Pathogen pipelines

- Research Update-

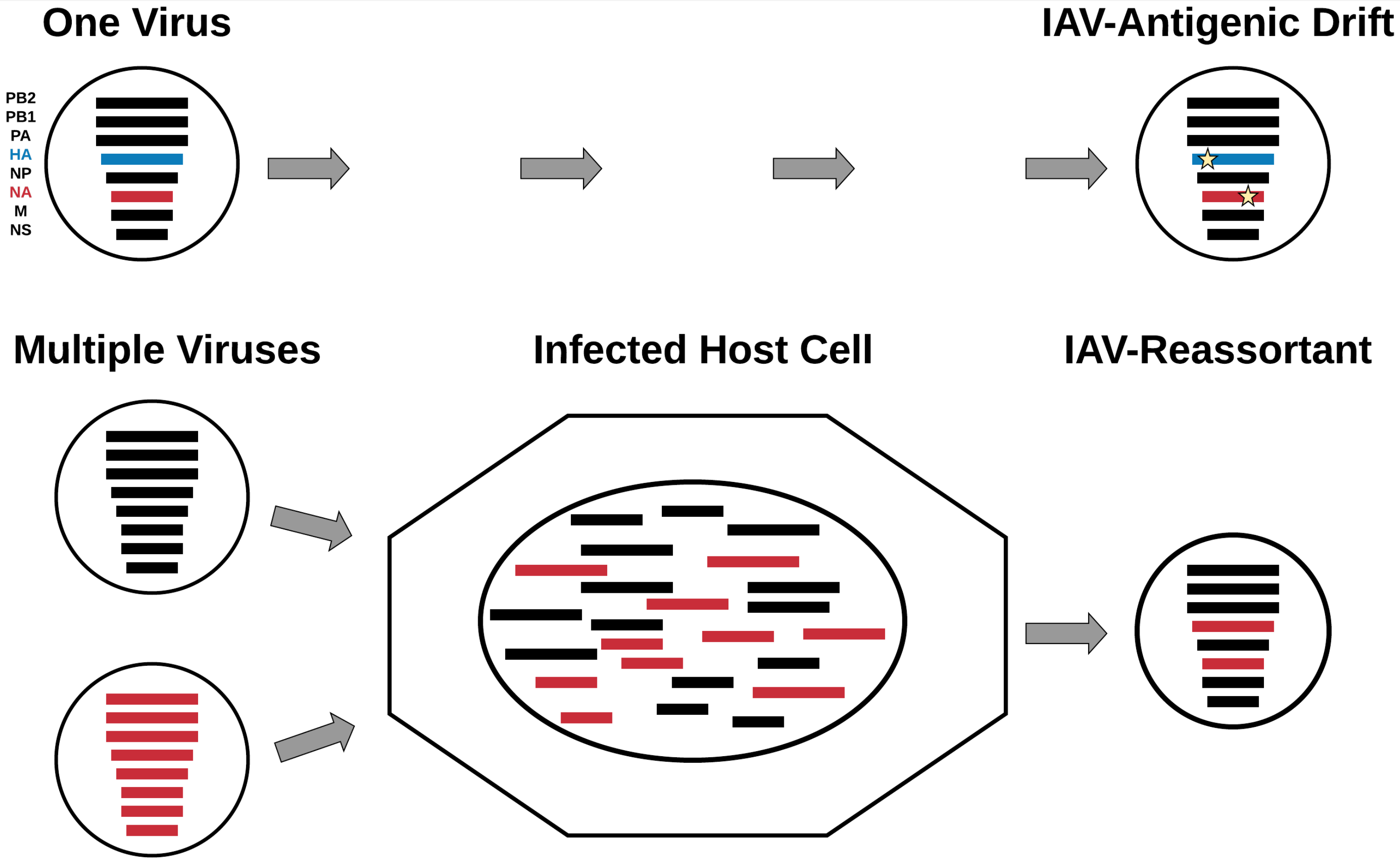
Jennifer Chang, Ph.D.

Bioinformatician, Bedford Lab
ORISE-Post Doc, Vincent Lab
Bioinformatics & Computational Biology Program, ISU

Outline

- Influenza A Viruses in Swine
  - Swine Surveillance key question
  - Reassortment estimates
- Zika and Measles
  - What are the key questions?
  - Metadata merging
- Conclusions and Future Directions

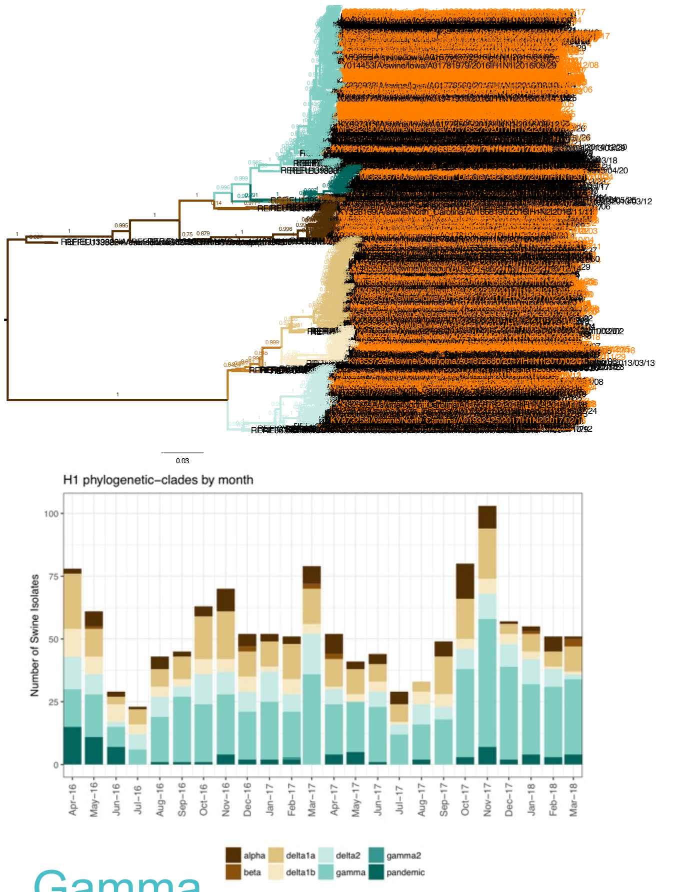
Influenza A Viruses (IAV)



- Influenza A Viruses (IAV) contains 8 RNA segments: PB2, PB1, PA, HA, NP, NA, M, NS
- Evolutionary changes accrue via **Antigenic Drift** or **Antigenic Shift** (reassortment)

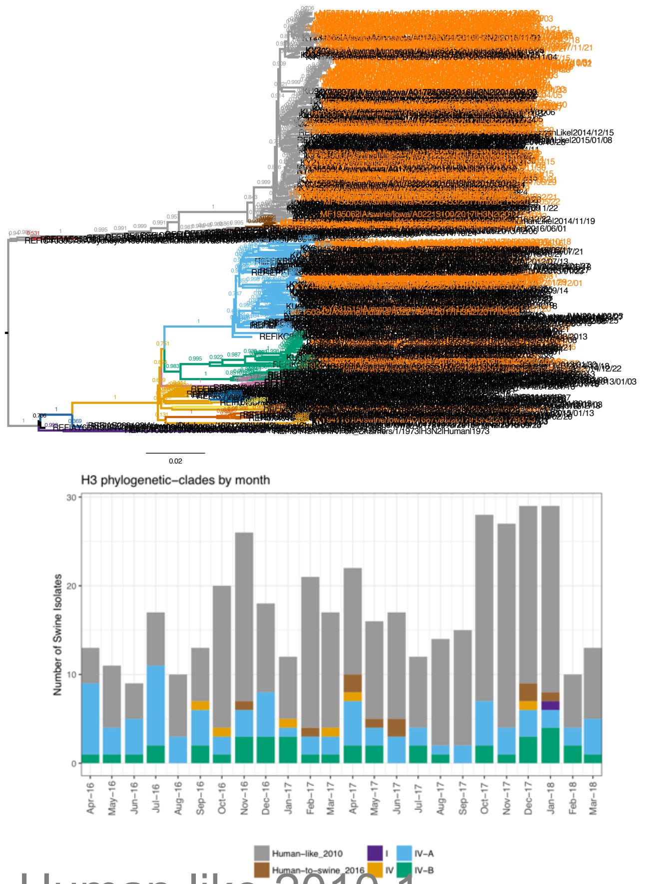
Hemagglutinin in Swine IAV (H1 or H3)

swine H1



H1. Gamma

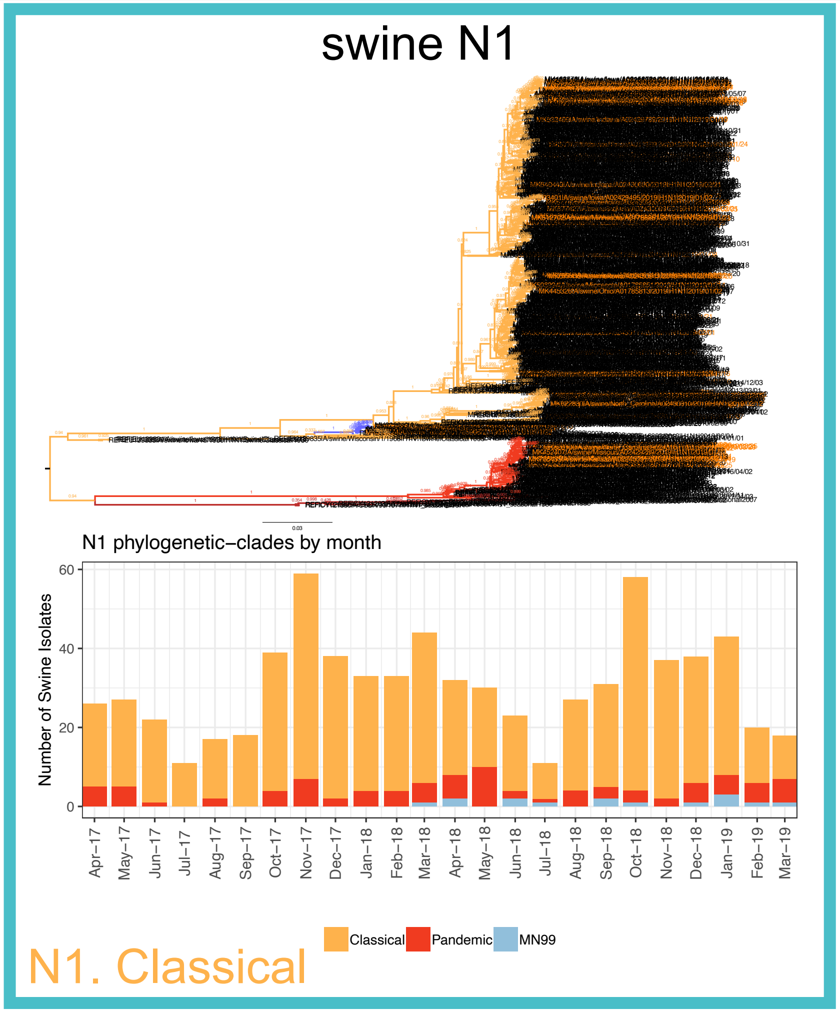
swine H3



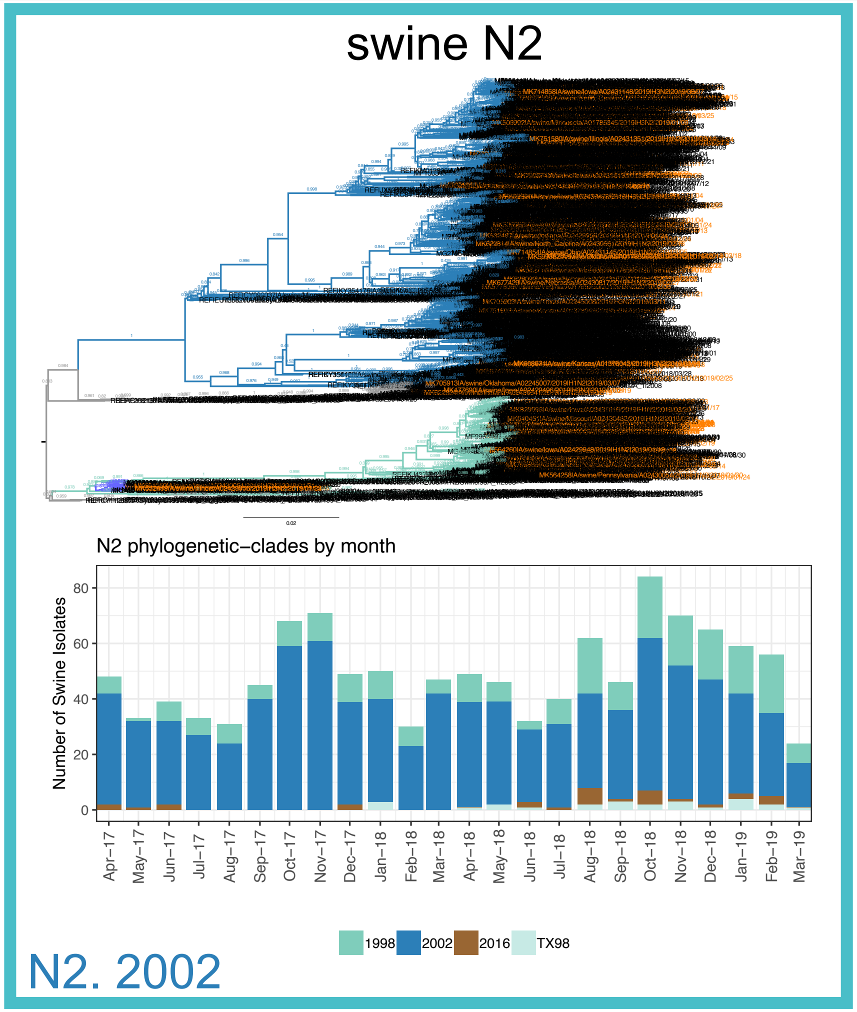
H3. Human-like 2010.1

- Hemagglutinin is a surface protein of IAV responsible for binding sialic acid on the host cell
- Under evolutionary pressure to evade host immune response

Neuraminidase in Swine IAV (N1 or N2)



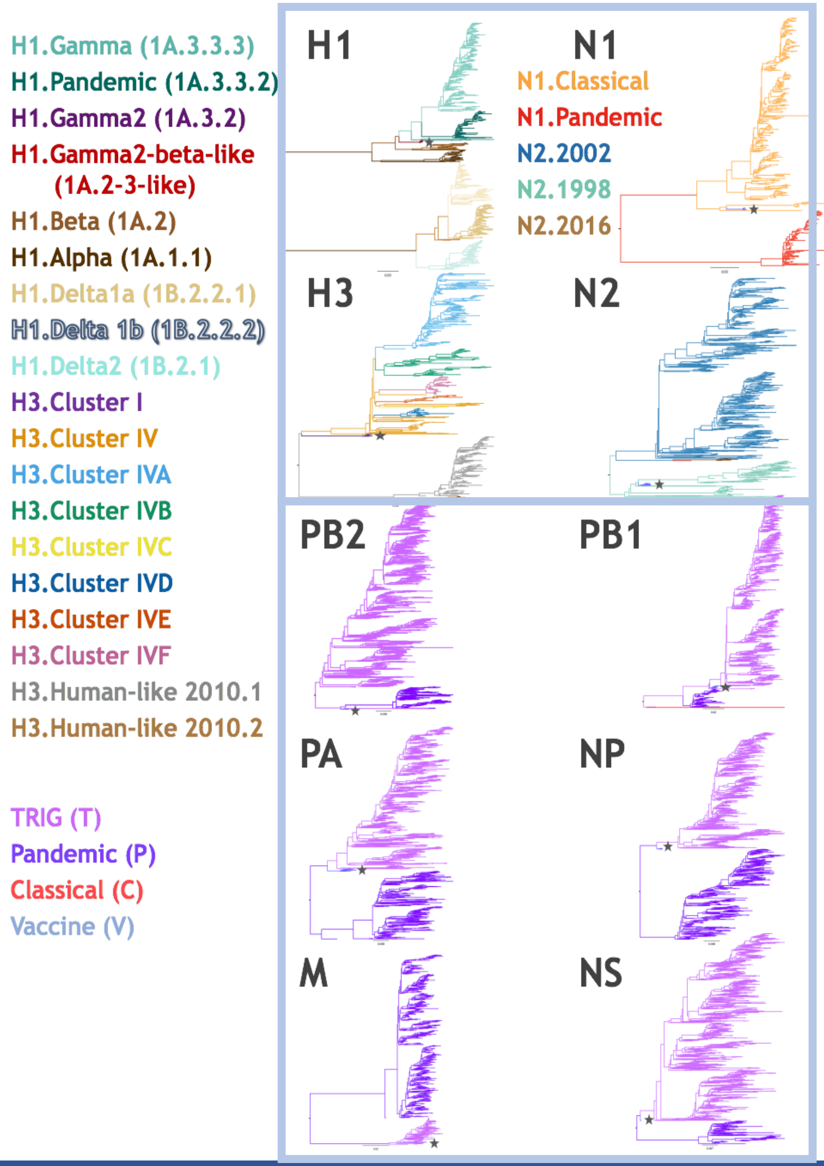
N1. Classical



N2. 2002

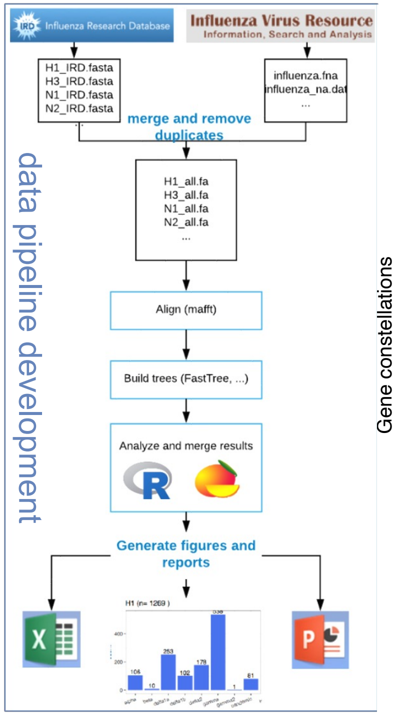
- Neuraminidase is a surface protein of IAV responsible for release of the virus into the host cell
- Under evolutionary pressure to evade host immune response

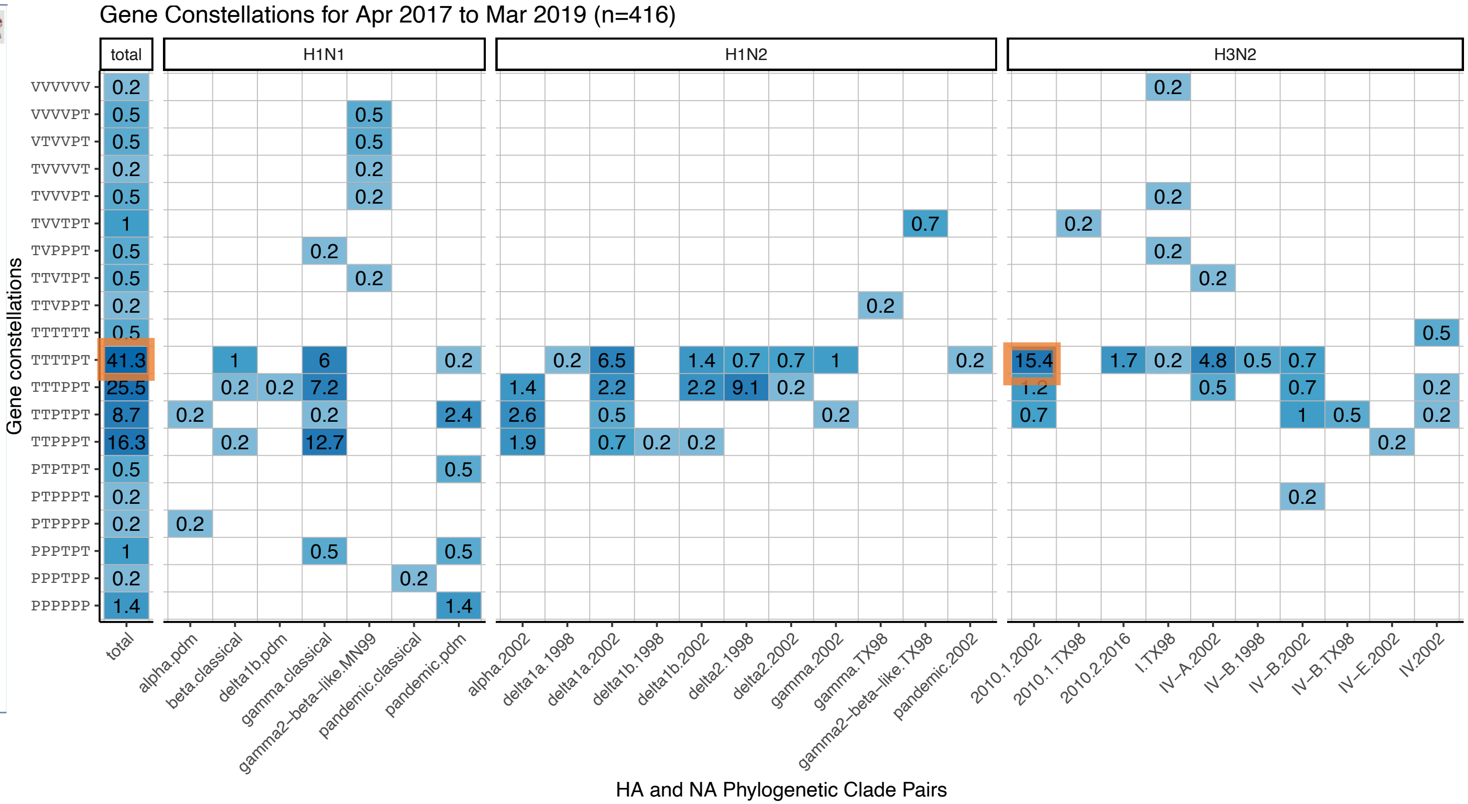
Influenza A Viruses



- Influenza A Virus contains 8 RNA segments
- Internal Genes (PB2 – NS) are phylogenetically classified as TRIG (T) or PDM (P)
- The 6-letter code of internal genes is considered the gene constellation of a strain (e.g. TTTTPT, TTPPPT)

Gene constellations of Influenza A Viruses in Swine





Gene Constellations for Apr 2017 to Mar 2019 (n=416)

Gene constellations	total
VVVVVV	0.2
VVVVPT	0.5
VTVVPT	0.5
TVVVVT	0.2
TVVVPT	0.5
TVVTPT	1
TVPPPT	0.5
TTVTPT	0.5
TTVPPT	0.2
TTTTTT	0.5
TTTTPT	41.3
TTTPPT	25.5
TTPTPT	8.7
TTPPPT	16.3
PTPTPT	0.5
PTPPPT	0.2
PTPPPP	0.2
PPPTPT	1
PPPTPP	0.2
PPPPPP	1.4

H1N1

Gene constellations	alpha.pdm	beta.classical	delta1b.pdm	gamma.classical	gamma2–beta–like.MN99	pandemic.classical	pandemic.pdm
VVVVVV							
VVVVPT					0.5		
VTVVPT					0.5		
TVVVVT					0.2		
TVVVPT					0.2		
TVVTPT							
TVPPPT				0.2			
TTVTPT					0.2		
TTVPPT							
TTTTTT							
TTTTPT		1		6			0.2
TTTPPT		0.2	0.2	7.2			
TTPTPT	0.2			0.2			2.4
TTPPPT		0.2		12.7			
PTPTPT							0.5
PTPPPT							
PTPPPP	0.2						
PPPTPT				0.5			0.5
PPPTPP						0.2	
PPPPPP							1.4

H1N2

Gene constellations	alpha.2002	delta1a.1998	delta1a.2002	delta1b.1998	delta1b.2002	delta2.1998	delta2.2002	gamma.2002	gamma.TX98	gamma2–beta–like.TX98	pandemic.2002
VVVVVV											
VVVVPT											
VTVVPT											
TVVVVT											
TVVVPT											
TVVTPT										0.7	
TVPPPT											
TTVTPT											
TTVPPT									0.2		
TTTTTT											
TTTTPT		0.2	6.5		1.4	0.7	0.7	1			0.2
TTTPPT	1.4		2.2		2.2	9.1	0.2				
TTPTPT	2.6		0.5					0.2			
TTPPPT	1.9		0.7	0.2	0.2						
PTPTPT											
PTPPPT											
PTPPPP											
PPPTPT											
PPPTPP											
PPPPPP											

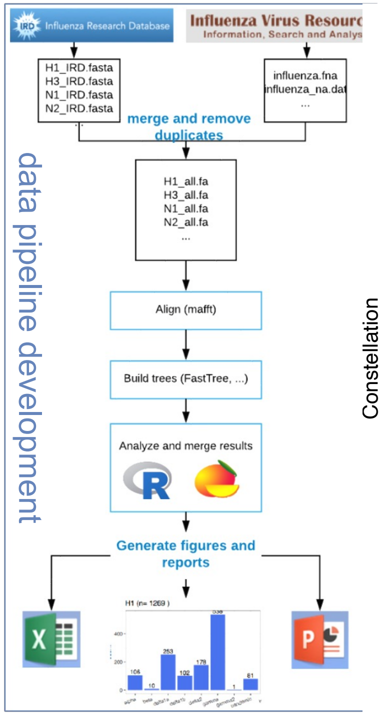
H3N2

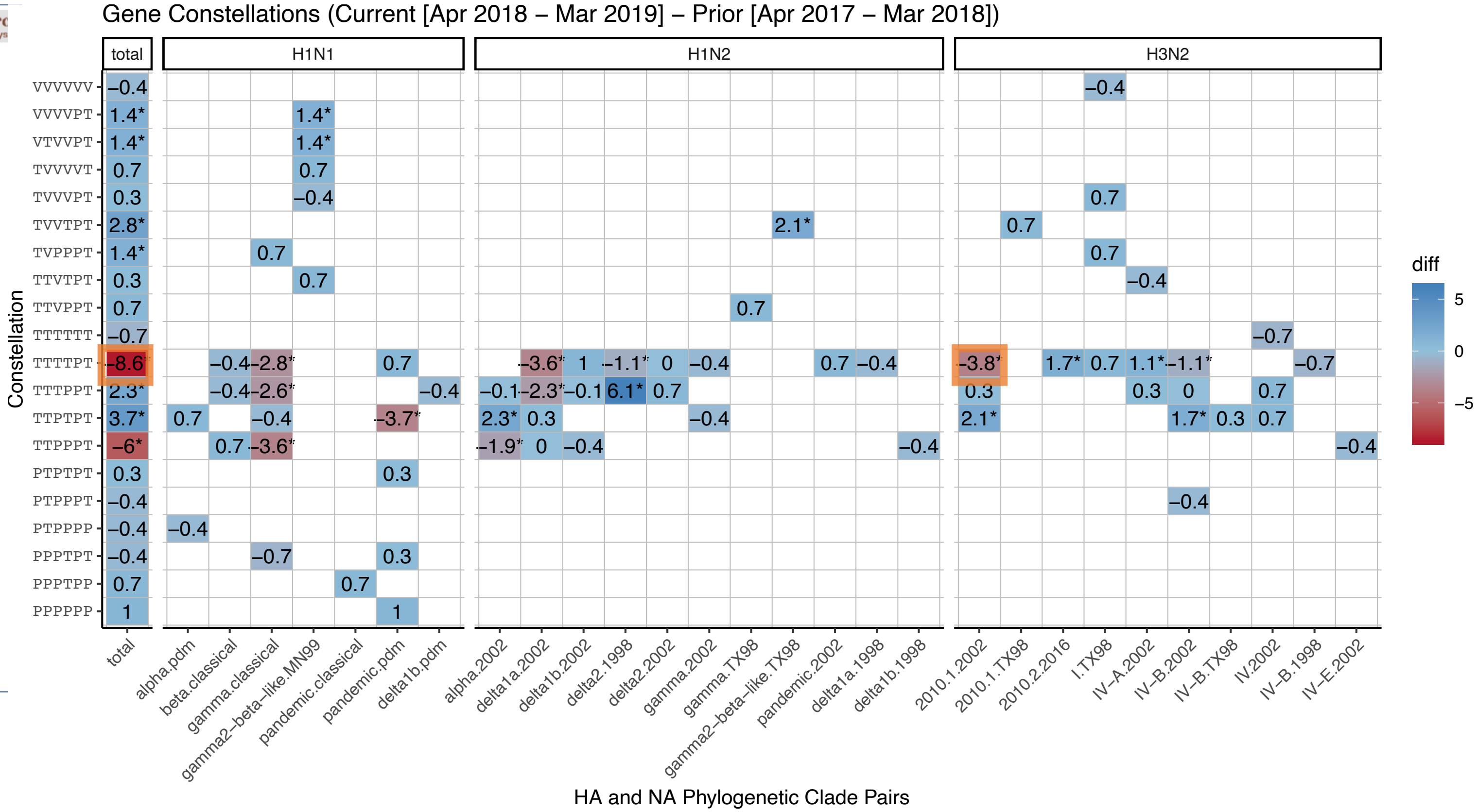
Gene constellations	2010.1.2002	2010.1.TX98	2010.2.2016	I.TX98	IV–A.2002	IV–B.1998	IV–B.2002	IV–B.TX98	IV–E.2002	IV.2002
VVVVVV				0.2						
VVVVPT										
VTVVPT										
TVVVVT										
TVVVPT				0.2						
TVVTPT		0.2								
TVPPPT				0.2						
TTVTPT					0.2					
TTVPPT										
TTTTTT										0.5
TTTTPT	15.4		1.7	0.2	4.8	0.5	0.7			
TTTPPT	1.2				0.5		0.7			0.2
TTPTPT	0.7						1	0.5		0.2
TTPPPT									0.2	
PTPTPT										
PTPPPT							0.2			
PTPPPP										
PPPTPT										
PPPTPP										
PPPPPP										

HA and NA Phylogenetic Clade Pairs

- **T**=Triple Reassortment (TRIG); **P**=Pandemic; **V**=LAIV-related
- Listed in order: PB2, PB1, PA, NP, M, NS

Gene constellations of Influenza A Viruses in Swine




Gene Constellations (Current [Apr 2018 – Mar 2019] – Prior [Apr 2017 – Mar 2018])

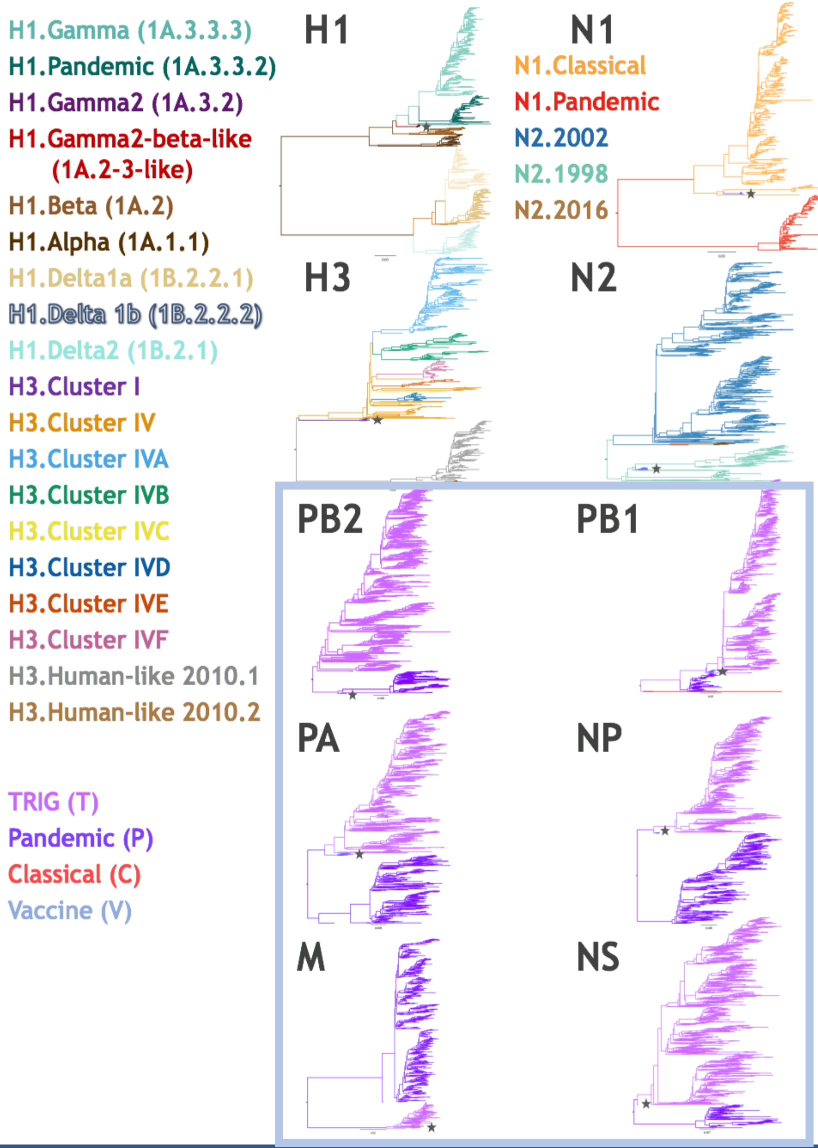
Constellation	total	H1N1 alpha.pdm	H1N1 beta.classical	H1N1 gamma.classical	H1N1 gamma2−beta−like.MN99	H1N1 pandemic.classical	H1N1 pandemic.pdm	H1N1 delta1b.pdm
VVVVVV	−0.4							
VVVVPT	1.4*				1.4*			
VTVVPT	1.4*				1.4*			
TVVVVT	0.7				0.7			
TVVVPT	0.3				−0.4			
TVVTPT	2.8*							
TVPPPT	1.4*			0.7				
TTVTPT	0.3				0.7			
TTVPPT	0.7							
TTTTTT	−0.7							
TTTTPT	−8.6*		−0.4	−2.8*			0.7	
TTTPPT	2.3*		−0.4	−2.6*				−0.4
TTPTPT	3.7*	0.7		−0.4			−3.7*	
TTPPPT	−6*		0.7	−3.6*				
PTPTPT	0.3						0.3	
PTPPPT	−0.4							
PTPPPP	−0.4	−0.4						
PPPTPT	−0.4			−0.7			0.3	
PPPTPP	0.7					0.7		
PPPPPP	1						1	

Constellation	H1N2 alpha.2002	H1N2 delta1a.2002	H1N2 delta1b.2002	H1N2 delta2.1998	H1N2 delta2.2002	H1N2 gamma.2002	H1N2 gamma.TX98	H1N2 gamma2−beta−like.TX98	H1N2 pandemic.2002	H1N2 delta1a.1998	H1N2 delta1b.1998
VVVVVV											
VVVVPT											
VTVVPT											
TVVVVT											
TVVVPT											
TVVTPT								2.1*			
TVPPPT											
TTVTPT											
TTVPPT							0.7				
TTTTTT											
TTTTPT		−3.6*	1	−1.1*	0	−0.4			0.7	−0.4	
TTTPPT	−0.1	−2.3*	−0.1	6.1*	0.7						
TTPTPT	2.3*	0.3				−0.4					
TTPPPT	−1.9*	0	−0.4								−0.4
PTPTPT											
PTPPPT											
PTPPPP											
PPPTPT											
PPPTPP											
PPPPPP											

Constellation	H3N2 2010.1.2002	H3N2 2010.1.TX98	H3N2 2010.2.2016	H3N2 I.TX98	H3N2 IV−A.2002	H3N2 IV−B.2002	H3N2 IV−B.TX98	H3N2 IV.2002	H3N2 IV−B.1998	H3N2 IV−E.2002
VVVVVV				−0.4						
VVVVPT										
VTVVPT										
TVVVVT										
TVVVPT				0.7						
TVVTPT		0.7								
TVPPPT				0.7						
TTVTPT					−0.4					
TTVPPT										
TTTTTT								−0.7		
TTTTPT	−3.8*		1.7*	0.7	1.1*	−1.1*			−0.7	
TTTPPT	0.3				0.3	0		0.7		
TTPTPT	2.1*					1.7*	0.3	0.7		
TTPPPT										−0.4
PTPTPT										
PTPPPT						−0.4				
PTPPPP										
PPPTPT										
PPPTPP										
PPPPPP										

- **T**=Triple Reassortment (TRIG); **P**=Pandemic; **V**=LAIV-related
- Listed in order: PB2, PB1, PA, NP, M, NS

Influenza A Viruses

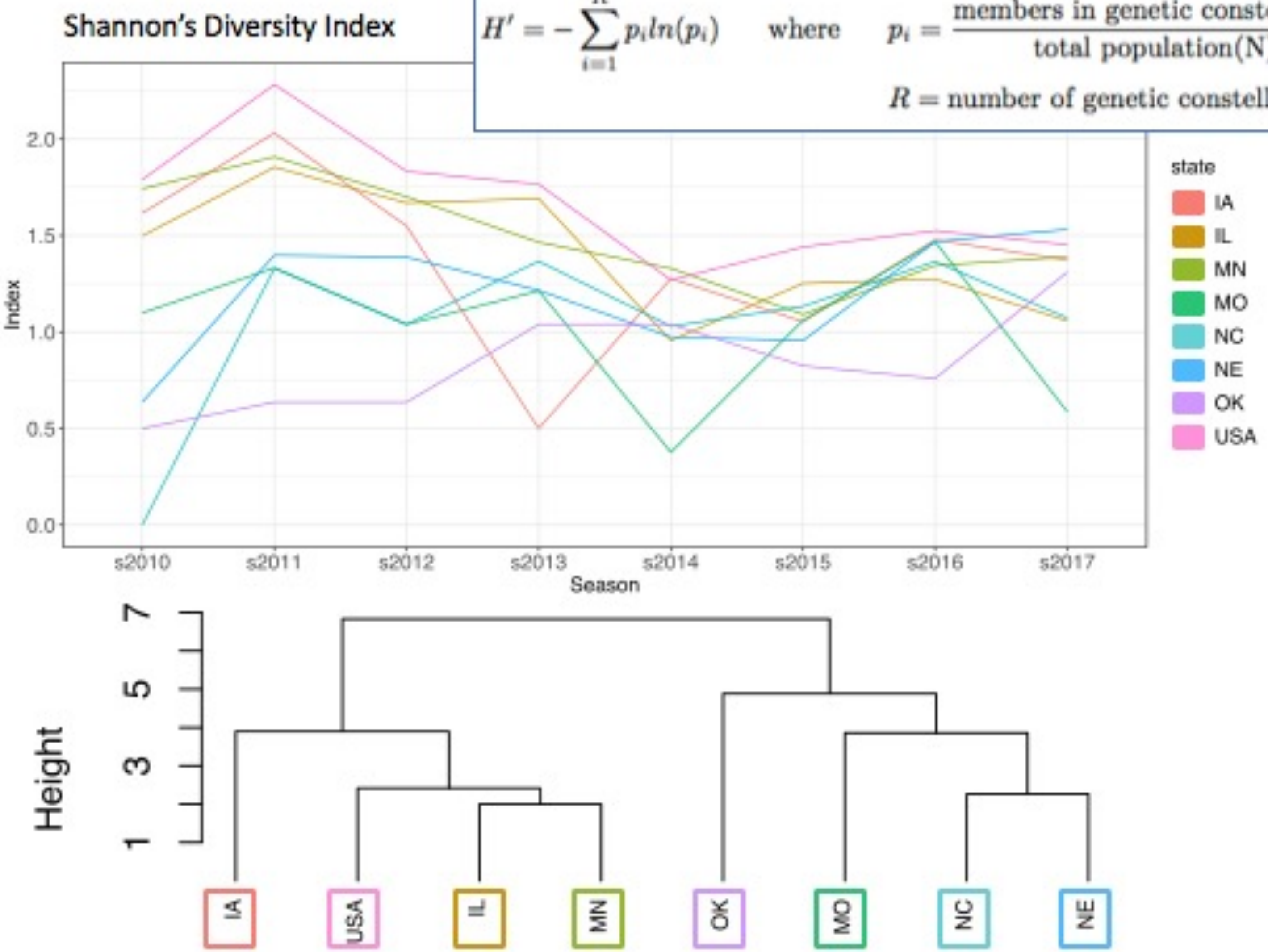


- Influenza A Virus contains 8 RNA segments
- Internal Genes (PB2 – NS) are phylogenetically classified as TRIG (T) or PDM (P)
- The 6-letter code of internal genes is considered the gene constellation of a strain (e.g. TTTTPT, TTPPPT)
- How might we measure gene constellation diversity across U.S. States?
- Since these 8 segments can reassort, can we identify reassortment events?
- Can we identify spatial and temporal patterns of reassortment?

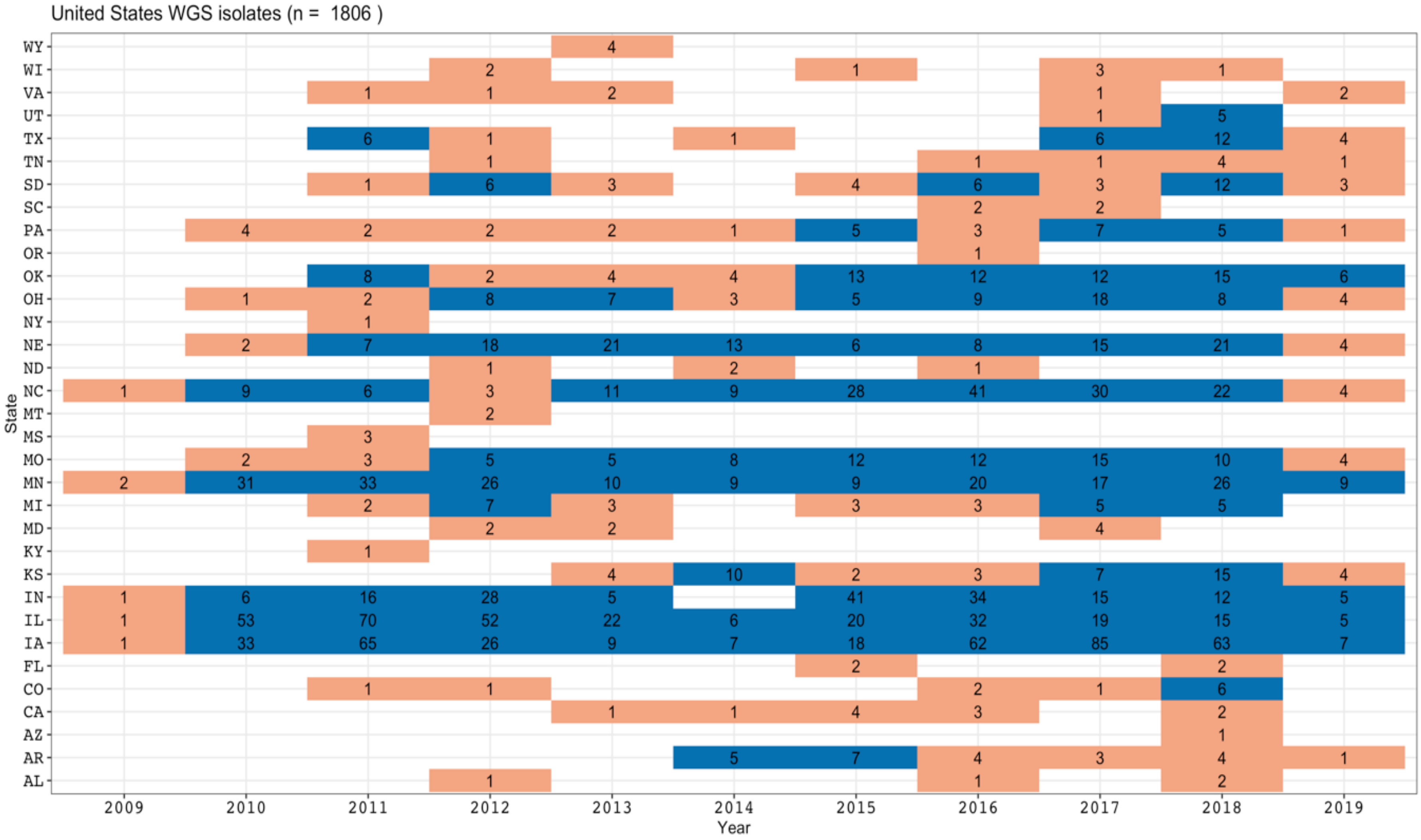
Measuring gene constellation diversity across states

$$H' = -\sum_{i=1}^{R} p_i ln(p_i) \quad \text{where} \quad p_i = \frac{\text{members in genetic constellation}}{\text{total population(N)}}$$

$$R = \text{number of genetic constellations}$$



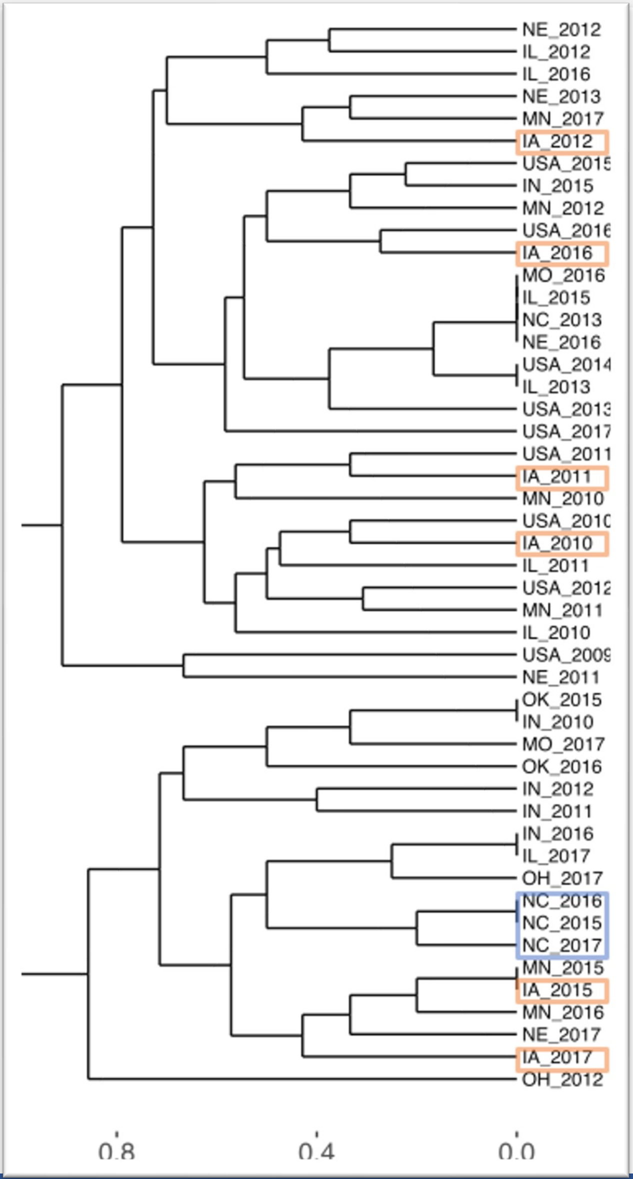
However several states are not included



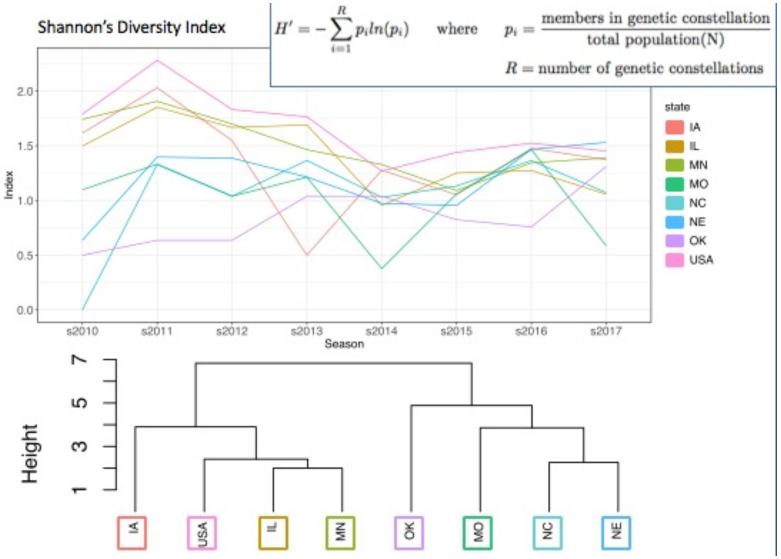
United States WGS isolates (n = 1806)

State \ Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
WY					4						
WI				2			1		3	1	
VA			1	1	2				1		2
UT									1	5	
TX			6	1		1			6	12	4
TN				1				1	1	4	1
SD			1	6	3		4	6	3	12	3
SC								2	2		
PA		4	2	2	2	1	5	3	7	5	1
OR								1			
OK			8	2	4	4	13	12	12	15	6
OH		1	2	8	7	3	5	9	18	8	4
NY			1								
NE		2	7	18	21	13	6	8	15	21	4
ND				1		2		1			
NC	1	9	6	3	11	9	28	41	30	22	4
MT				2							
MS			3								
MO		2	3	5	5	8	12	12	15	10	4
MN	2	31	33	26	10	9	9	20	17	26	9
MI			2	7	3		3	3	5	5	
MD				2	2				4		
KY			1								
KS					4	10	2	3	7	15	4
IN	1	6	16	28	5		41	34	15	12	5
IL	1	53	70	52	22	6	20	32	19	15	5
IA	1	33	65	26	9	7	18	62	85	63	7
FL							2			2	
CO			1	1				2	1	6	
CA					1	1	4	3		2	
AZ										1	
AR						5	7	4	3	4	1
AL				1				1		2	

Measuring gene constellation diversity across states

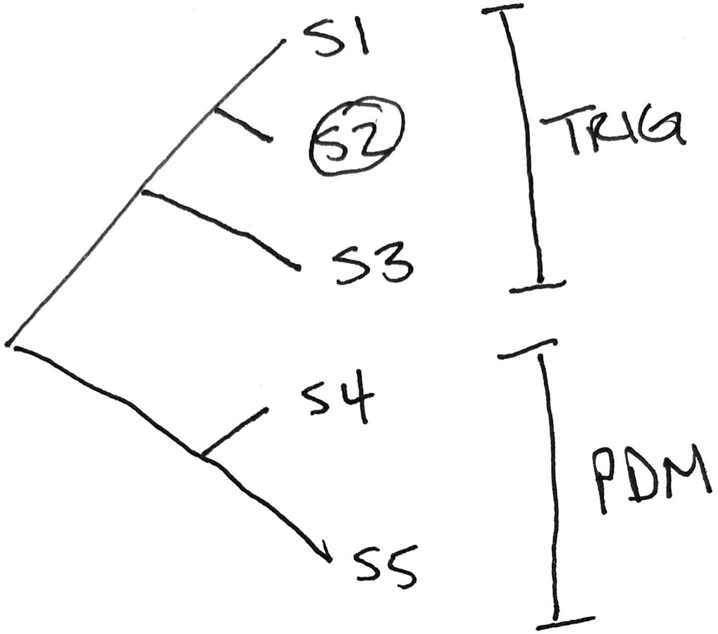


Binary distance of gene constellations exhibited in states across years. If constellations are consistent across time, we expect them to cluster together (NC). States with more variable gene constellations across time are spread out (IA).

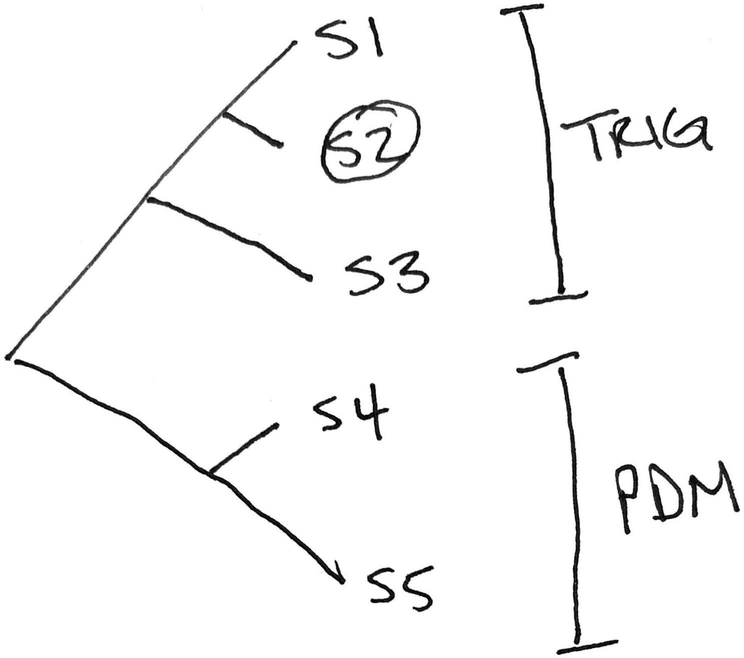


Identifying reassortment events

M



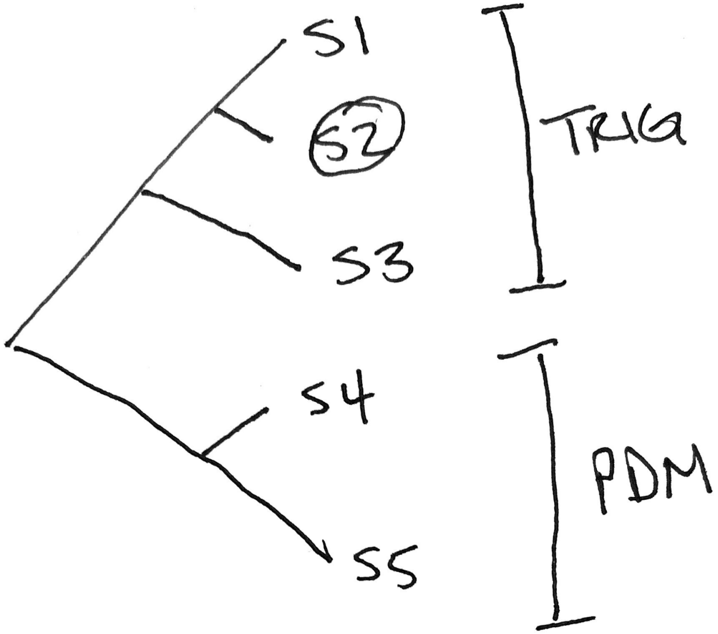
NS



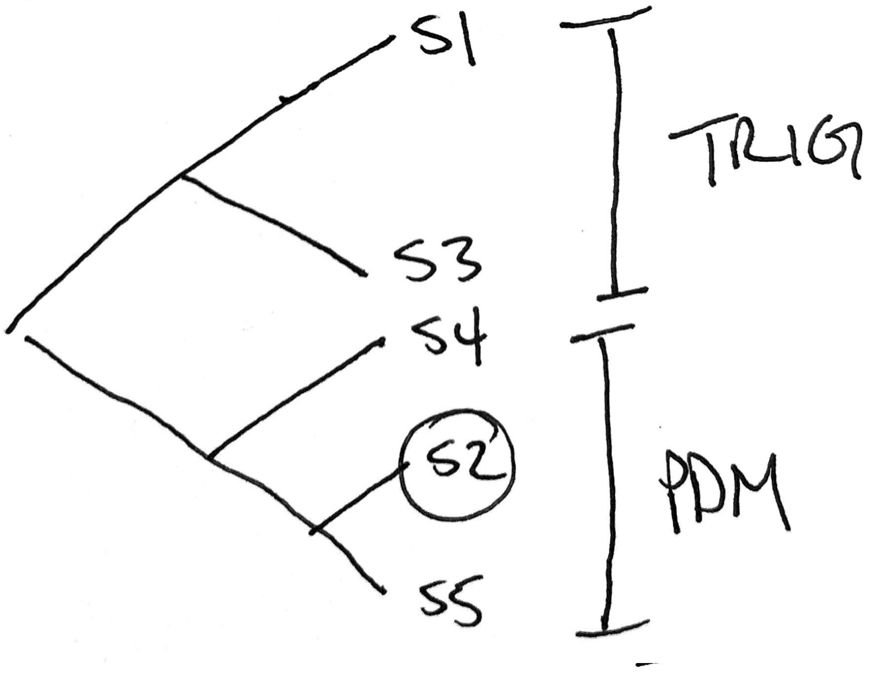
Focusing on only two genes (M and NS), assuming reassortment does not occur, their gene trees should be consistent.

Identifying reassortment events

M



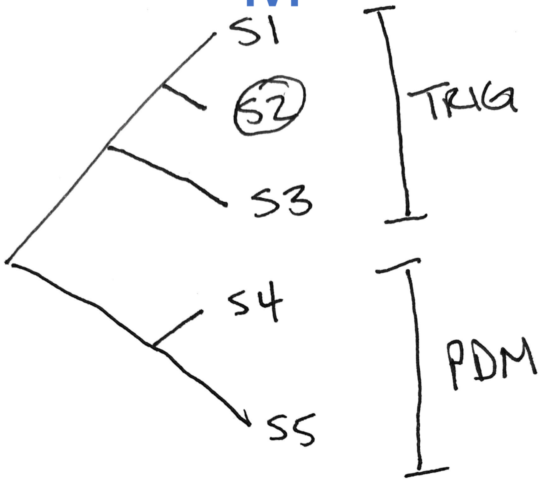
NS



However if the tree topology is not consistent, it is an indication of possible reassortment. Possible antigenic shift occurring.

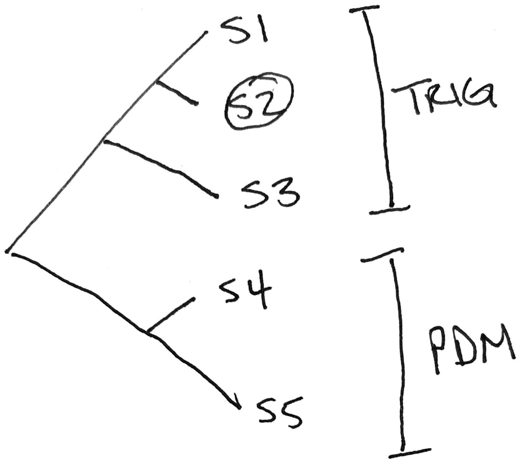
Identifying reassortment events

M



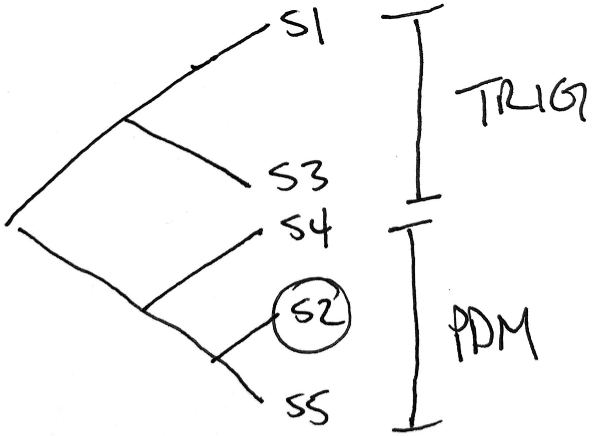
S2 M gene
- S1
- S3
- S4

NS



S2 NS gene
- S1
- S3
- S4

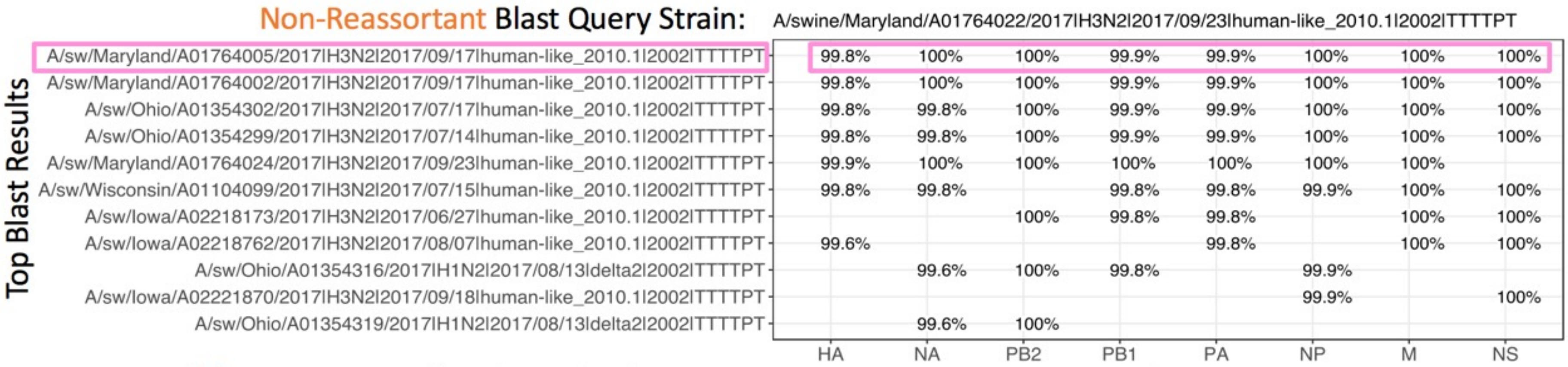
NS-reassorted



S2 NS gene
- S5
- S4
- S3

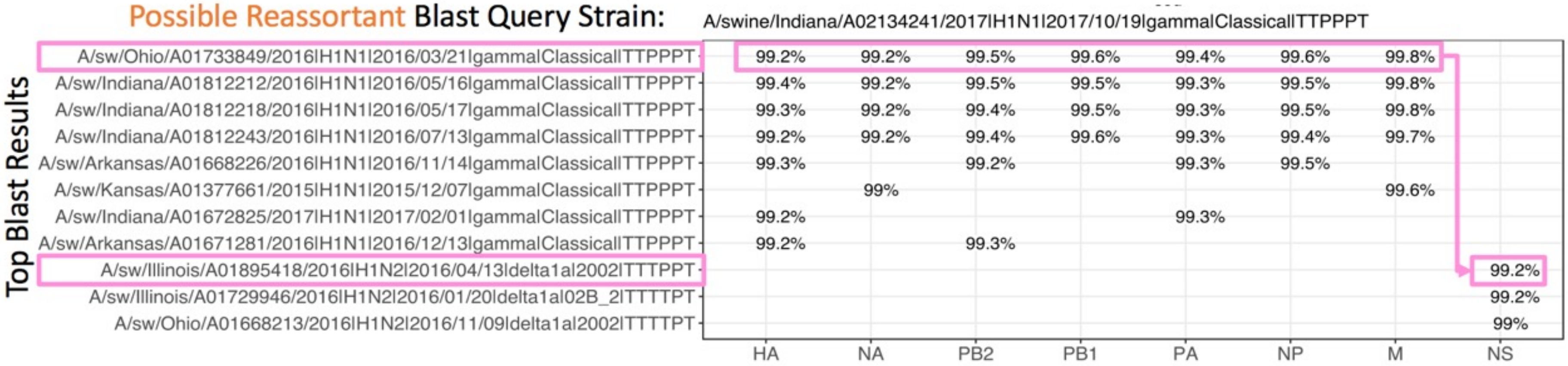
The 3 blast hits for each individual gene of S2 should be the the same.
If it is not, then reassortment may have occurred.

Identifying reassortment events



Non-Reassortant Blast Query Strain: A/swine/Maryland/A01764022/2017\|H3N2\|2017/09/23\|human-like_2010.1\|2002\|TTTTPT

Top Blast Results	HA	NA	PB2	PB1	PA	NP	M	NS
A/sw/Maryland/A01764005/2017|H3N2|2017/09/17|human-like_2010.1|2002|TTTTPT	99.8%	100%	100%	99.9%	99.9%	100%	100%	100%
A/sw/Maryland/A01764002/2017|H3N2|2017/09/17|human-like_2010.1|2002|TTTTPT	99.8%	100%	100%	99.9%	99.9%	100%	100%	100%
A/sw/Ohio/A01354302/2017|H3N2|2017/07/17|human-like_2010.1|2002|TTTTPT	99.8%	99.8%	100%	99.9%	99.9%	100%	100%	100%
A/sw/Ohio/A01354299/2017|H3N2|2017/07/14|human-like_2010.1|2002|TTTTPT	99.8%	99.8%	100%	99.9%	99.9%	100%	100%	100%
A/sw/Maryland/A01764024/2017|H3N2|2017/09/23|human-like_2010.1|2002|TTTTPT	99.9%	100%	100%	100%	100%	100%	100%	
A/sw/Wisconsin/A01104099/2017|H3N2|2017/07/15|human-like_2010.1|2002|TTTTPT	99.8%	99.8%		99.8%	99.8%	99.9%	100%	100%
A/sw/Iowa/A02218173/2017|H3N2|2017/06/27|human-like_2010.1|2002|TTTTPT			100%	99.8%	99.8%		100%	100%
A/sw/Iowa/A02218762/2017|H3N2|2017/08/07|human-like_2010.1|2002|TTTTPT	99.6%				99.8%		100%	100%
A/sw/Ohio/A01354316/2017|H1N2|2017/08/13|delta2|2002|TTTTPT		99.6%	100%	99.8%		99.9%		
A/sw/Iowa/A02221870/2017|H3N2|2017/09/18|human-like_2010.1|2002|TTTTPT						99.9%		100%
A/sw/Ohio/A01354319/2017|H1N2|2017/08/13|delta2|2002|TTTTPT		99.6%	100%					



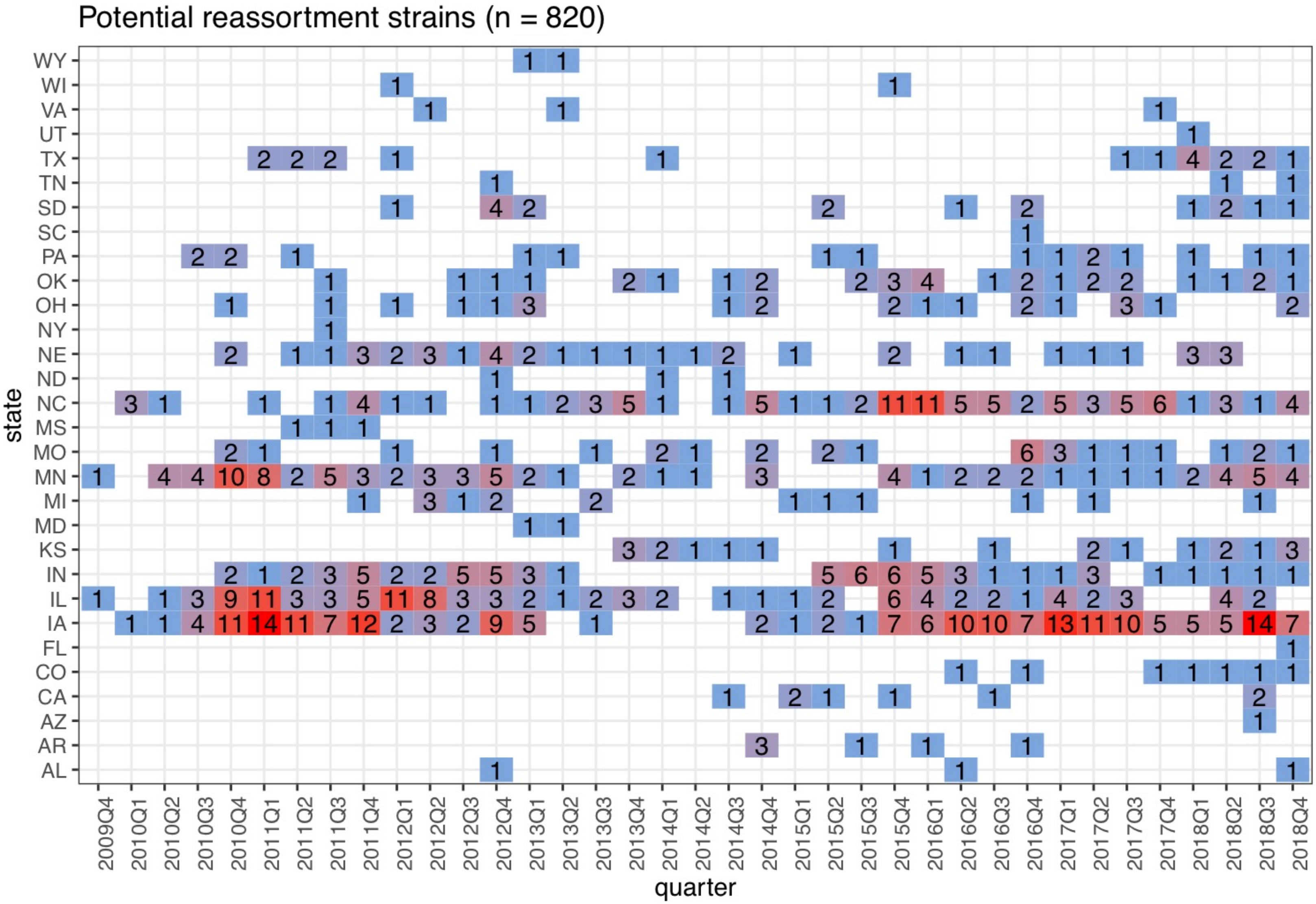
Possible Reassortant Blast Query Strain: A/swine/Indiana/A02134241/2017\|H1N1\|2017/10/19\|gamma\|Classical\|TTPPPT

Top Blast Results	HA	NA	PB2	PB1	PA	NP	M	NS
A/sw/Ohio/A01733849/2016|H1N1|2016/03/21|gamma|Classical|TTPPPT	99.2%	99.2%	99.5%	99.6%	99.4%	99.6%	99.8%	
A/sw/Indiana/A01812212/2016|H1N1|2016/05/16|gamma|Classical|TTPPPT	99.4%	99.2%	99.5%	99.5%	99.3%	99.5%	99.8%	
A/sw/Indiana/A01812218/2016|H1N1|2016/05/17|gamma|Classical|TTPPPT	99.3%	99.2%	99.4%	99.5%	99.3%	99.5%	99.8%	
A/sw/Indiana/A01812243/2016|H1N1|2016/07/13|gamma|Classical|TTPPPT	99.2%	99.2%	99.4%	99.6%	99.3%	99.4%	99.7%	
A/sw/Arkansas/A01668226/2016|H1N1|2016/11/14|gamma|Classical|TTPPPT	99.3%		99.2%		99.3%	99.5%		
A/sw/Kansas/A01377661/2015|H1N1|2015/12/07|gamma|Classical|TTPPPT		99%					99.6%	
A/sw/Indiana/A01672825/2017|H1N1|2017/02/01|gamma|Classical|TTPPPT	99.2%				99.3%			
A/sw/Arkansas/A01671281/2016|H1N1|2016/12/13|gamma|Classical|TTPPPT	99.2%		99.3%					
A/sw/Illinois/A01895418/2016|H1N2|2016/04/13|delta1a|2002|TTTPPT								99.2%
A/sw/Illinois/A01729946/2016|H1N2|2016/01/20|delta1a|02B_2|TTTTPT								99.2%
A/sw/Ohio/A01668213/2016|H1N2|2016/11/09|delta1a|2002|TTTTPT								99%

Blast-based algorithm for detecting reassortants.

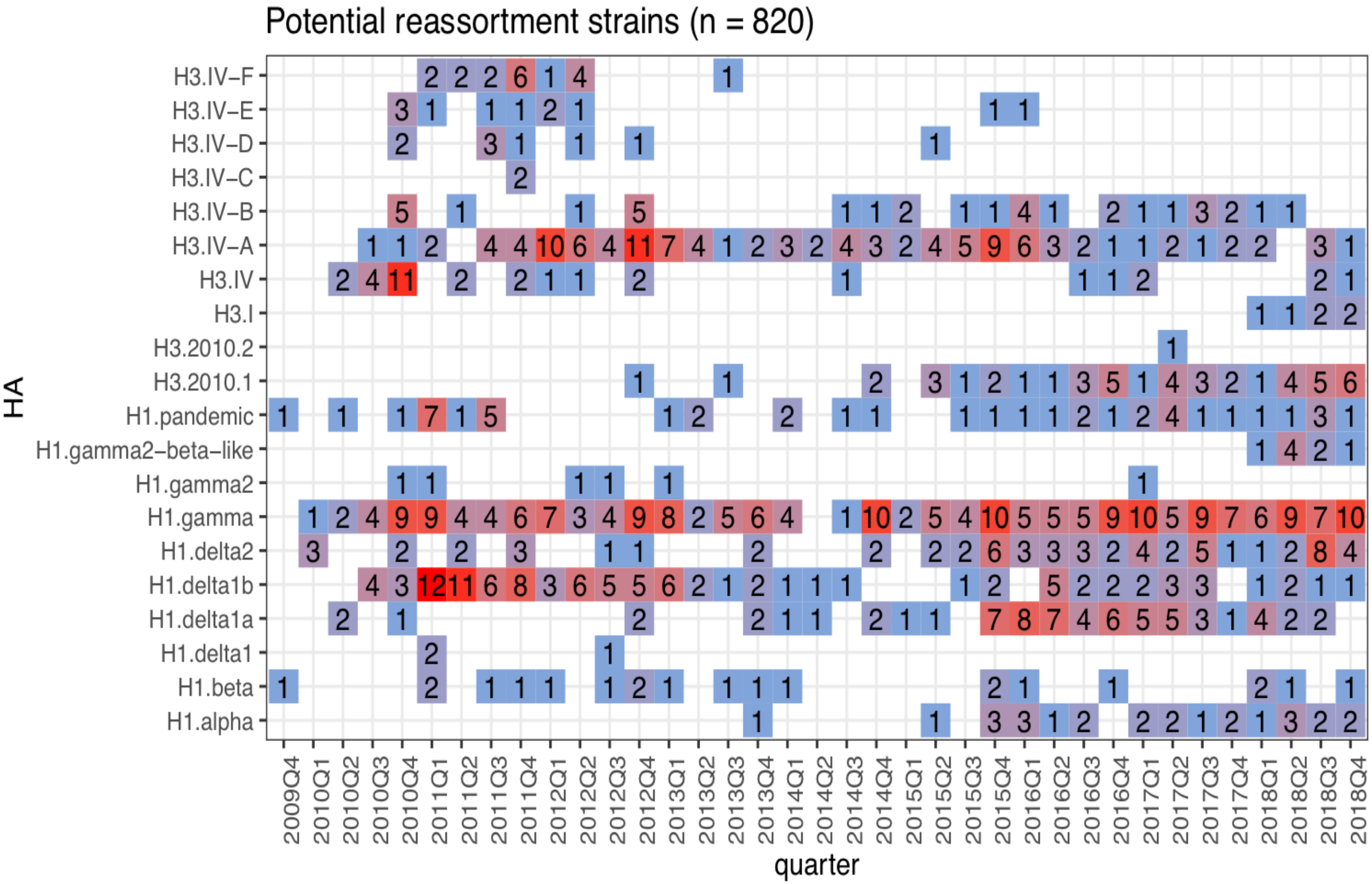
Most likely source strain or gene reassortant combination highlighted in pink

Reassortant strains by U.S. state


Potential reassortment strains (n = 820)

state	2009Q4	2010Q1	2010Q2	2010Q3	2010Q4	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2	2012Q3	2012Q4	2013Q1	2013Q2	2013Q3	2013Q4	2014Q1	2014Q2	2014Q3	2014Q4	2015Q1	2015Q2	2015Q3	2015Q4	2016Q1	2016Q2	2016Q3	2016Q4	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4
WY														1	1																						
WI										1															1												
VA											1				1																		1				
UT																																		1			
TX						2	2	2		1								1														1	1	4	2	2	1
TN													1																						1		1
SD										1			4	2									2				1		2					1	2	1	1
SC																													1								
PA				2	2		1							1	1								1	1					1	1	2	1		1		1	1
OK								1				1	1	1			2	1		1	2			2	3	4		1	2	1	2	2		1	1	2	1
OH					1			1		1		1	1	3						1	2				2	1	1		2	1		3	1				2
NY								1																													
NE					2		1	1	3	2	3	1	4	2	1	1	1	1	1	2		1			2		1	1		1	1	1		3	3		
ND													1					1		1																	
NC		3	1			1		1	4	1	1		1	1	2	3	5	1		1	5	1	1	2	11	11	5	5	2	5	3	5	6	1	3	1	4
MS							1	1	1																												
MO					2	1				1			1			1		2	1		2		2	1					6	3	1	1	1		1	2	1
MN	1		4	4	10	8	2	5	3	2	3	3	5	2	1		2	1	1		3				4	1	2	2	2	1	1	1	1	2	4	5	4
MI									1		3	1	2			2						1	1	1					1		1					1	
MD														1	1																						
KS																	3	2	1	1	1				1			1			2	1		1	2	1	3
IN					2	1	2	3	5	2	2	5	5	3	1								5	6	6	5	3	1	1	1	3		1	1	1	1	1
IL	1		1	3	9	11	3	3	5	11	8	3	3	2	1	2	3	2		1	1	1	2		6	4	2	2	1	4	2	3			4	2	
IA		1	1	4	11	14	11	7	12	2	3	2	9	5		1					2	1	2	1	7	6	10	10	7	13	11	10	5	5	5	14	7
FL																																					1
CO																											1		1				1	1	1	1	1
CA																				1		2	1		1			1								2	
AZ																																				1	
AR																					3			1		1			1								
AL													1														1										1

Reassortant strains by HA Clade



Potential reassortment strains (n = 820)

HA	2009Q4	2010Q1	2010Q2	2010Q3	2010Q4	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2	2012Q3	2012Q4	2013Q1	2013Q2	2013Q3	2013Q4	2014Q1	2014Q2	2014Q3	2014Q4	2015Q1	2015Q2	2015Q3	2015Q4	2016Q1	2016Q2	2016Q3	2016Q4	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4
H3.IV−F						2	2	2	6	1	4					1																					
H3.IV−E					3	1		1	1	2	1														1	1											
H3.IV−D					2			3	1		1		1										1														
H3.IV−C									2																												
H3.IV−B					5		1				1		5							1	1	2		1	1	4	1		2	1	1	3	2	1	1		
H3.IV−A				1	1	2		4	4	10	6	4	11	7	4	1	2	3	2	4	3	2	4	5	9	6	3	2	1	1	2	1	2	2		3	1
H3.IV			2	4	11		2		2	1	1		2							1								1	1	2						2	1
H3.I																																		1	1	2	2
H3.2010.2																															1						
H3.2010.1													1			1					2		3	1	2	1	1	3	5	1	4	3	2	1	4	5	6
H1.pandemic	1		1		1	7	1	5						1	2			2		1	1			1	1	1	1	2	1	2	4	1	1	1	1	3	1
H1.gamma2−beta−like																																		1	4	2	1
H1.gamma2					1	1					1	1		1																1							
H1.gamma		1	2	4	9	9	4	4	6	7	3	4	9	8	2	5	6	4		1	10	2	5	4	10	5	5	5	9	10	5	9	7	6	9	7	10
H1.delta2		3			2		2		3			1	1				2				2		2	2	6	3	3	3	2	4	2	5	1	1	2	8	4
H1.delta1b				4	3	12	11	6	8	3	6	5	5	6	2	1	2	1	1	1				1	2		5	2	2	2	3	3		1	2	1	1
H1.delta1a			2		1								2				2	1	1		2	1	1		7	8	7	4	6	5	5	3	1	4	2	2	
H1.delta1						2						1																									
H1.beta	1					2		1	1	1		1	2	1		1	1	1							2	1			1					2	1		1
H1.alpha																	1						1		3	3	1	2		2	2	1	2	1	3	2	2

quarter

Gene Constellation Conclusion

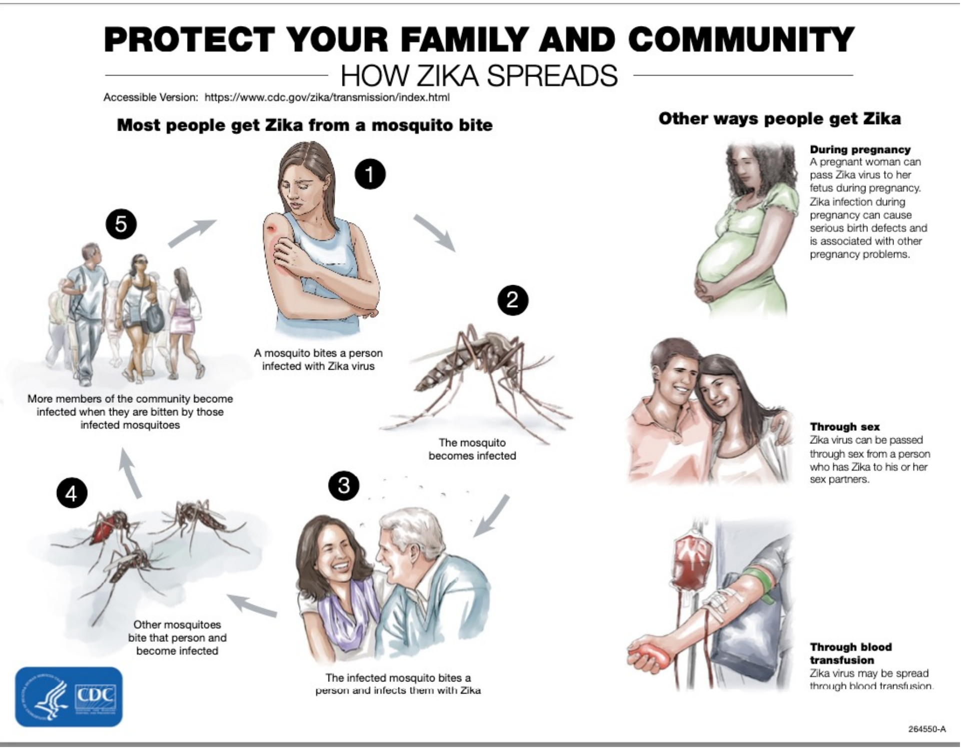
- A blast-based method may be a scalable way to identify reassortment events across all 8 gene segments
- Identifying potential reassortant strains can indicate which U.S. states may be hotspots of reassortment
  - Although counts of reassortment should be normalized by number of isolates collected in that state
- Certain combinations of 8 genes may be more prone to reassortment but further analysis is needed
- Future directions include streamlining a pipeline to tag new isolates that are potential reassortments

Outline

- Influenza A Viruses in Swine
  - Swine Surveillance Pipeline
  - Gene Constellation Manuscript
- Zika and Measles
  - What are the key questions?
- Conclusions and Future Directions

Zika

Measles



PROTECT YOUR FAMILY AND COMMUNITY

HOW ZIKA SPREADS

Accessible Version: https://www.cdc.gov/zika/transmission/index.html

Most people get Zika from a mosquito bite

1. A mosquito bites a person infected with Zika virus
2. The mosquito becomes infected
3. The infected mosquito bites a person and infects them with Zika
4. Other mosquitoes bite that person and become infected
5. More members of the community become infected when they are bitten by those infected mosquitoes

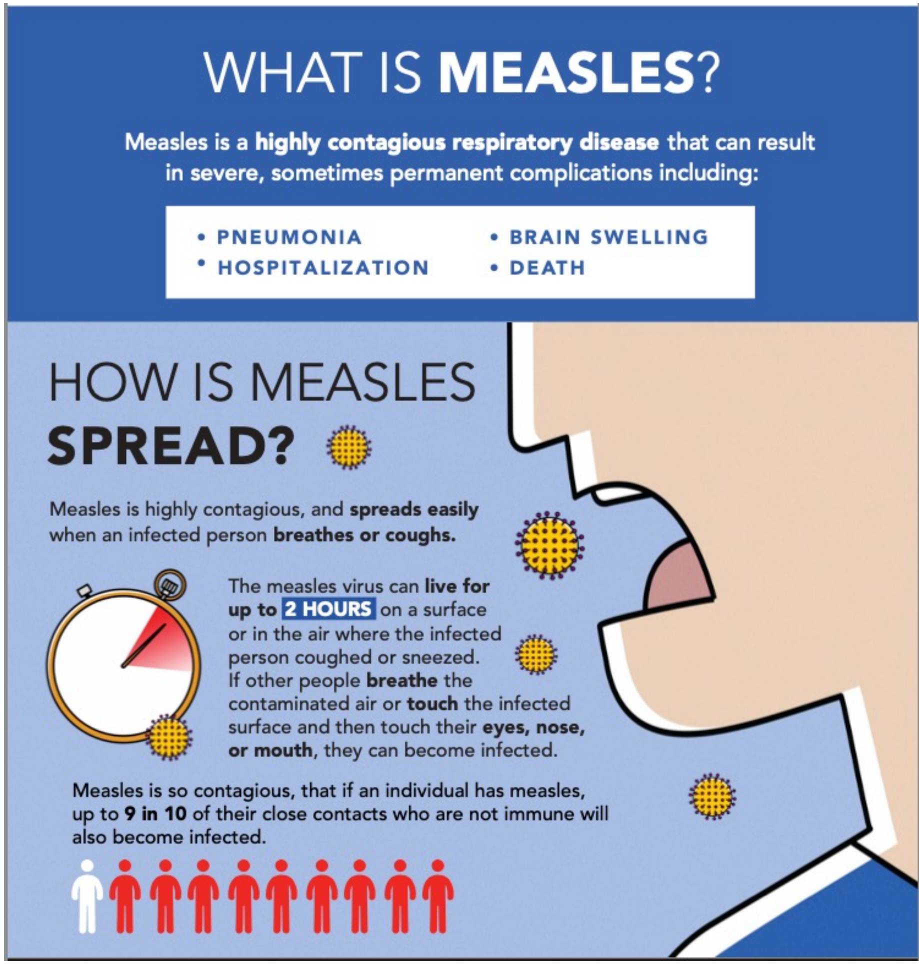
Other ways people get Zika

During pregnancy
A pregnant woman can pass Zika virus to her fetus during pregnancy. Zika infection during pregnancy can cause serious birth defects and is associated with other pregnancy problems.

Through sex
Zika virus can be passed through sex from a person who has Zika to his or her sex partners.

Through blood transfusion
Zika virus may be spread through blood transfusion.

264550-A



WHAT IS MEASLES?

Measles is a **highly contagious respiratory disease** that can result in severe, sometimes permanent complications including:

- PNEUMONIA
- HOSPITALIZATION
- BRAIN SWELLING
- DEATH

HOW IS MEASLES SPREAD?

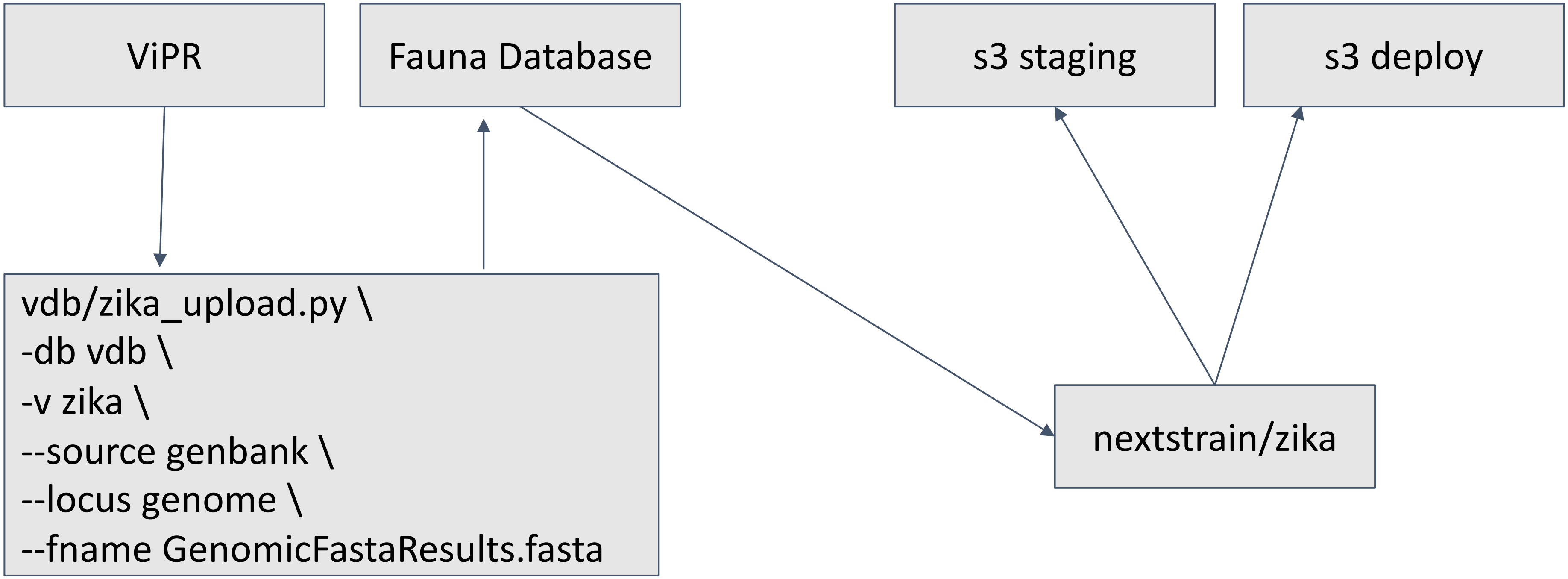
Measles is highly contagious, and **spreads easily** when an infected person **breathes or coughs.**

The measles virus can **live for up to 2 HOURS** on a surface or in the air where the infected person coughed or sneezed. If other people **breathe** the contaminated air or **touch** the infected surface and then touch their **eyes, nose, or mouth**, they can become infected.

Measles is so contagious, that if an individual has measles, up to **9 in 10** of their close contacts who are not immune will also become infected.

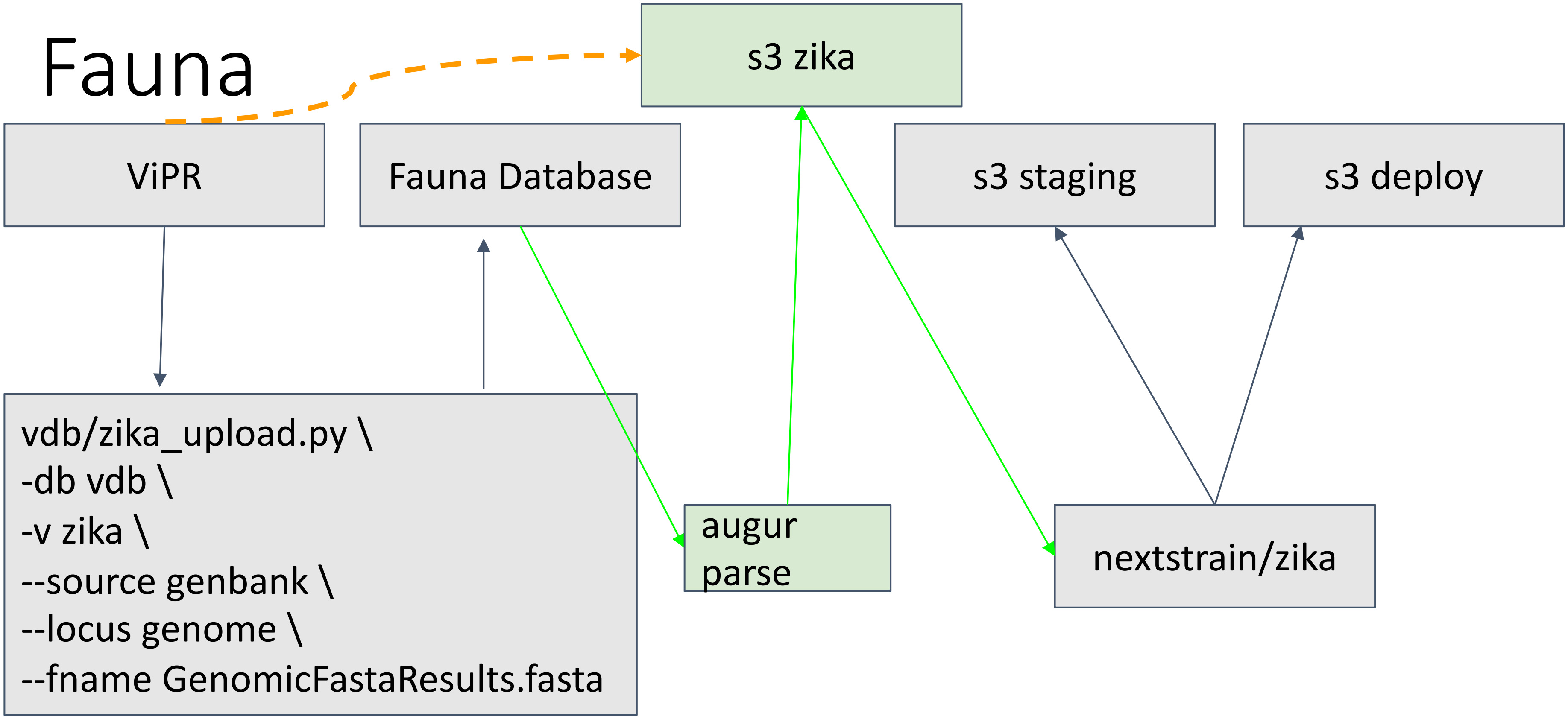
Viral Genome

Fauna



```
vdb/zika_upload.py \
-db vdb \
-v zika \
--source genbank \
--locus genome \
--fname GenomicFastaResults.fasta
```

Fauna



s3 zika

ViPR

Fauna Database

s3 staging

s3 deploy

```
vdb/zika_upload.py \
-db vdb \
-v zika \
--source genbank \
--locus genome \
--fname GenomicFastaResults.fasta
```

augur parse

nextstrain/zika

Merging Metadata

Genbank	Strain	Date	Country	Notes
				REF_TX98

cache

Accession	Strain	Country	Date	Clade

NCBI, ViPR, fauna, other…

Harmonized cache

Genbank	Strain	Date	Country	Notes	Clade

← Importance

Merging Metadata

cache

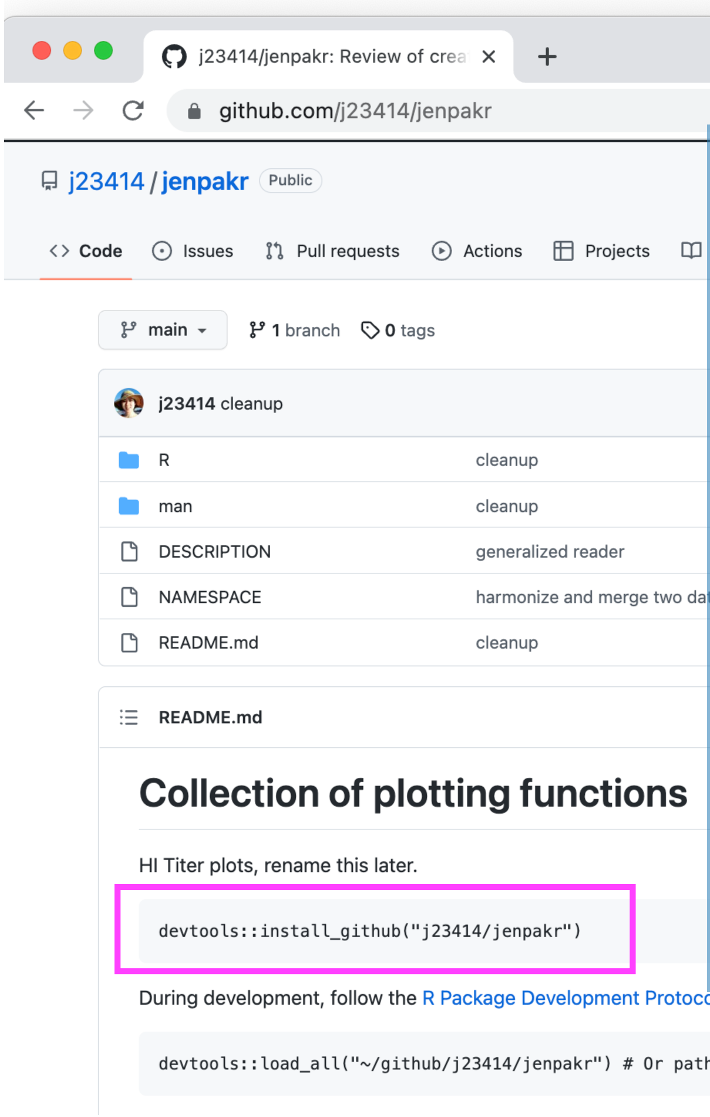
Genbank	Strain	Date	Country	Notes
MG1111	a/wa/17		USA	

NCBI, ViPR, fauna, other…

Accession	Strain	Country	Date	Clade
MG1111	a/wa/17	IND		

Harmonized cache

Genbank	Strain	Date	Country	Notes	Clade
MG1111	a/wa/17		USA,IND		



j23414 cleanup

R	cleanup
man	cleanup
DESCRIPTION	generalized reader
NAMESPACE	harmonize and merge two data sou
README.md	cleanup

README.md

Collection of plotting functions

HI Titer plots, rename this later.

```
devtools::install_github("j23414/jenpakr")
```

During development, follow the R Package Development Protocol.

```
devtools::load_all("~/github/j23414/jenpakr") # Or path to your c
```

Merging Datasets

```
library(jenpakr)
library(readxl)
library(tidyverse)
library(magrittr)

# ================= Read in Data to merge

# Normal excel cache
cache_data <- readxl::read_excel("path/to/cache.xlsx")

# Possible special cases
vipr_data <- jenpakr::read_delim_file("path/to/vipr.tsv", type="vipr")
ncbi_data <- jenpakr::read_delim_file("path/to/ncbi.tsv")
fauna_data <- jenpakr::read_delim_file("path/to/fauna.tsv", type="fauna")
clades_data <- jenpakr::read_delim_file("path/to/clades.csv", delim=",")

# ================== Merge files
merged_df <- cache_data %>%
  jenpakr::merge_two(., vipr_data) %>%
  jenpakr::merge_two(., ncbi_data) %>%
  jenpakr::merge_two(., fauna_data) %>%
  jenpakr::merge_two(., clades_data)

# ================== Save new cache
writexl::write_xlsx(merged_df, "new_cache.xlsx")
```

Thank you!