to use other software, like TreeAnnotator [4], to produce a single ("representative") maximum clade credibility (MCC) tree. Even then, the evolutionary spread of the virus can be difficult to interpret via a simple tree viewer. There is software (such as SpreaD3 [5]) for visualizing a tree within a geographic context, yet for novice users, it might not be easy to use. Currently, there are only a few systems designed to automate these types of tasks for virus surveillance and phylogeography.

Methods

We have developed ZooPhy, a pipeline for sequence-informed surveillance and phylogeography [1]. It is designed for health agency personnel that do not have expertise in bioinformatics or phylogeography. We created a large database of all virus sequences and metadata from GenBank [6] as well as a smaller database for selected viruses perceived to be of great interest for health agencies including: influenza (A, B, and C), Ebola, rabies, West Nile virus, and Zika virus.

In Figure 1A, we show our front-end architecture, created in the style of the influenza research database [7], that enables the user to search by: virus, gene name, host, time-frame, and geography. We also allow users to upload their own list of GenBank accessions or unpublished sequences. Hitting "Search" produces a Results tab which includes the metadata of the sequences. We provide a feature to randomly down-sample by a specified percentage or number. We also allow the user to download the metadata in CSV format or the unaligned sequences in FASTA format.

The final tab, "Run", includes a text box for specifying an email in order to send job updates and final results on virus spread. We also enable for the user to study the influence of predictors on virus spread (via a generalized linear model). Currently, we have predictors such as temperature, great circle distance, population, and sample size for selected countries. We also offer experts the ability to specify advanced modeling parameters including the molecular clock type (strict vs. relaxed), coalescent tree prior, and chain length and sampling frequency for the Markov-chain Monte Carlo. When the user selects "Start ZooPhy", a pre-processor eliminates incomplete or non-disjoint record locations and sends the rest for analysis.

Results

When initiated, the ZooPhy pipeline includes sequence alignment via Mafft [8] and creation of an XML template via BEASTGen for input into BEAST for discrete phylogeography. It then uses TreeAnnotator [3] to create an MCC tree from the posterior distribution of sampled trees. ZooPhy uses the MCC as input into SpreaD3 for a recreation of the time-estimated migration via a map. If the user selects the GLM option, the system runs an R script to calculate the Bayes factor of the inclusion probability for



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online Journal of Public Health Informatics * ISSN 1947-2579 * http://ojphi.org * 11(1): e301, 2019

ISDS 2019 Conference Abstracts



each predictor and draws a plot including the regression coefficient and its 95% Bayesian credible interval. We are currently working on new visualization techniques such as those demonstrated by Dudas et al. that combine time-oriented spread via a map and evolution on a phylogenetic tree annotated by discrete locations [9].

Conclusions

Recent advances in phylodynamics, bioinformatics, and visualization have demonstrated the potential of pipelines to support surveillance. One example is NextStrain which can perform real-time virus phylodynamics [10]. The system has recently been added as an app to the Global Initiative on Sharing Avian Influenza Data (GISAID) database for influenza tracking using DNA sequences [11]. This presentation will highlight a pipeline for virus phylogeography designed for epidemiologists who are not experts in bioinformatics but wish to leverage virus sequence data as part of routine surveillance. We will describe the development and implementation of our system, ZooPhy, and use real-world case studies to demonstrate its functionality. We invite ISDS delegates to use the system via our web portal, https://zodo.asu.edu/zoophy/ and provide feedback on system utilization.

Acknowledgement

This work was supported by the National Library of Medicine of the NIH under award R01LM012080 (to MS).

References

- Scotch M, Mei C, Brandt C, Sarkar IN, Cheung K. 2010. At the intersection of public-health informatics and bioinformatics: using advanced Web technologies for phylogeography. *Epidemiology*. 21(6), 764-68. <u>PubMed</u> <u>https://doi.org/10.1097/EDE.0b013e3181f534dd</u>
- Gardy JL, Loman NJ. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet. 19, 9-20. <u>PubMed https://doi.org/10.1038/nrg.2017.88</u>
- 3. Avise JC. Phylogeography: the history and formation of species. 2000, Cambridge, Mass.: Harvard University Press.
- 4. Suchard MA, Lemey P, Baele G, et al. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4(1), vey016. <u>PubMed https://doi.org/10.1093/ve/vey016</u>
- 5. Bielejec F, Baele G, Vrancken B, et al. 2016. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol Biol Evol*. 33(8), 2167-69. <u>PubMed https://doi.org/10.1093/molbev/msw082</u>
- 6. Benson DA, Cavanaugh M, Clark K, et al. 2018. GenBank. *Nucleic Acids Res.* 46, D41-47. <u>PubMed https://doi.org/10.1093/nar/gkx1094</u>
- 7. Zhang Y, et al. 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.* 45, D466-74. <u>PubMed https://doi.org/10.1093/nar/gkw857</u>
- 8. Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol*. 1079, 131-46. <u>PubMed https://doi.org/10.1007/978-1-62703-646-7_8</u>
- 9. Dudas G, et al. 2017. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*. 544(7650), 309-15. <u>PubMed https://doi.org/10.1038/nature22040</u>
- 10. Hadfield J, Megill C, Bell SM, Huddleston J, et al. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 34(23), 4121-23. <u>PubMed https://doi.org/10.1093/bioinformatics/bty407</u>
- 11. NextFlu. 2018; Available from: https://www.gisaid.org/epiflu-applications/nextflu-app/.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



ISDS 2019 Conference Abstracts

arch Reads Rom Arcs Root Roman Roma	6	Rec	onstruc	cting Virus Sp	read using Phyl	ogeography	- 1	1		A	out Zoophy	
North North Continent Pom To Number 2A + Human + North America - 2017 2017 Number 2A - <td< th=""><th>earch</th><th>Res</th><th>ults</th><th>Run</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></td<>	earch	Res	ults	Run								
Influenzal A Image: Barrier and Statistics of America H North America Minimum Sequence Lea Andraga and Darbuda Andraga and Darbuda <	Virus				Host			Continent		From	То	
Number of Subsection Participation Number of Subsection Number of Subsection Number of Subsection Participation	Infic	Jenza A		\$	Human		¢	North America	a \$	2017	2017	
Image: Space Anguila TOD Image: Space Anguila TOD Image: Space Regions Regions Image: Space Regions Regions </td <td>Flu A</td> <td>Sub Tyr</td> <td>10</td> <td></td> <td></td> <td></td> <td></td> <td>Countries</td> <td></td> <td>Minimum Sec</td> <td>wence Lengt</td>	Flu A	Sub Tyr	10					Countries		Minimum Sec	wence Lengt	
1 1		н		N				Anguilla		170	0	
Bahamas PBG PBG PBG Automas PBG Automas PBG Automas PBG Control PBG Control Control PBG Control Control PBG Control Control Control PBG Control Control Control Control PBG Control Control Control Contro <td>1</td> <td>\$</td> <td></td> <td>1 ¢</td> <td></td> <td></td> <td></td> <td>Antigua and Aruba</td> <td>Barbuda</td> <td></td> <td></td>	1	\$		1 ¢				Antigua and Aruba	Barbuda			
Construction All Animal Analysis PB-1 All Animal Analysis PA Analysis PA All Animal Analysis PA Analysis <td></td> <td>pdm0</td> <td>9 Only</td> <td>? 📴</td> <td></td> <td></td> <td></td> <td>Bahamas Regions</td> <td></td> <td></td> <td></td>		pdm0	9 Only	? 📴				Bahamas Regions				
PB3 PA Alabama A	Gene	95						All United St	ates of America			
PA Ations St Advanced Options St Records St Records COUNT Pace send any questions or concerns to coophylob@gmil.com Descendance About 2000 Conffy Descendance B Descendance Conffy Descendance B Descendance Descendance Conffy Descendance B Descendance Descendance Conffy Descendance Descendance Descendance Descendance Conffy Desc	PB PB	12						Alabama Alaska				
Image: States Control Descendary questions or concerts to scoophydol@granicor Control Descend	HA							Arizona				
Proventions Statements Control Processional and questions or concerns to 2000/hytob/gmail.com Advanced or pytobs Advanced or pytobs Processional and questions or concerns to 2000/hytob/gmail.com Advanced or pytobs Advanced or pytobs Processional and questions or concerns to 2000/hytob/gmail.com Advanced or pytobs Advanced or pytobs Processional and questions or concerns to 2000/hytob/gmail.com Advanced or pytobs Advanced or pytobs Processional and questions and the pytobs of pytobs Procession of pytobs Advanced or pytobs Processional and questions and the pytobs of pytobs Procession of pytobs Procession of pytobs Processional and questions and the pytobs of pytobs Procession of pytobs Procession of pytobs Processional and questions and pytobs Procession of pytobs Procession of pytobs Processional and question of pytobs Procession of pytobs Procession of pytobs Processional and question of pytobs Procession of pytobs Procession of pytobs Processional and question of pytobs Procession of pytobs Procession of pytobs Procession of pytobs Processional and question of pytobs Procession of pytobs Procession of pytobs Procession of pytobs Proc	+	Advances	Onto									
Results COMM Place send any questions or concerns to zoophylobiligmal.com Advanced any questions on questions to for the span="2">Advanced any question content to span="2">Advanced any question span="2">Advanced any question content to span="2" <th any<="" codd="" td=""><td>21</td><td>чочалсек</td><td>J Opuo</td><td>ns</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th>	<td>21</td> <td>чочалсек</td> <td>J Opuo</td> <td>ns</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	21	чочалсек	J Opuo	ns							
Conference of processing Phylogeography B Advacation Note Network Reference 0.2005Phylogeography B Advacation Note Network Reference 0.2005Phylogeography B Advacation Note Network Reference 0.2005Phylogeography Brandon Balance D Note Network Reference 0.2005Phylogeography Brandon Balance D D Note Network Reference 0.2005Phylogeography Brandon Balance D	*	Rosot				36	61 Recor	rds			Q Search	
And the series of the serie					Please se	nd any questions	or concer	ns to zoophylab@g	mail.com			
Reconstructing Winis Bynead using Phylogrography Model 2006 Results Run Results Run Instances A monode selected Instances A monode select		ZooPhy									(and a start file)	
N Result Run 11 Influenza A moords selected meller moords selected moords selected moordse selected moords selected moords selected moords selected moords	2)	Reconstru	acting Vi	irus Spread using	g Phylogeography		Ľ	5			About 200phy	
11 Influenza A moords selected 0 Zooffry Jobs must contain 5 to 1000 Records	ch	Results	Run									
Umport PASTA Umport PASTA Umport PASTA 11 10 ° 0 anno 2017 Porto segimer; gender m; ige 50 United States 1752 12 12 0 ° 0 anno 2017 Porto segimer; gender m; ige 50 United States 1752 12 12 14 0 ° - Jan - 2017 Porto segimer; gender m; igg 60 United States 1752 12 12 14 0 ° - Jan - 2017 Porto segimer; gender m; igg 60 United States 1752 12 12 Porto segimer; gender m; igg 60 United States 1752 Issue : United Viewer; State Viewer; Viewer; State Viewer; Viewer; State Viewer; Viewer; State Viewer; Viewer	361 Infli complet	uenza A reci te records se	ords selec slected	ted		O ZooPhy Job	s must conta 0 1 distin	in 5 to 1000 Records of locations selected	Re	cord Details: CY21433	8	
C1721432 HA 60-Jan-2017 homo sagients: gender m; age 63 1752 C17214330 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214331 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214336 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214336 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214528 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214528 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214528 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214528 HA 60-Jan-2017 homo sagients: gender f; age 63 1752 C17214528 HA 60-Jan-2017 homo sagients; gender 1752 Tomo sagients; gender 1752 C17214529 HA 60-Jan-2017 homo sagients; gender 1752 Tomo sagients; gender 1752 C17214581 HA 60-Jan-2017 homo sagients; gender 1752 Tomo sagients; gender 1752 C17214581 HA 60-Jan-2017 homo sagients; gender 1	Export	Errogeni erro	ASTA	t Date	t Host	* Country	# Leogth	Random Sample	Date: 05-Jan-2017	PubMed ID: n/a		
GT214330 HA Q2-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT214336 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT214336 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT214328 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT214238 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT214258 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT2142573 HA Q5-Jan-2017 forms septems, genote fr. age 63 United Bates 1752 GT2142573 HA Q5-Jan-2017 form septems, genote fr. age 63 United Bates 1752 GT2142573 HA Q5-Jan-2017 form septems, genote fr. age 63 United Bates 1752 GT2142573 HA Q6-Jan-2017 form septems, genote fr. age 36 United Bates 1752 GT2142573 HA Q6-Jan-2017 form septems, genote fr. age 36 United States 1752	۵	CY214322	на	02-Jan-2017	homo sapiens; geoder mi ana 67	United States	1752		Taxon: 1940351 Isolate: Unknown	Strain: A/Texas/03/2017 Host: Homo sapiens; ge	nder F; age 9	
Ch214358 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214358 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214358 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214526 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214526 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214527 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214573 HA 05-Jan-2017 Nome septence Variated States 1752 Ch214505 HA 06-Jan-2017 Nome septence Variated States 1752 Ch214505 HA 06-Jan-2017 Nome septence Variated States 1752 Ch214505 HA 06-Jan-2017 Nome septence Variated States 1752 Oromo septence		GY214330	на	02-Jan-2017	homo sapiens;	United States	1752		Location: texas,US Virus: Influenza A vir	Genes: HA us (A/Texas/03/2017(H1N1)	Viruses; ssRNA	
CY215218 HA Op-Jan-2017 Promo sapients: gender m; tage 63 United States 1752 CY215218 HA Ob-Jan-2017 Tomo sapients: gender m; tage 63 United States 1752 CY215218 HA Ob-Jan-2017 Tomo sapients: gender m; tage 63 United States 1752 CY215218 HA Ob-Jan-2017 Tomo sapients: gender m; tage 63 United States 1752 CY215218 HA Ob-Jan-2017 Tomo sapients: gender m; tage 73 United States 1752 CY215218 HA Ob-Jan-2017 Tomo sapients: gender m; tage 73 United States 1752 CY215285 HA Ob-Jan-2017 Tomo sapients: gender m; tage 73 United States 1752 CY215285 HA Ob-Jan-2017 Tomo sapients: gender m; tage 73 United States 1752 CY215285 HA Ob-Jan-2017 Tomo sapients: gender m; tage 58 United States 1752 CY215285 HA Ob-Jan-2017 Tomo sapients: gender m; tage 58 United States 1752 CY215285 HA Ob-Jan		CY214338	на	05-Jan-2017	homo sapiens;	United States	1752		viruses; ssRNA nega Influenzavirus A.	tive-strand viruses; Orthorn	yxoviridae;	
GV21526 HA G2-Jan-2017 born sapients: Qu'hide Bates 1752 GV21526 HA G4-Jan-2017 horn sapients: Qu'hide Bates 1752 GV21527 HA G3-Jan-2017 horn sapients: Qu'hide Bates 1752 GV21528 HA G4-Jan-2017 horn sapients: QU'hide Bates 1752 HA G4-Jan-2017 horn sapients: QU'hide Bates 1752 HA G4-Jan-2017 horn sapi		CY216218	на	02-Jan-2017	homo sapiens;	United States	1752		Definition: Influenza hemagglutinin (HA) g	A virus (A/Texas/03/2017(H ene, complete cds.	1N1))	
CV21526 HA 04-Jan-2017 homo sapients: gap 6 United States 1752 CV21526 HA 04-Jan-2017 homo sapients: gap United States 1752 CV215689 HA 04-Jan-2017 homo sapients: gap United States 1752 CV215689 HA 04-Jan-2017 homo sapients: United States 1752 CV215689 HA 04-Jan-2017 homo sapients: United States 1752 gender tr: gap 8 United States 1752 CV215689 HA 04-Jan-2017 homo sapients: United States 1752 CV215689 HA 04-Jan-2017 homo sapients: United States 1752 gender tr: gap 8 United States 1752 gender tr: gap 8 United States 1752		CV215226	HA	02-Jan-2017	homo sapiens;	United States	1752			View Genbank Record		
CV216873 HA 03-Jan-2017 homo septients: age: United Bates 1752 CV216881 HA 06-Jan-2017 homo septients: age: United Bates 1752 CV216881 HA 06-Jan-2017 homo septients: age: United Bates 1752 CV216881 HA 06-Jan-2017 homo septients: age: 30 1752 CV216803 HA 06-Jan-2017 homo septients: age: 38 1752 CV216005 HA 06-Jan-2017 homo septients: age: 38 1752 CV216005 HA 06-Jan-2017 homo septients: age: 38 1752 OC7216005 HA 06-Jan-2017 homo septients: age: 38 1752 CV216005 HA 06-Jan-2017 homo septients: age: 38 1752 OC7216005 HA 06-Jan-2017 homo septient: age: 38 1752 OC7216005 HA 06-Jan-2017 homo septient: age: 38 1752 OC OC OC Show Heatmap Q Viewing Q Selected		CY215258	на	04-Jan-2017	homo sapiens;	United States	1752			Record Location		
CY216811 HA 06-Jan-2017 formo sapiens; age United States 1752 GY216801 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 39 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 gender m: age 59 CY216805 HA 04-Jan-2017 formo sapiens; United States 1752 CY216805 HA 04-Jan-2		CY216873	на	03-Jan-2017	homo sapiens; age	United States	1762		1 to			
CV216669 HA 04-Jan-2017 homo sapines: United Bates 1752 GV216005 HA 04-Jan-2017 homo sapines: United States 1752 GV216005 HA 04-Jan-2017 homo sapines: United States 1752 gender m; age 59 United States 1752 GC UNITED States 0 United States 1752 GC UNITED States 0 Un		CY216881	на	08-Jan-2017	homo sapiene; age	United States	1752		6.	9.00 -	625	
CY21603 HA 05-Jan-2017 horo saperns: United States 1752 CY21605 HA 04-Jan-2017 horo saperns: United States 1752 gender m: sge 59 CY21605 HA 04-Jan-2017 horo saperns: United States 1752 gender m: sge 59 CC		CY216689	на	04-Jan-2017	homo sapiens;	United States	1752		- 0" UNI	115	15	
CY216005 HA 04-Jan-2017 horn sage 55 United States 1752 C Show Heatmap Q Viewing Q Selected		CY216807	НА	03-Jan-2017	homo sapiens;	United States	1752		· • • •			
C Show Heatmap 9 Viewing 9 Selected		CY216805	на	04-Jan-2017	homo sapiens;	United States	1752		11.	•	1	
C				ta a land	gender m; age 59	11.12			Show Heatmap	Viewing Selected		
		CY216873 CY216881 CY216881 CY216883 CY216803	HA HA HA HA	03-Jan-2017 06-Jan-2017 05-Jan-2017 04-Jan-2017	homo sapirati, age 47 homo sapirati, age homo sapirati, gender m; age 3 homo sapirati, gender f; age 36 homo sapirati, gender m; age 59	United States United States United States United States United States	1752 1752 1752 1752		Show Heatmap (Viewing © Selected		
	-	- ??	L		-	(ST)	5 -	-11				
	-	Y			5	3-2	21	-F	-		=	
	-	5-			Y	2 hor	hat	me		-	10	
	2 Lis	Y	-		X	For	R	A A			E H	
	1	-	-		A	211	2	18				
	1.545				1.7		21			L	0	
	me: 20	17.056			4		3				-LFF	
	and the diverse	and realized by 25	- if it bing	50								



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online Journal of Public Health Informatics * ISSN 1947-2579 * http://ojphi.org * 11(1): e301, 2019



Figure1. ZooPhy search portal. A) A search for pdm09 H1N1 hemagglutinin (HA) sequences for 2017 in the U.S. B) The returned search result, including a U.S. heatmap of 361 pdm09 sequences. Before running the analytic pipeline, we show the geographic distribution of samples and enable the user to download the metadata and unaligned (FASTA) sequence file. C) We demonstrate geospatial results pertaining to spread and evolution that we will soon implement into ZooPhy.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online Journal of Public Health Informatics * ISSN 1947-2579 * http://ojphi.org * 11(1): e301, 2019