



# Rabies virus Long-Read Sequencing & Automated Bioinformatics Workflow

**Austin Rueda, MS**  
APHL-CDC Microbiology Fellow  
Arizona State Public Health Laboratory



# Agenda

## 1. Introduction to rabies virus

- Impact
- Variants
- Phylogenetics

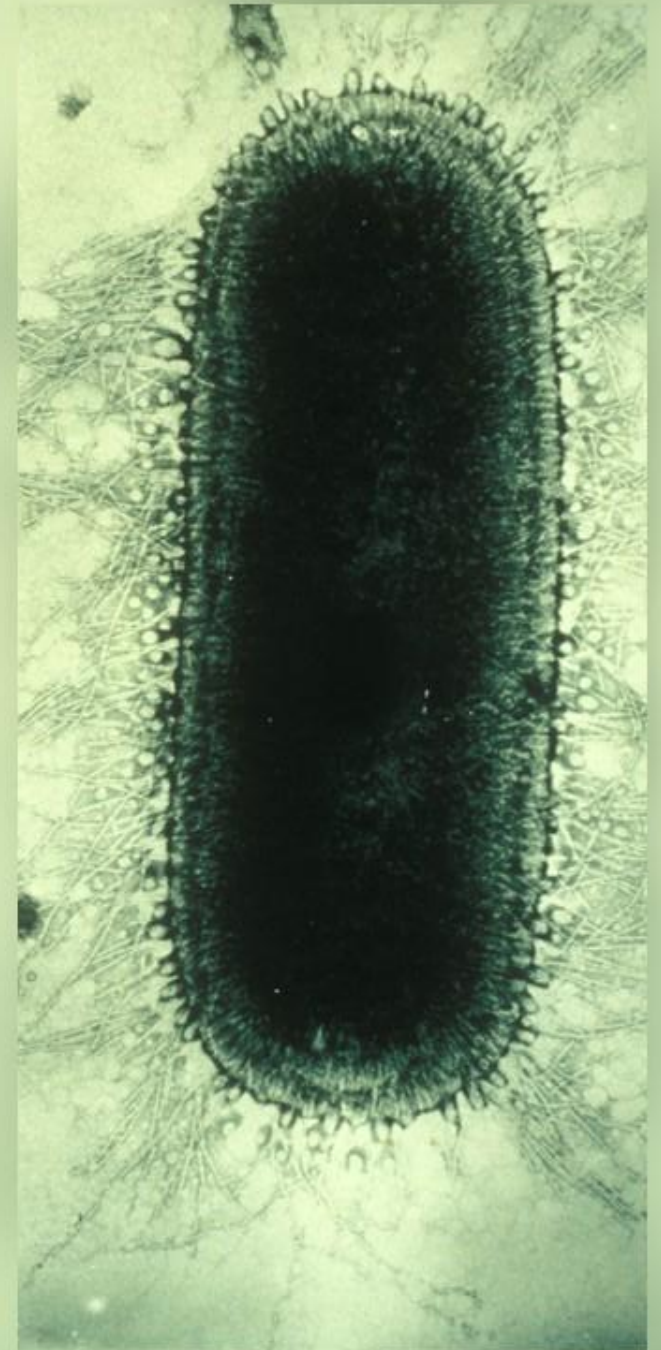
## 2. Rabies Surveillance

- Antigenic Methods
- Targeted Amplicon NGS
- Metagenomic WGS
- NGS Analysis Workflow

## 3. Case Study

- Reanalyzing AZ Rabies Samples

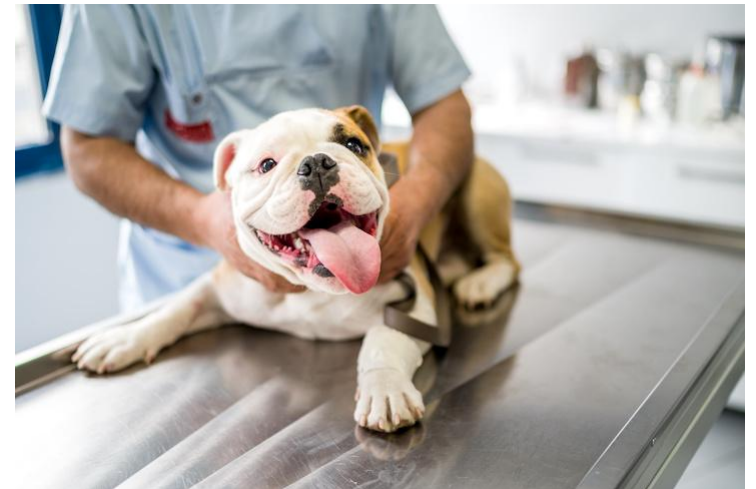
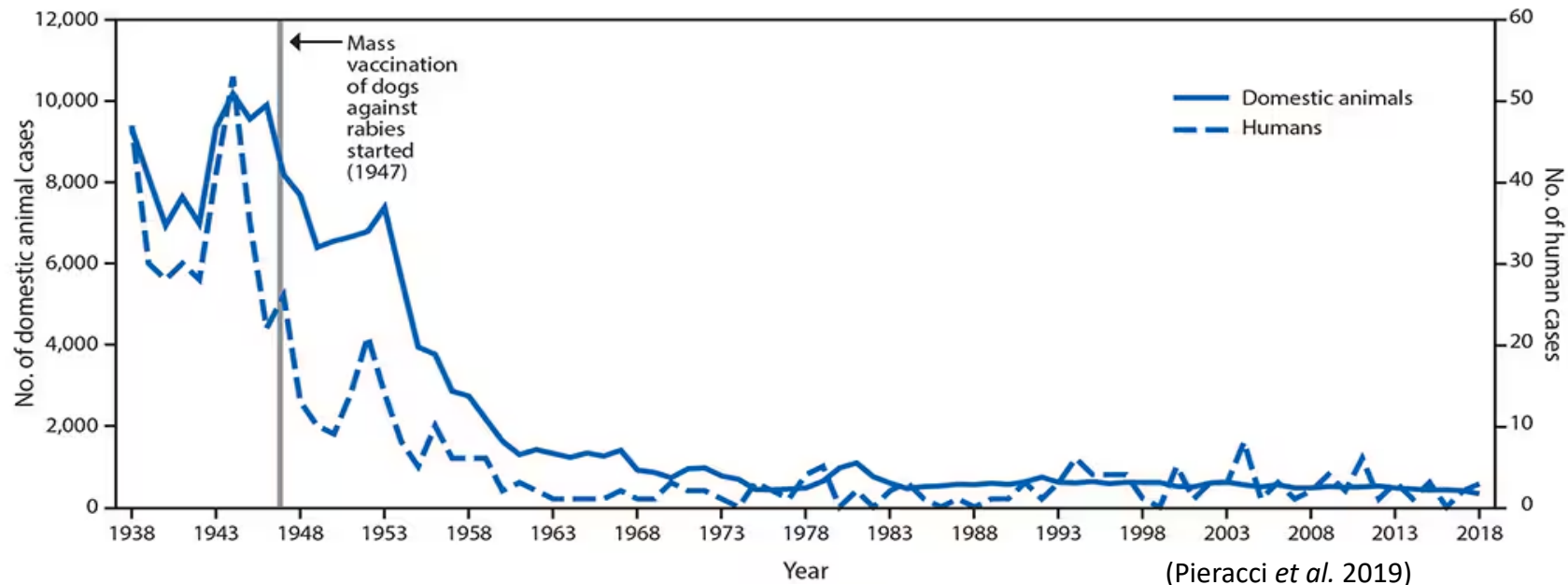
*Electron microscope image of rabies virus.  
(Norden, 1985)*



# Rabies Disease – Global Impacts

- Highest case fatality rate of any known disease (~100%)
- Global death rate: >59,000 / year
- 98% caused by canine rabies virus variant (CRVV)

*Mass vaccination of Dogs led to significant reductions in human rabies infections*





# Rabies Disease – US Impacts



*Myotis spp. (Bat Conservation Intl.)*



*Arizona Gray Fox (US FWS)*

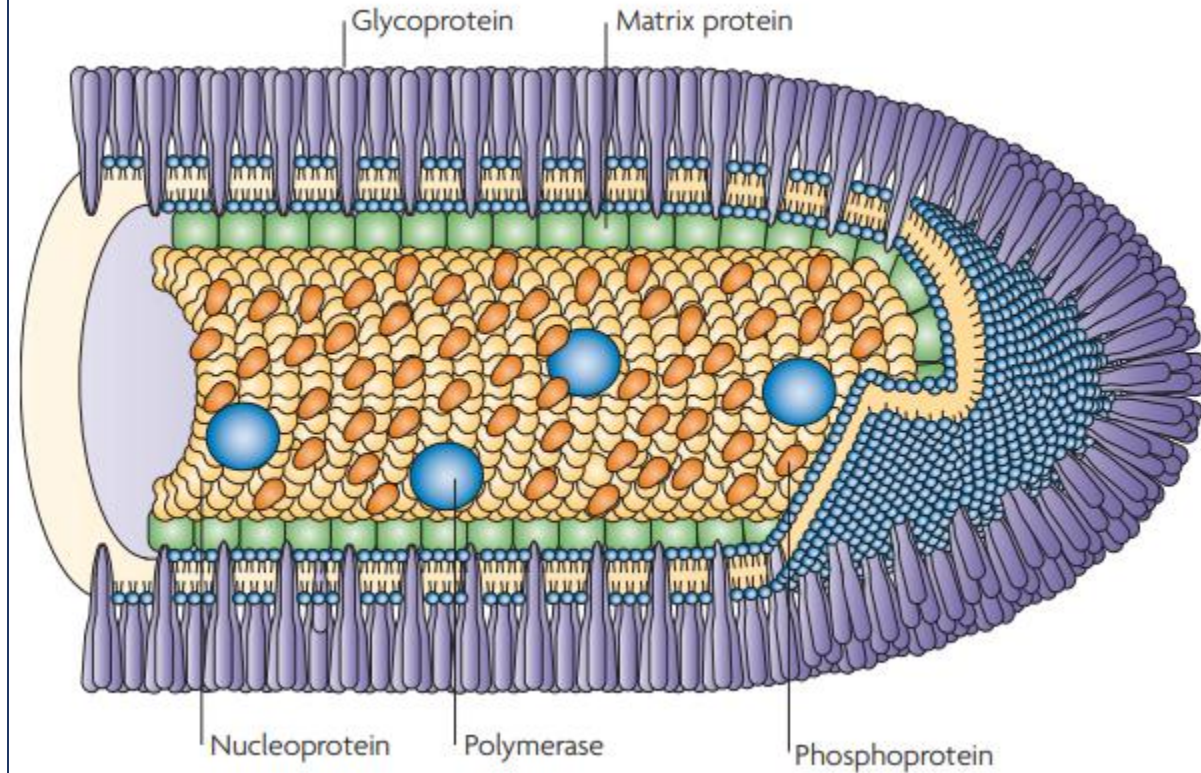
- <10 human cases / year
- 70% caused by bat bites
- 4,000 annual animal cases
  - Mesocarnivores
  - Bats
  - Skunks



*Striped Skunk (Vermont FWS)*

# Rabies Virus (RABV)

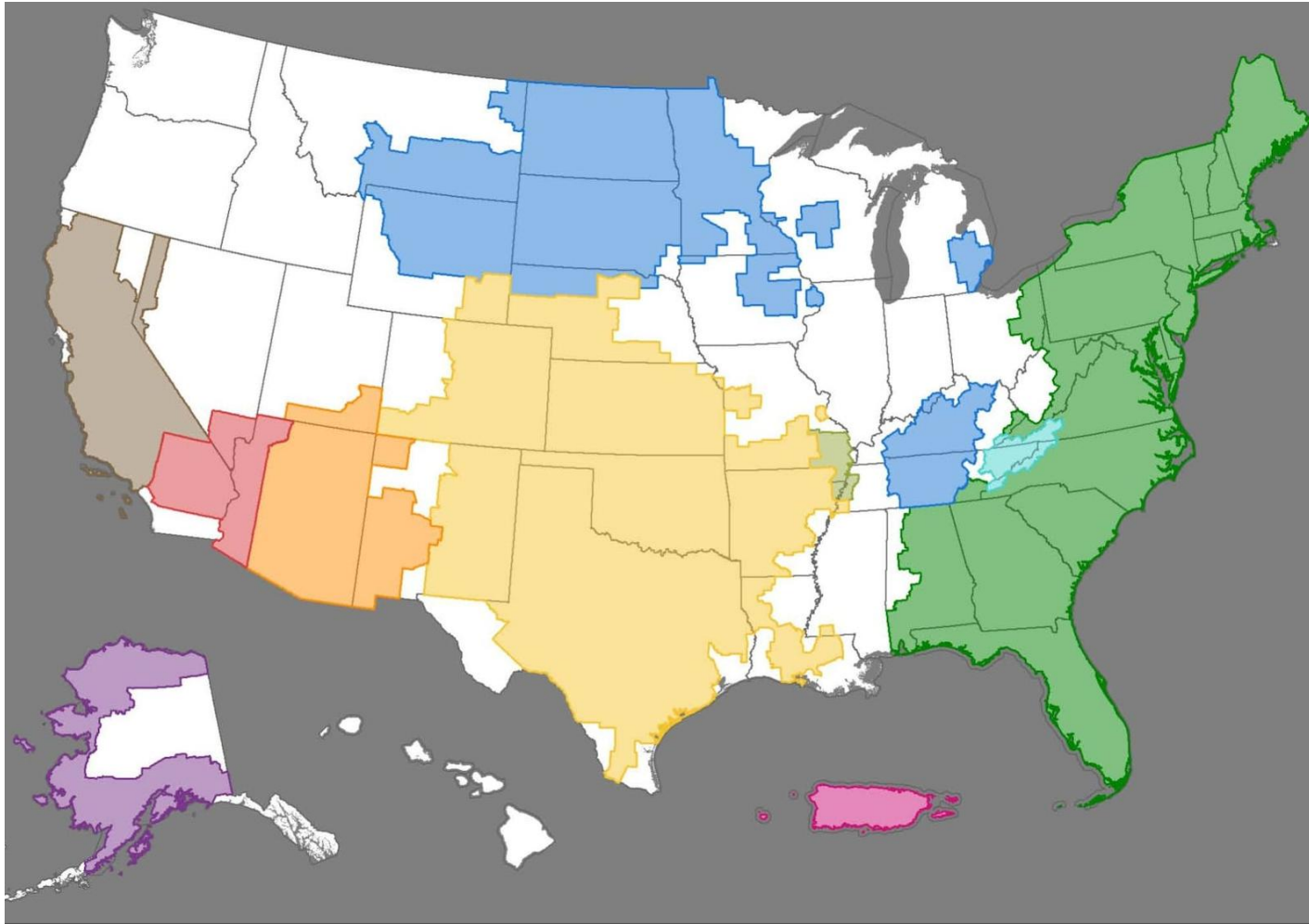
- Nonsegmented, Negative-stranded RNA Virus
- ~12knt – 5 Genes
  - I. Nucleoprotein (N)
  - II. Phosphoprotein (P)
  - III. RNA polymerase (L)
  - IV. Matrix Protein (M)
  - V. Glycoprotein (G)



*Rabies virion*  
(Nat Rev Microbiol., 2010)



# RABV Variants - US



Rabies Virus Variant



Rabies variant territories (CDC, 2024)

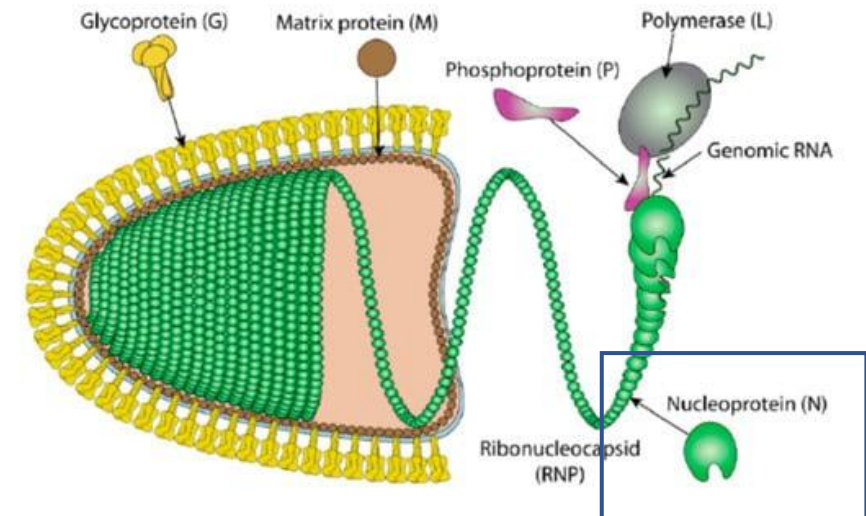
## Canine Variant

- I. Arctic Fox
- II. California Skunk
- III. Arizona Fox
- IV. South Central Skunk
- V. North Central Skunk
- VI. Eastern Raccoon
- VII. Mongoose

## Bat Variant

- I. Silver-haired Bat
- II. Hoary Bat
- III. Western Yellow Bat
- IV. Big Brown Bat
- V. Pallid Bat
- VI. Myotis
- VII. *Et al.*

A.



B.

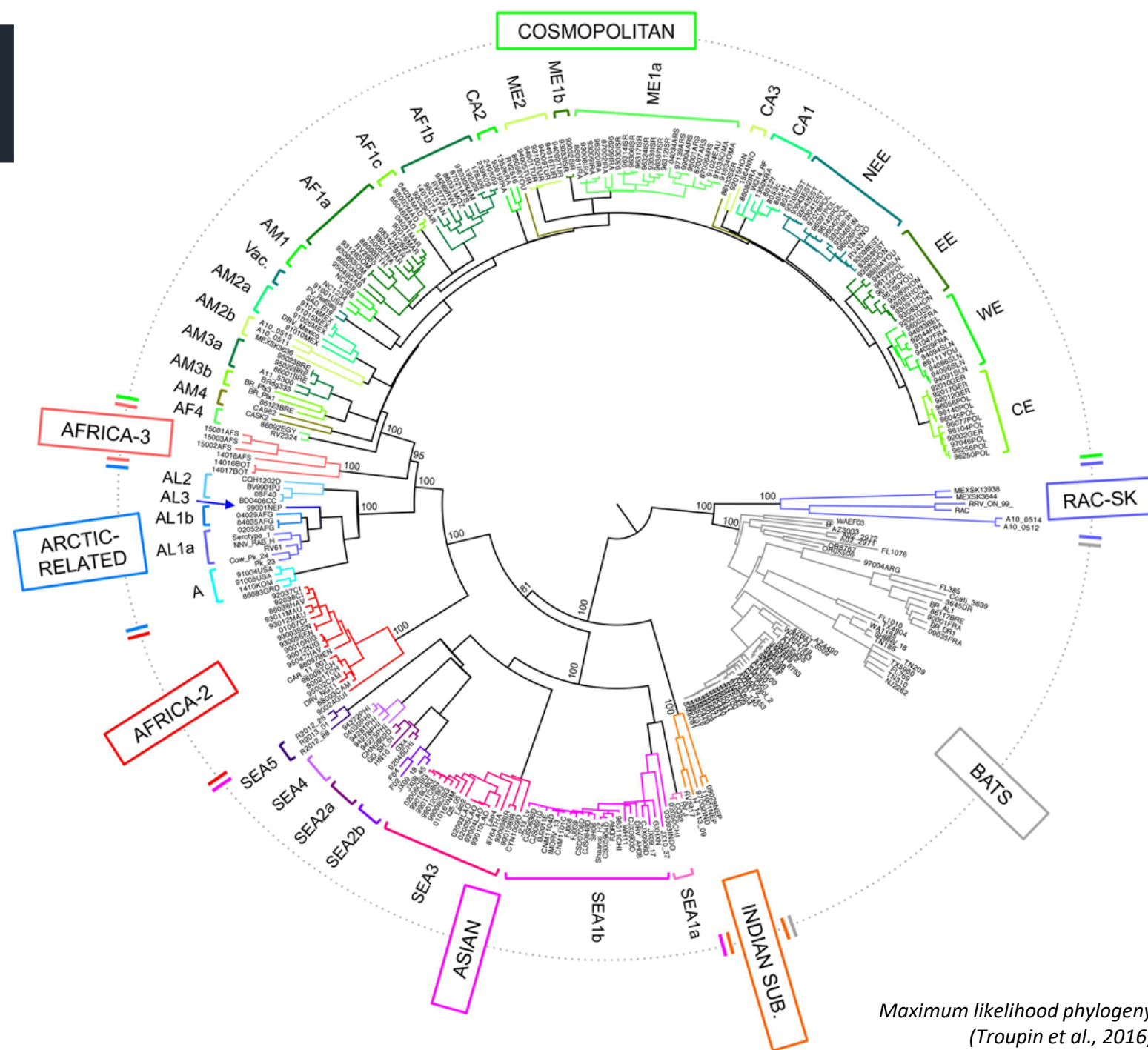
Rabies virion (Kiflu, 2024)

# RABV Phylogenetics

## Dog-Related

- I. Indian Sub
- II. Asian
- III. Africa-2
- IV. Africa-3
- V. Arctic-Related
- VI. Cosmopolitan

- 22 subclades – 3 US Circulating
- AM1 – Skunk & Woodchucks
- Am2(b) – US/ Mexico Wildlife
- AM4 – CA Skunks



Maximum likelihood phylogeny  
(Troupin et al., 2016)

# RABV Phylogenetics

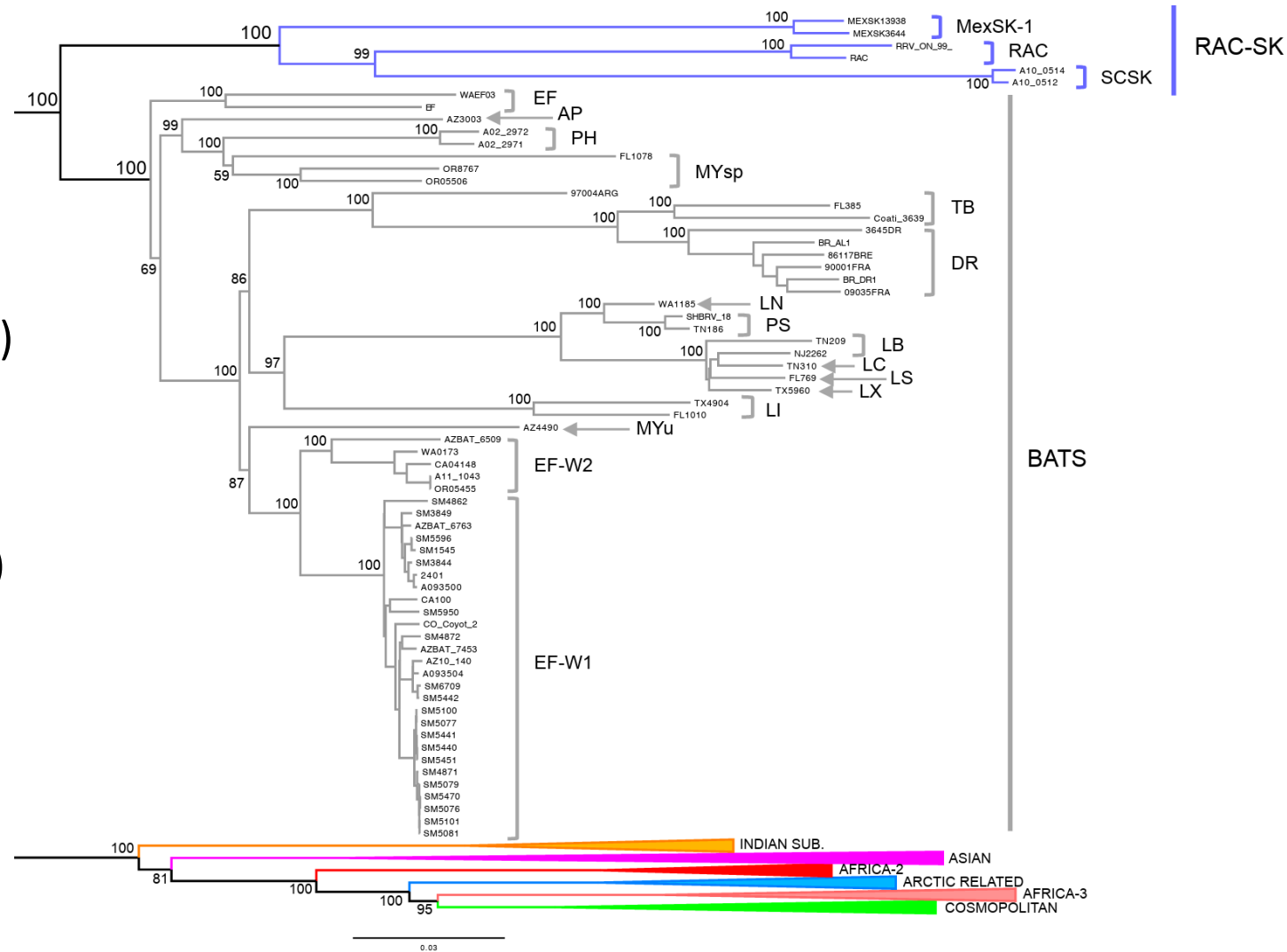
## Bat-Related

### I. Bats

- I. EF-W1/2 – *Eptesicus fuscus* (big brown bat)
- II. MYu – *Myotis yumanensis* (Yuma myotis)
- III. LX – *Lasiurus xanthinus* (Western yellow bat)
- IV. LC – *Lasiurus cinereus* (Hoary bat)
- V. LB – *Lasiurus blosevila* (Western red bat)
- VI. TB – *Tadarida brasiliensis* (Mexican freetail)

### II. RAC-SK

- I. SCSK – South Central Skunk
- II. RAC – Raccoon
- III. MexSK-1 – Mexican Skunk







# Rabies Surveillance – Current Methods

# DFA, rRT-PCR, & LFD

Direct Fluorescent  
Antibody (DFA) /

Direct Rapid  
Immunohistochemistry  
Test (DRIT)

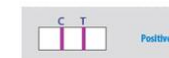
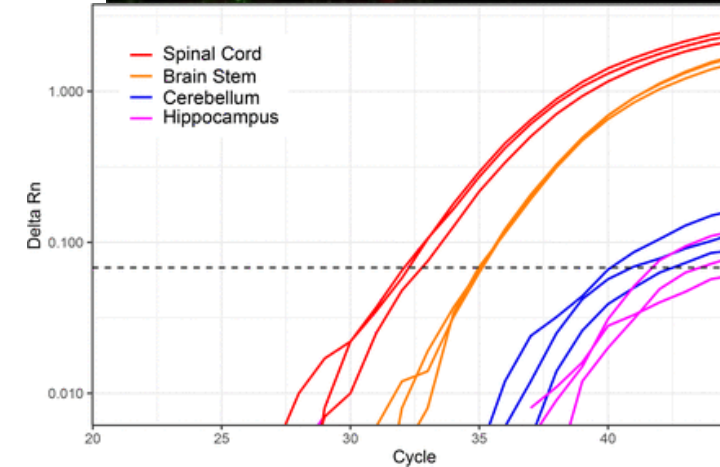
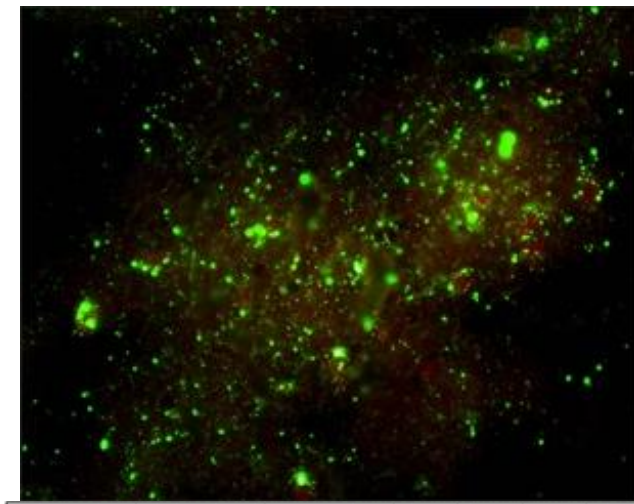
- Gold-Standard of rabies testing
- Subjective
- Sensitive
- Time-Consuming

LN34 rRT-PCR

- Highly Sensitive
- Less hands-on time than DFA
- Confirmatory testing only

Lateral Flow  
Device (LFD)

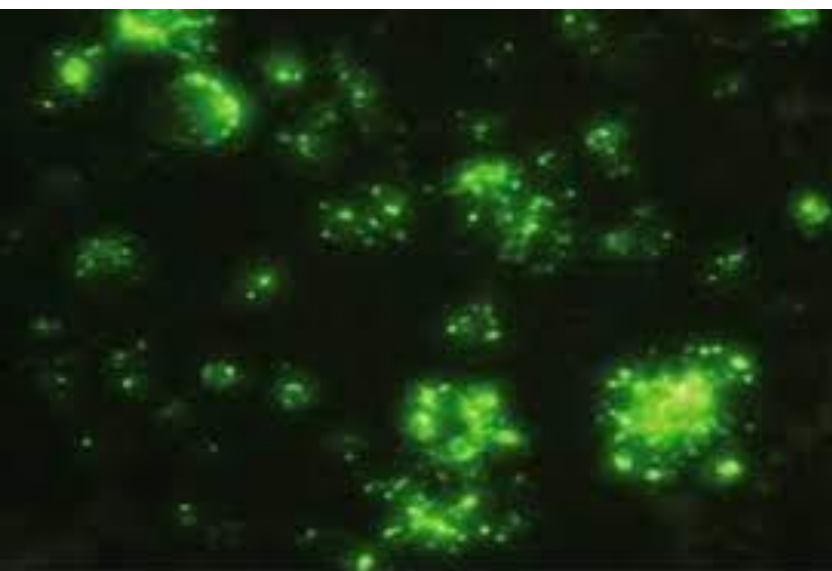
- Rapid
- Cheap
- Simple
- Low sensitivity
- Not widely available



# Antigenic Variant Typing

RABIES VARIANT TYPING CHART:

Cat. No.	Clone No.	Major Terrestrial Animal Reservoirs			Minor Terrestrial Animal Reservoirs				Insectivorous Bats		Other Reactivity
		Raccoon E US	Skunk NC US & CA	Skunk SC US	Dog/ Coyote Texas	Fox TX	Fox New England	Fox AZ	Bat Migratory	Bat Resident	Non-Rabies Lyssavirus
MAB8716	C4-62-15-2	pos.	pos.	pos.	pos.	pos.	pos.	pos.	pos.	pos.	NEG.
MAB8717	CR54-62-29-54	pos.	NEG.	varies	NEG.	NEG.	NEG.	NEG.	NEG.	NEG.	varies
MAB8718	C1-62-3-1	NEG.	pos.	NEG.	pos.	pos.	NEG.	pos.	NEG.	varies	varies
MAB8719	C2-62-8-2	pos.	pos.	NEG.	pos.	pos.	pos.	pos.	pos.	varies	varies
MAB8722	C10-62-52-2	pos.	pos.	pos.	pos.	pos.	pos.	NEG.	pos.	varies	varies
MAB8723	C15-62-97-3	pos.	pos.	pos.	pos.	pos.	NEG.	pos.	NEG.	varies	varies
MAB8724	C18-62-143-2	pos.	NEG.	pos.	NEG.	NEG.	NEG.	NEG.	NEG.	varies	varies



DFA positive for rabies virus  
(Tesfaye, 2017)

- Phenotypic, determined by Nucleoprotein
- 7 + monoclonal antibodies (MAB) tested against rabid brain tissue
- Results Limited
  - Arizona Gray Fox
  - South Central Skunk
  - Insectivorous Bat
  - Non-Endemic
- Sequencing send out to CDC required for further analysis



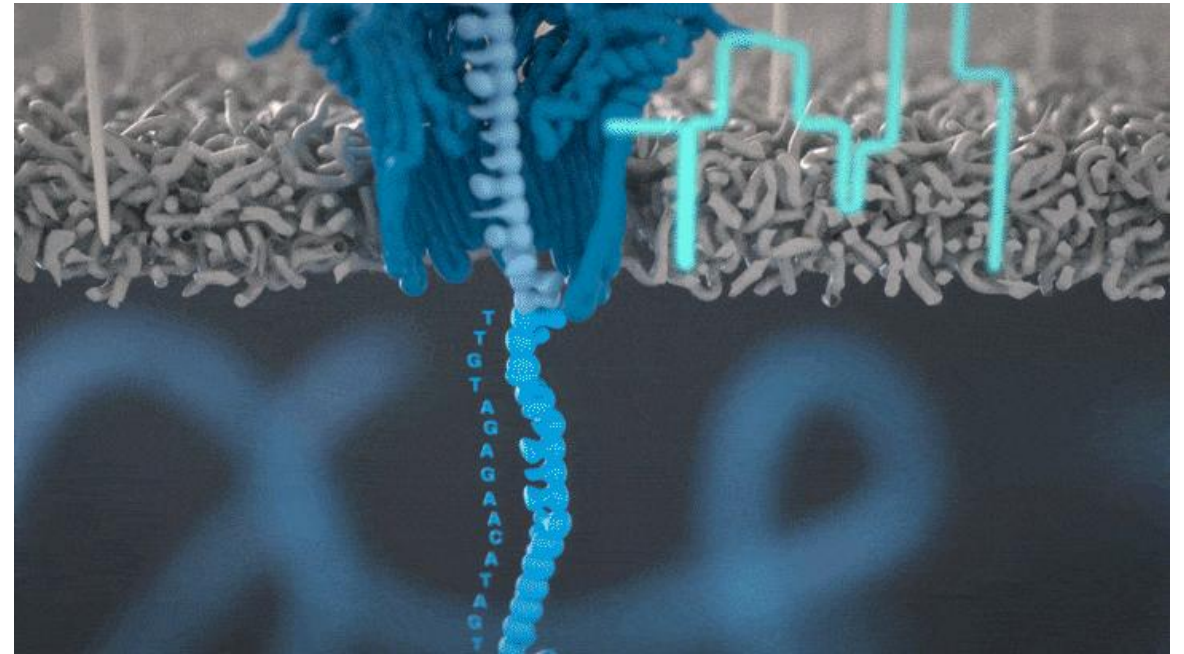
# Next-Generation Sequencing

Targeted Amplicon

Metagenomic WGS

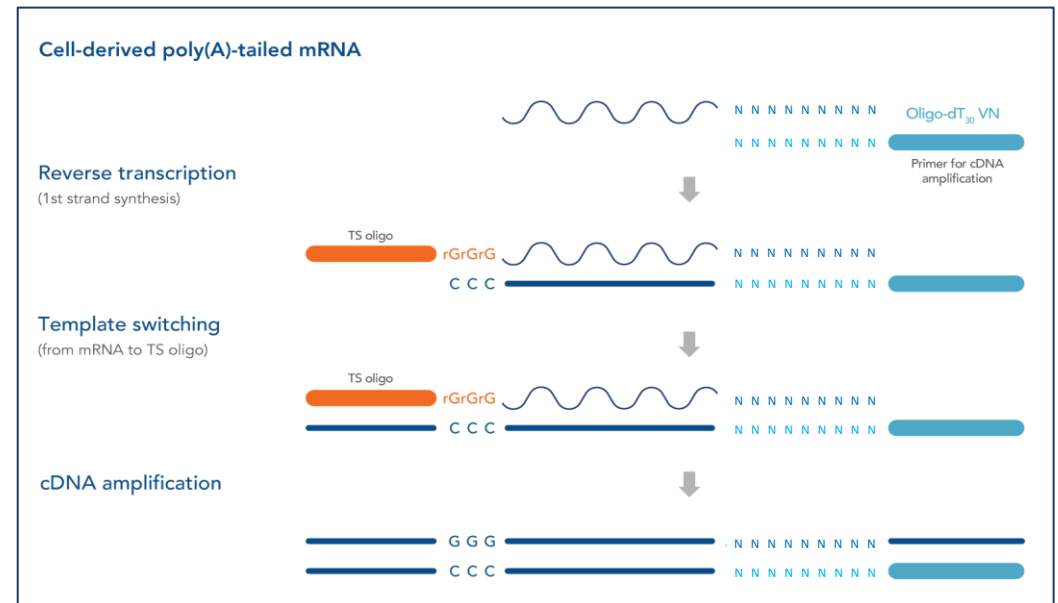
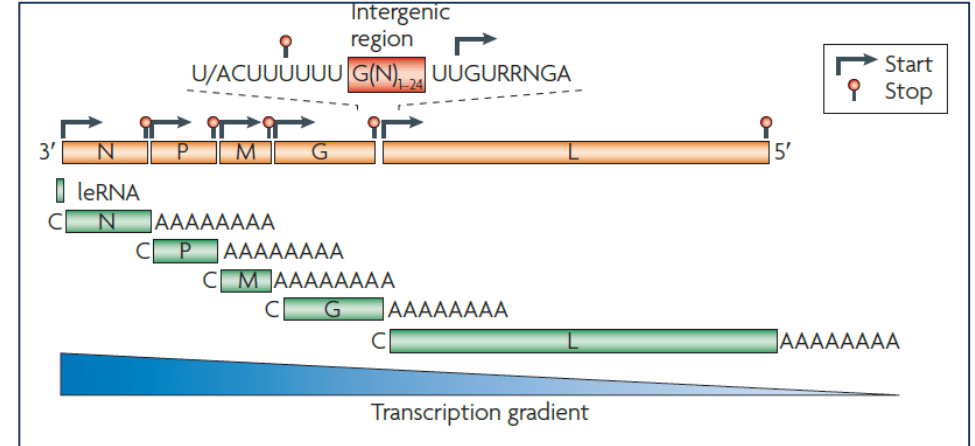
# Targeted Amplicon

- ONT Long-Read Sequencing
- Modified from CDC
- Targets:
  - N Gene (~1.5kb)
  - G Gene (~2kb)
- Single Primer Set
  - All US Canine & Bat Variants
  - Non US-Variants TBD
- Two separate rounds of PCR
  - barcoding acts as nested PCR primers



# Metagenomic WGS – Switching Mechanism At the 5' end of RNA Transcript (SMART-9N)

- ONT Long-Read Sequencing
- Variation of SMART method
  - dN first strand synthesis  
*Opposed to dT SMART method targeting polyadenylated transcripts*
- Metagenomic Approach
  - random priming (dN)
- Agnostic, Many Applications outside of RABV





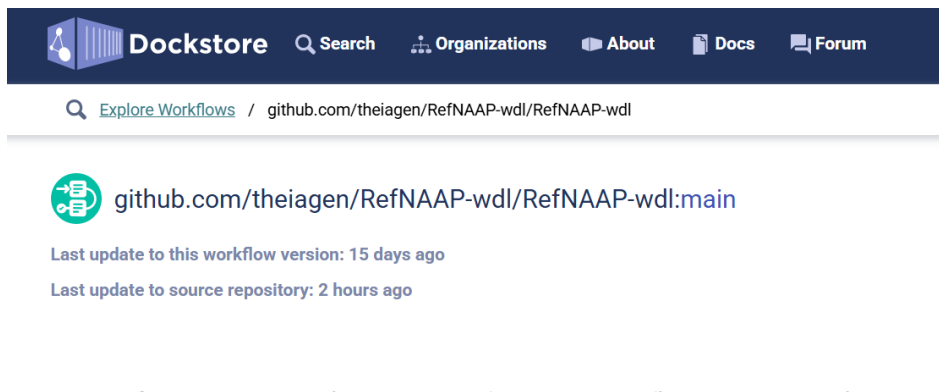
# Theiagen Genomics RefNAAP – WDL: Automated Bioinformatics Workflow

Developed in Collaboration with Theiagen Genomics

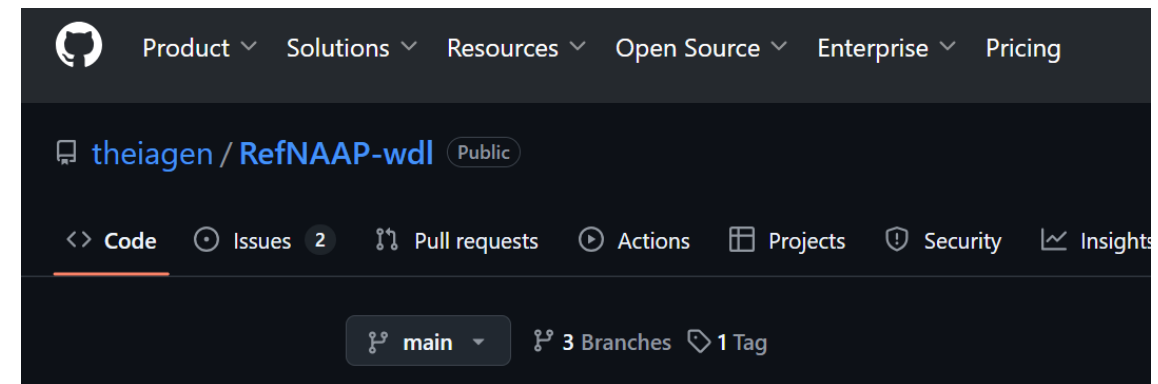
Fully automated Workflow hosted on Terra.bio

- QC w/ Report
- Alignment to Reference Sequence
- Closest NCBI Full Genome Accession
- Major & Minor Clades
- Coding Sequence Coverage

Available – *Currently in Development, Not Validated*



The screenshot shows the Dockstore website interface. The top navigation bar includes the Dockstore logo, a search bar, and links for Organizations, About, Docs, and Forum. Below the navigation bar, the breadcrumb trail reads "Explore Workflows / github.com/theiagen/RefNAAP-wdl/RefNAAP-wdl". The main content area displays the repository name "github.com/theiagen/RefNAAP-wdl/RefNAAP-wdl:main" with a green GitHub icon. Below the repository name, it states "Last update to this workflow version: 15 days ago" and "Last update to source repository: 2 hours ago".



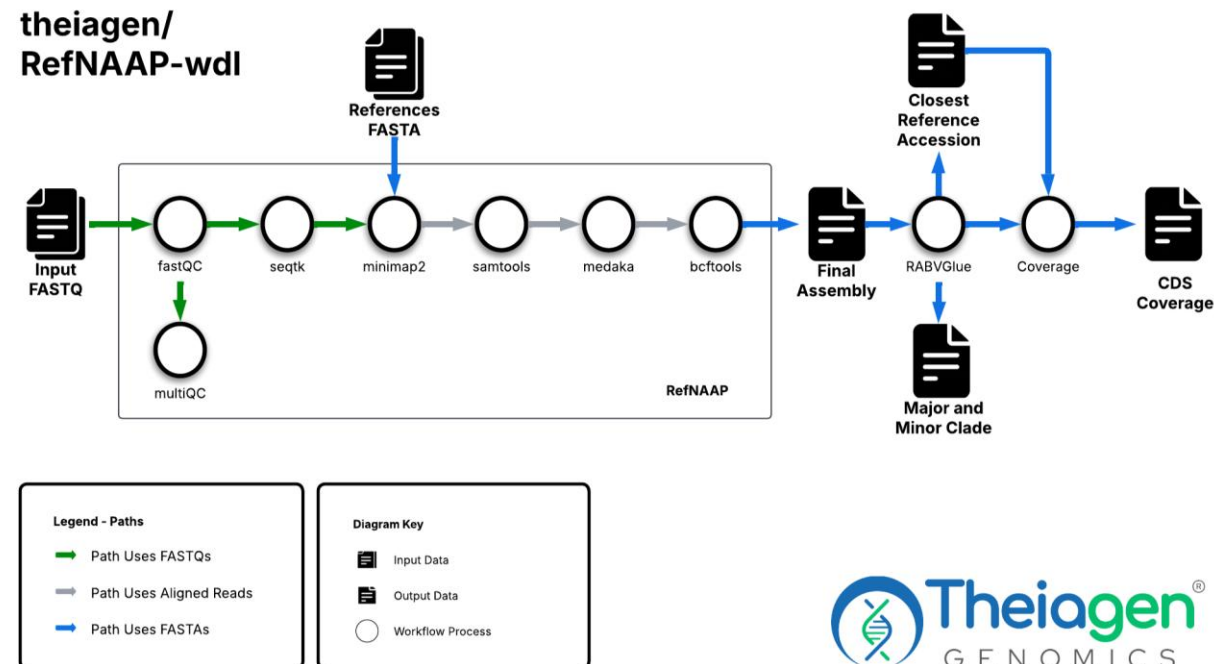
The screenshot shows the GitHub repository page for "theiagen / RefNAAP-wdl". The top navigation bar includes the GitHub logo and links for Product, Solutions, Resources, Open Source, Enterprise, and Pricing. The repository name "theiagen / RefNAAP-wdl" is displayed with a "Public" badge. Below the repository name, the navigation tabs include "Code", "Issues" (with a count of 2), "Pull requests", "Actions", "Projects", "Security", and "Insights". At the bottom, it shows the "main" branch selected, with "3 Branches" and "1 Tag" indicated.

## Reference-based Nanopore Amplicon Analysis Pipeline (RefNAAP) – *Modified from Yao & Gigante, CDC*

- QC Reads
  - FastQC & MultiQC to generate Report
- Trim & Filter
  - 25b off left & right end of reads
  - Filter reads <50bp
- Generates assembly
  - Reference based gap fixing
  - Medaka Assembly
  - Aligns to 1 of 14 Reference Seqs

## RABV-GLUE


- RABV centric bioinformatics resource & genotyping tool
  - Developed by University of Glasgow - BETA
- Blasts against NCBI RABV isolates
- Provides Major / Minor Clades & closest full genome reference sequence



# Automated Variant Typing Pipeline - Terra options

Parameters of Note:

- **Blast E Value** –  $1e^{-10}$  Default
- **Minimum Gene Coverage** – 75% Default
- **Minimum Percent Identity** – 75% Default
- **Minimum Depth of Coverage** – 5x Default
- **Medaka alignment Model**
- **Read Size**
- **Left and Right Read Trim**

Variable	Type	Input value
read1	File	<input type="text" value="this.fastq"/>  {...}
samplename	String	<input type="text" value="this.sample_id"/> {...}
blast_evalue	String	<input type="text" value="Optional"/> {...}
cpu	Int	<input type="text" value="Optional"/> {...}
disk_size	Int	<input type="text" value="Optional"/> {...}
docker	String	<input type="text" value="Optional"/> {...}
memory	Int	<input type="text" value="Optional"/> {...}
min_gene_coverage	Float	<input type="text" value="Optional"/> {...}
min_percent_identity	Float	<input type="text" value="Optional"/> {...}
cpu	Int	<input type="text" value="Optional"/> {...}
disk_size	Int	<input type="text" value="Optional"/> {...}
docker	String	<input type="text" value="Optional"/> {...}
memory	Int	<input type="text" value="Optional"/> {...}
cpu	Int	<input type="text" value="Optional"/> {...}
disk_size	Int	<input type="text" value="Optional"/> {...}
docker	String	<input type="text" value="Optional"/> {...}
memory	Int	<input type="text" value="Optional"/> {...}
min_coverage	Int	<input type="text" value="Optional"/> {...}
model	String	<input type="text" value="Optional"/> {...}
size	Int	<input type="text" value="Optional"/> {...}
trim_left	Int	<input type="text" value="Optional"/> {...}
trim_right	Int	<input type="text" value="Optional"/> {...}

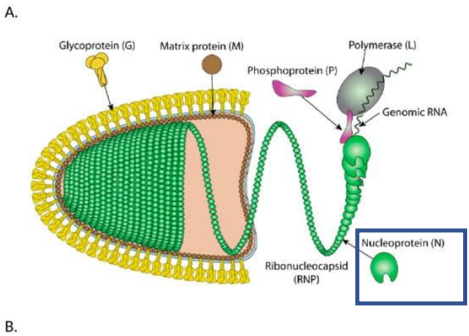


The background of the slide is a dark, high-resolution image of a star field. Numerous stars of varying brightness are visible, with a higher density in the upper right. Several stars are highlighted with a bright green color, creating a pattern that might represent a specific astronomical survey or data set. The overall tone is dark and scientific.

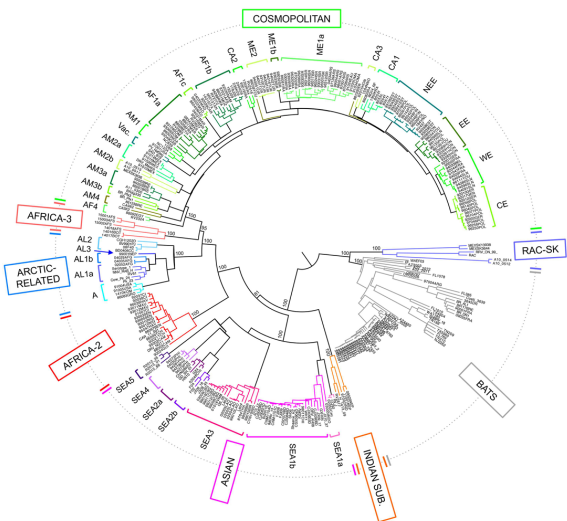
# Re-Analysis of ASPHL RABV Archive

# Correlating Antigenic Variant to Phylogenetic Clade

Variant = Result of Antigenic Testing



Major / Minor clade = Result of phylogenetic analysis via NGS

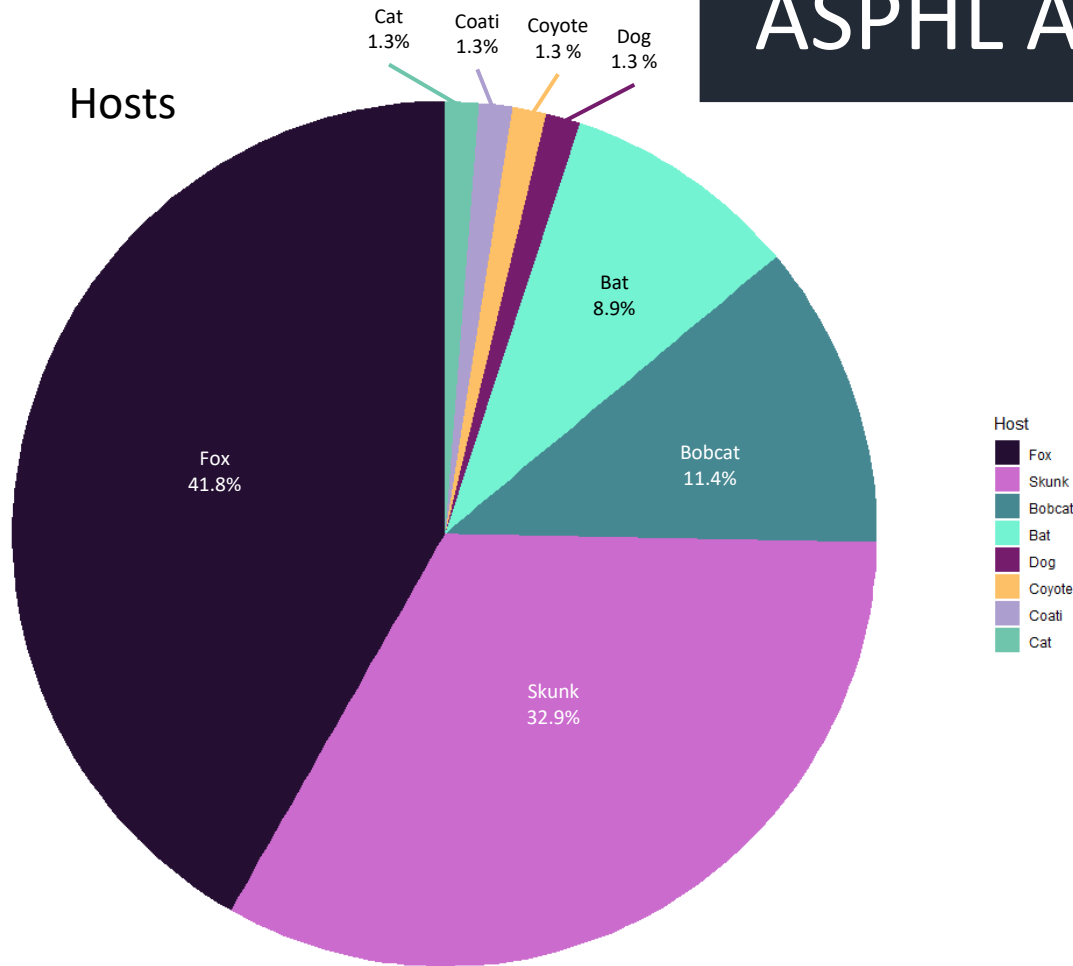


*Variant of the Closest Reference Sequence  
determined by RABV-GLUE*

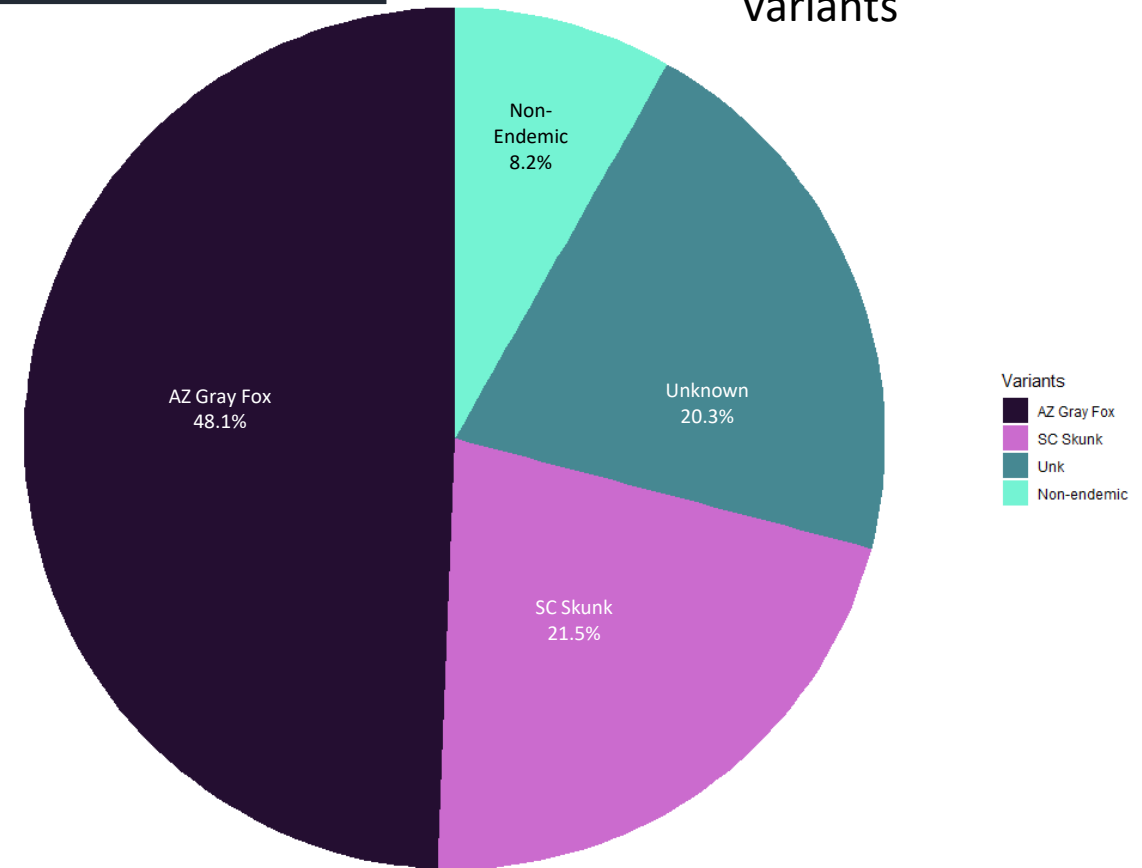
Variant	Clade
AZ Gray Fox	Cosmopolitan AM2b
SC Skunk	RAC-SK
Non-Endemic	Bats EF-W1
	Bats TB1
	Bats LX
	Bats MYu
	Cosmopolitan Vac2

# ASPHL Archived Tissue

Hosts



Variants



- All tissue samples previously confirmed positive for RABV *via DFA*
- Variant typed *via* MAB immunofluorescence
  - Bat specimens were not variant typed
- 80 samples sequenced... *And continuing*
  - Targeted Amplicon & Metagenomic WGS
- Analyzed Using Automated variant typing pipeline – Terra

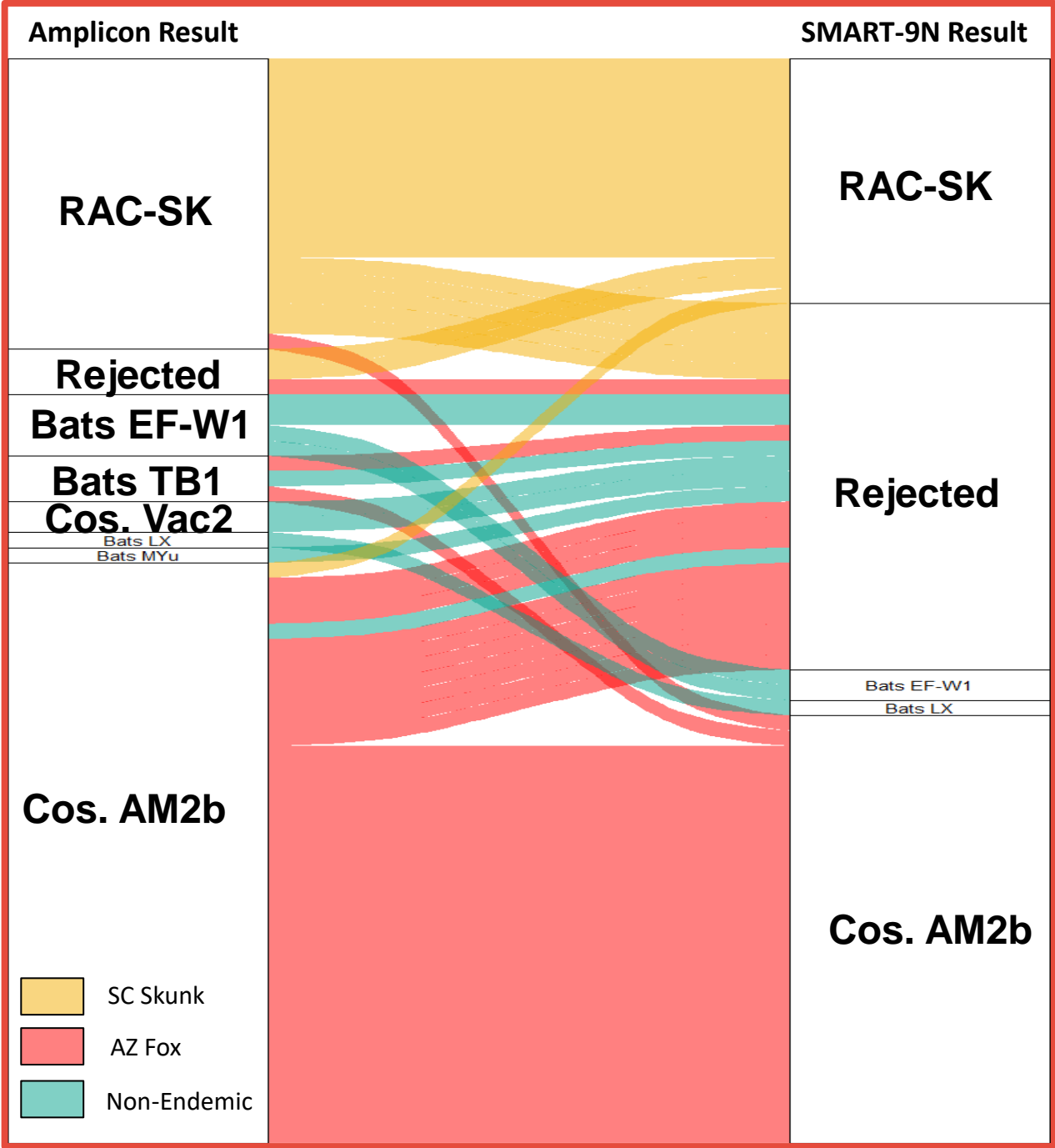


# Agreement to Antigenic Typing

Targeted Amplicon	
% Match	89% (67/75)
% Mismatch	6.5% (5/75)
% Rejected	4% (3/75)
Removed from analysis	(5) Not Variant Typed

SMART-9N WGS	
% Match	66% (44/68)
% Mismatch	0
% Rejected	35% (24/68)
Removed from analysis	(12) Not Variant typed / Tested

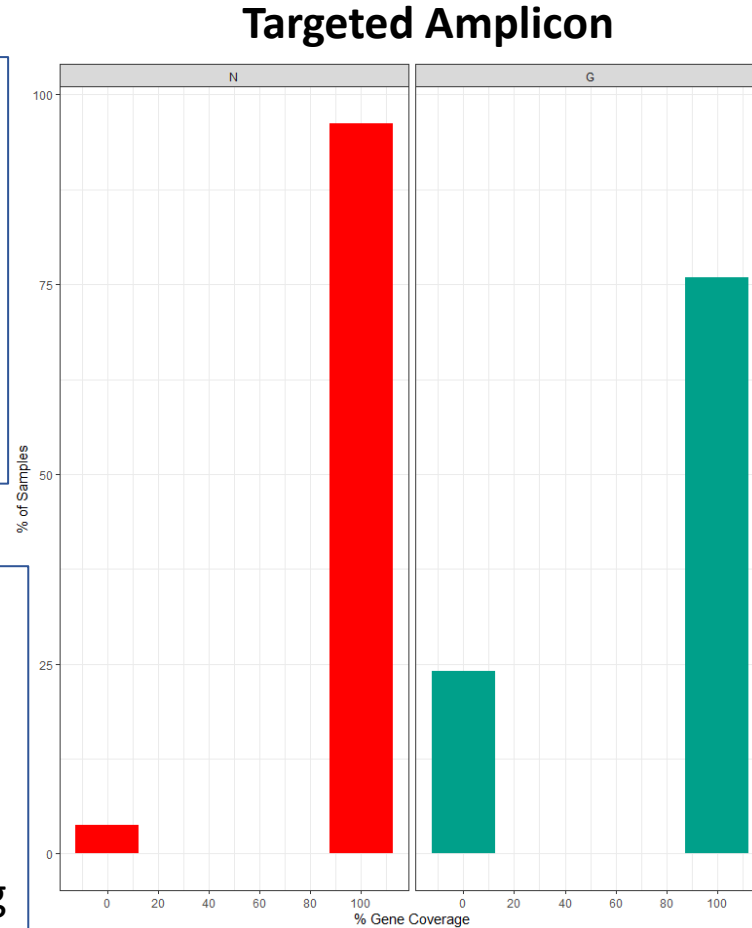
*\*Rejected: RefNAPP-WDL unable to process sample*



# Percent Gene Coverage

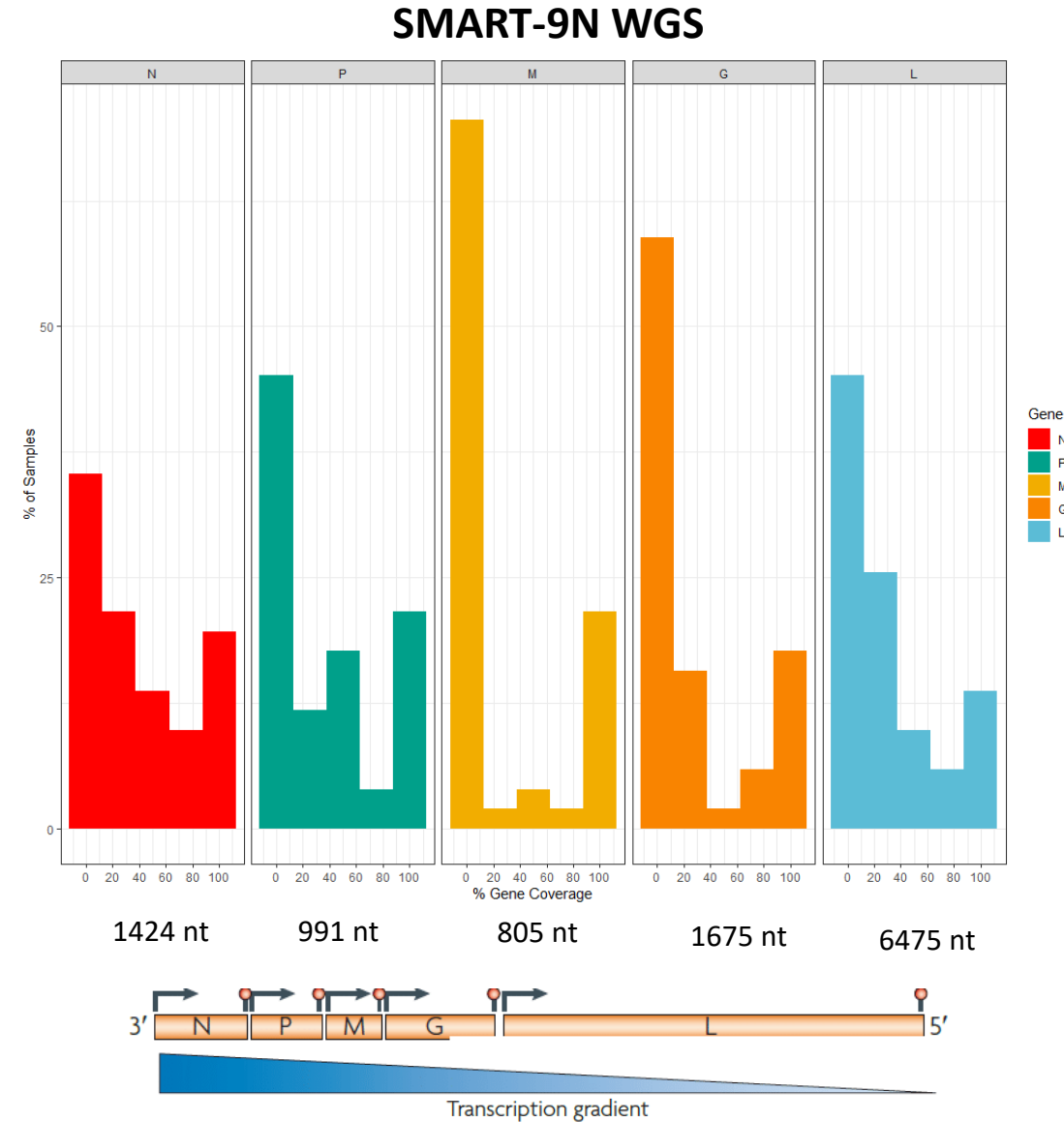
## Targeted Amplicon

- Complete coverage OR zero coverage
- Complete coverage of all N genes except failed sample



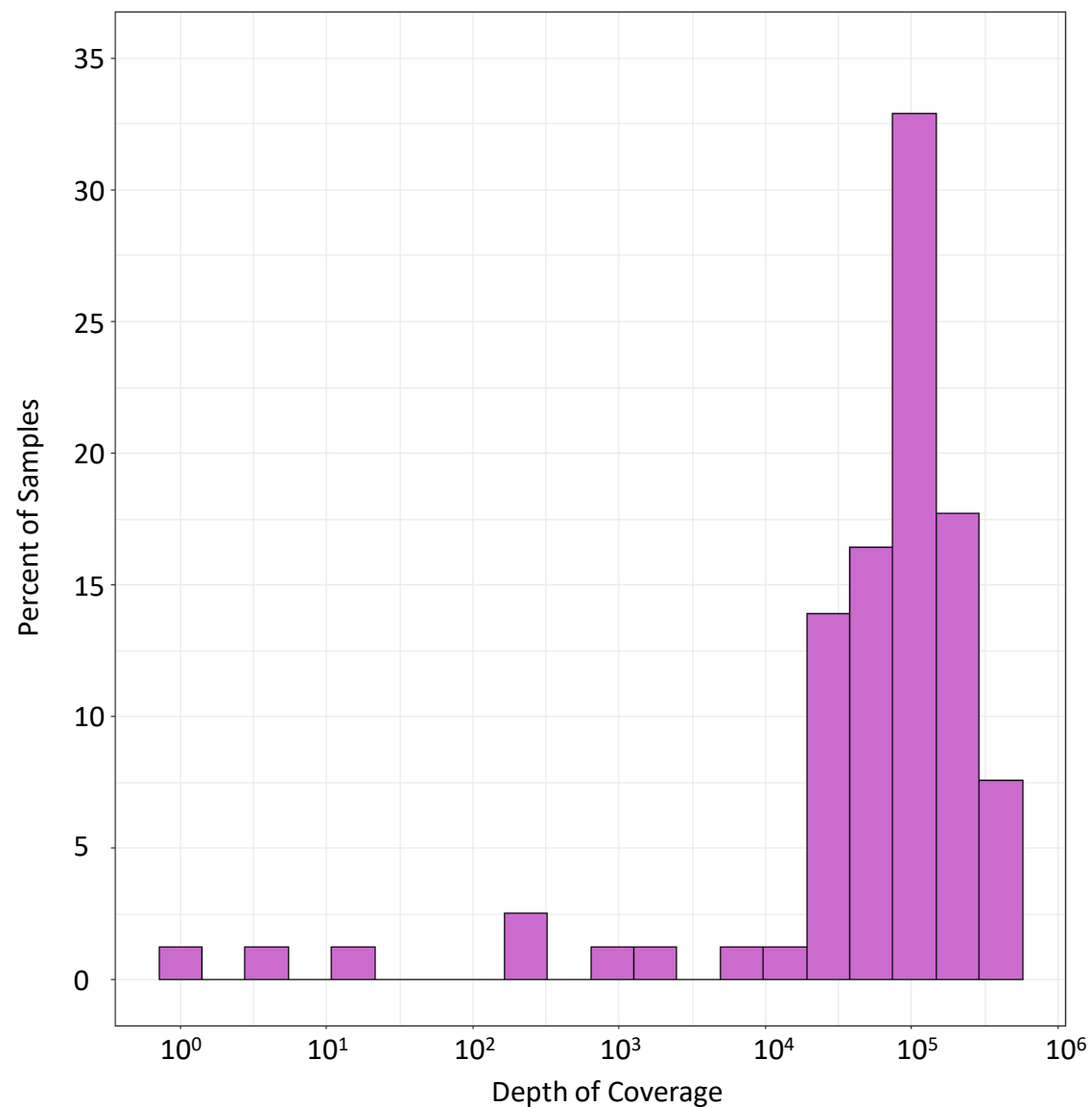
## SMART-9N

- Bi-modal, more distributed than Amp.
- Genes with zero coverage normally distributed along genome
- Gene length \* Transcription rate?

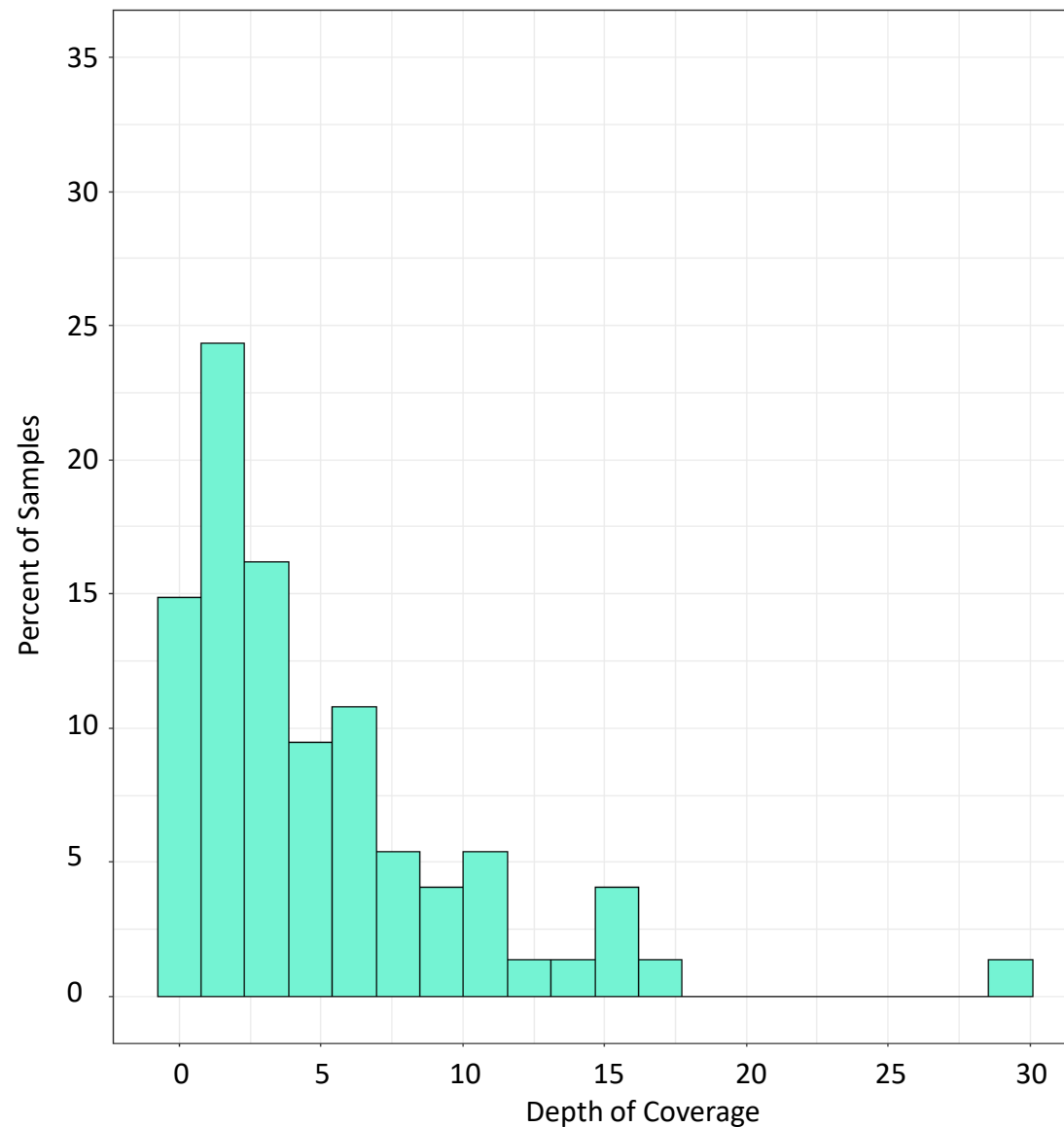


# Depth of Coverage

Targeted Amplicon



SMART-9N WGS



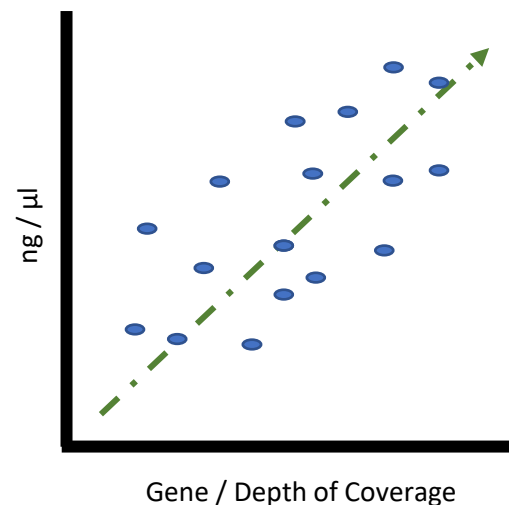
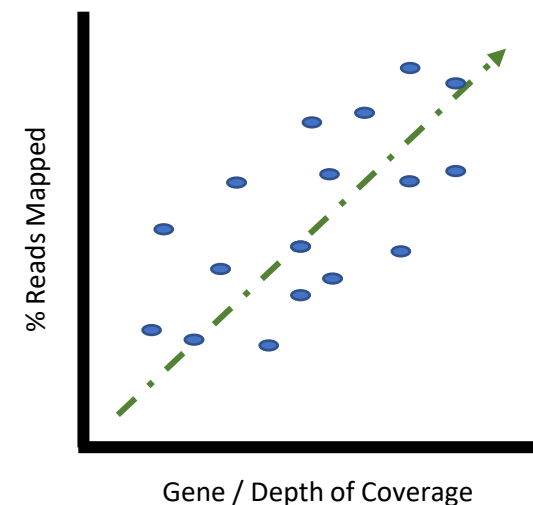
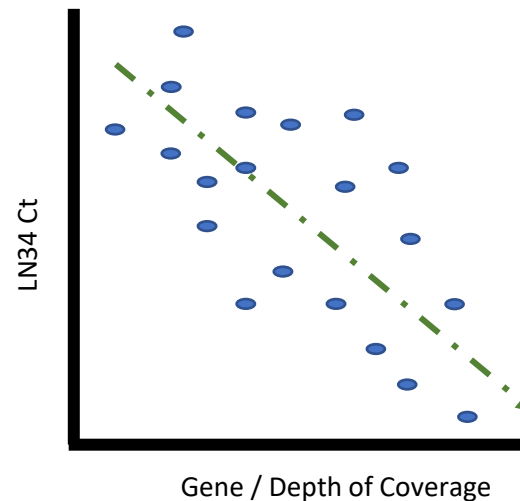


# Factors Influencing Sequencing Quality



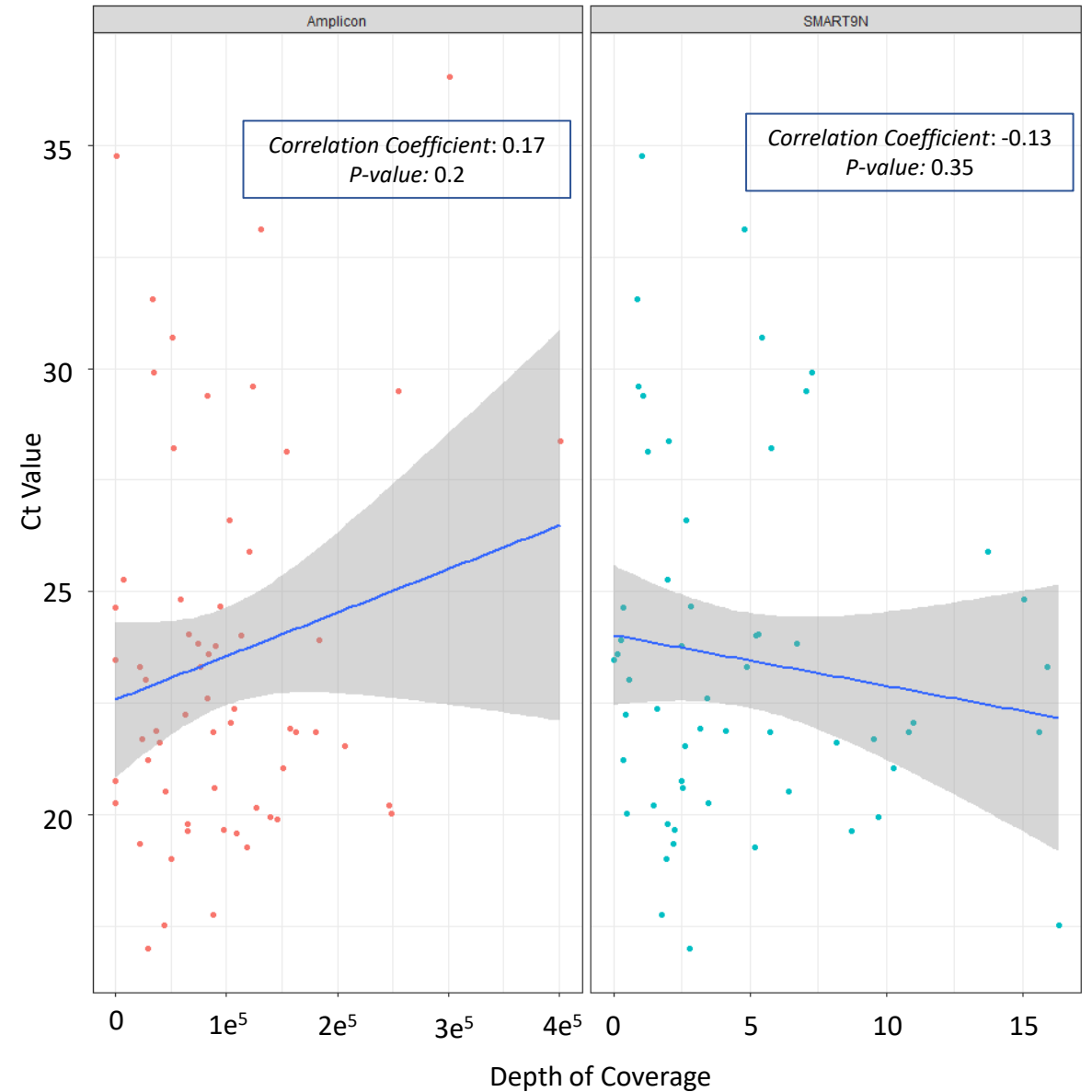
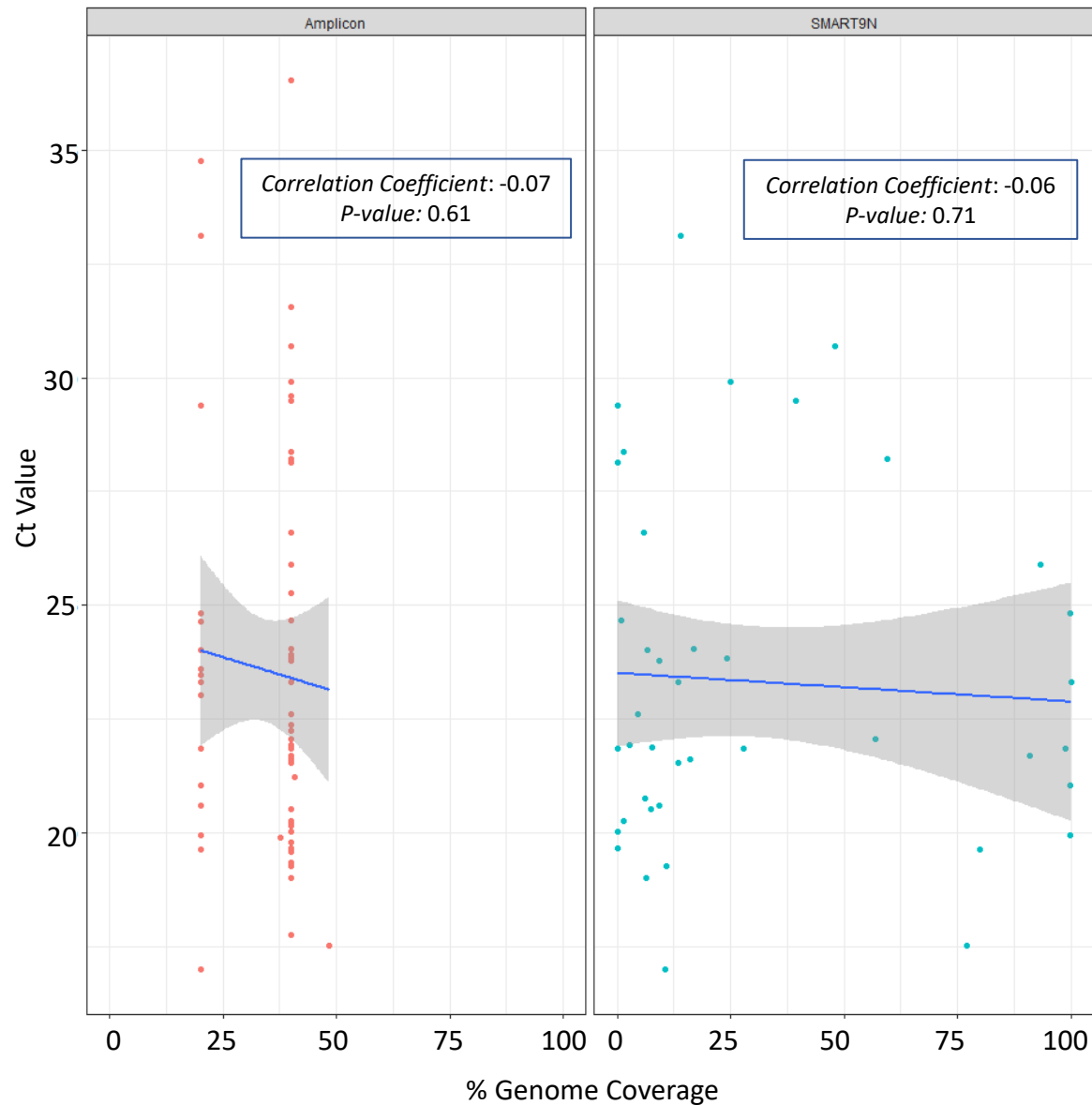
# Factors to Consider

1. Can **LN34 Ct** value predict gene coverage and/or depth of coverage?
  - Signal (RABV Reads) : Noise (Host Reads)
2. Does the proportion of **viral reads : host reads** affect gene coverage and / or depth of coverage?
  - Signal (RABV Reads) : Noise (Host Reads)
1. Can **PCR product mass** predict gene coverage and / or depth of coverage?
  - Used for library pooling normalization
  - *TBD*



# LN34 Ct

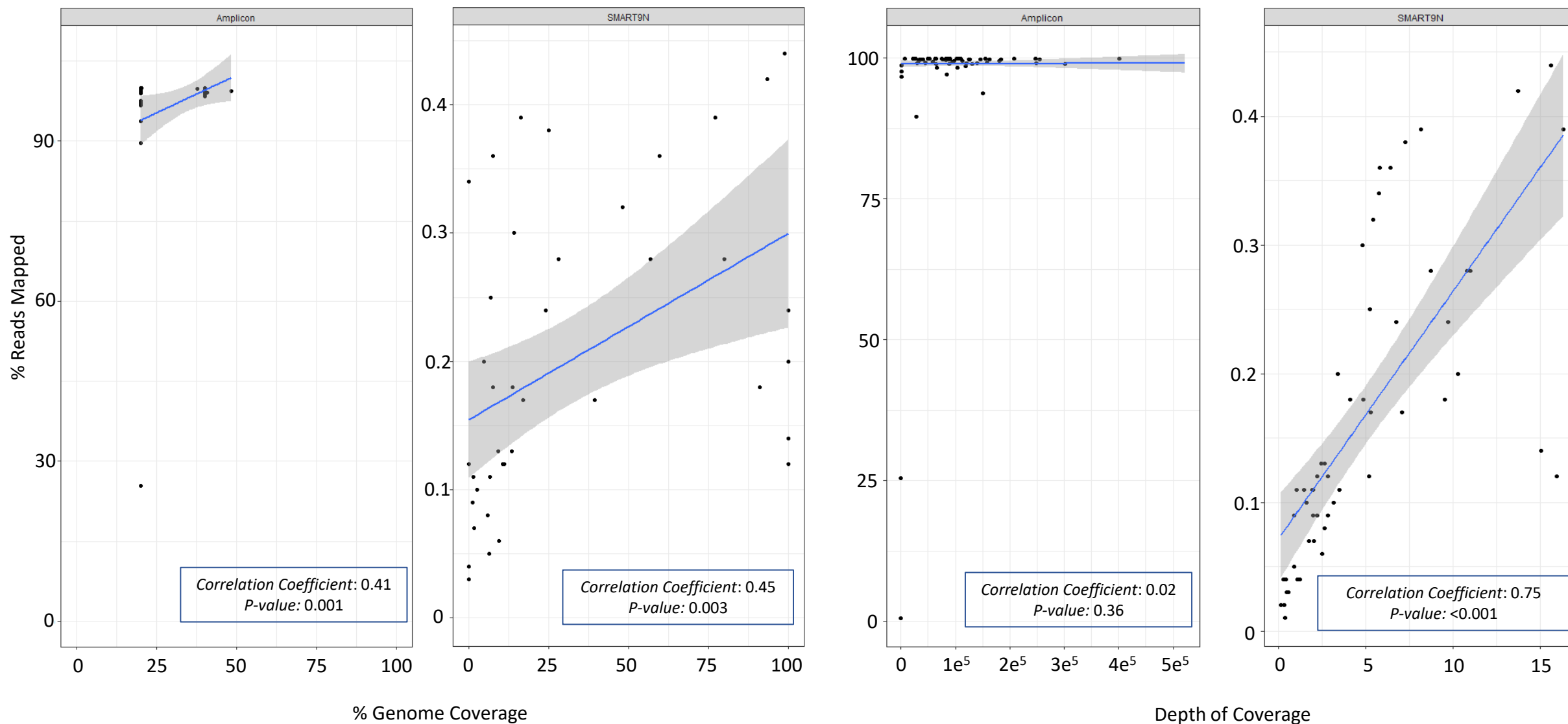
*LN34 Ct is not significantly correlated to Genome coverage /  
Depth of Coverage*



# % Reads Mapped

*% Reads Mapped to Ref Seq is significant predictor of genome coverage & SMRT-9N WGS Depth of Coverage*

*Viral reads : Host reads*





# Mismatched and Rejected Samples & Key Points



# Mismatched Samples

## Targeted Amplicon

	Count	Avg. Genome Coverage	Avg. Depth of Coverage
Match	91% (69/75)	36%	112,040 x
Mismatch	7% (5/75)	24%	92,039 x
Rejected	2% (2/75)	NA	2.5x
Removed from analysis	(4)	NA	NA

## SMART-9N WGS

	Count	Avg. Genome Coverage	Avg. Depth of Coverage
% Match	66% (44/68)	34%	7x
% Mismatch	0	NA	NA
% Rejected	35% (24/68)	NA	1x
Removed from analysis	(12)	NA	NA

- Genome Coverage: 36% - 24%
  - N gene is fully sequenced in all cases
- Massive Depth of Coverage in both cases
- Other Factors:
  1. Bioinformatics Error
  2. User Error

# Rejected Samples

## Targeted Amplicon

	Count	Avg. Genome Coverage	Avg. Depth of Coverage
Match	91% (69/75)	36%	112,040 x
Mismatch	7% (5/75)	24%	92,039 x
Rejected	2% (2/75)	NA	2.5x
Removed from analysis	(4)	NA	NA

## SMART-9N WGS

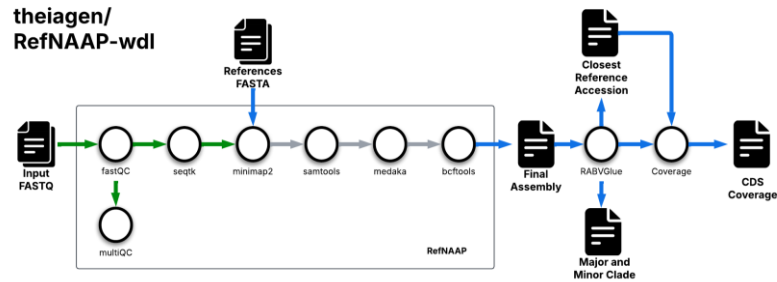
	Count	Avg. Genome Coverage	Avg. Depth of Coverage
% Match	66% (44/68)	34%	7x
% Mismatch	0	NA	NA
% Rejected	35% (24/68)	0	1x
Removed from analysis	(12)	NA	NA

- Significantly lower depth of coverage in rejected samples
- Proportion Host reads : Viral Reads Influences Depth of Coverage
- Targeted Amplicon Sequencing reliably sequenced more samples than SMART-9N WGS
- LN34 Ct value is not a useful predictor of sequencing success

# Key Takeaways

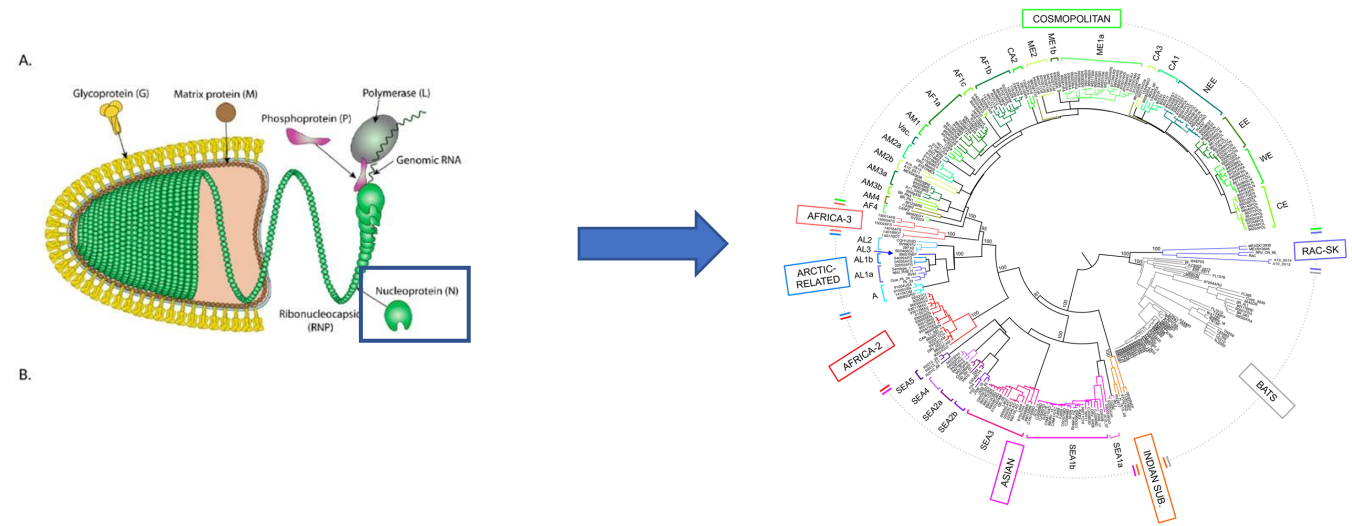
## Antigenic variant typing & NGS Phylogenetics

- Antigenic variant typing provides limited results
- Sequencing provides more detailed view RABV diversity, w/out CDC sendout



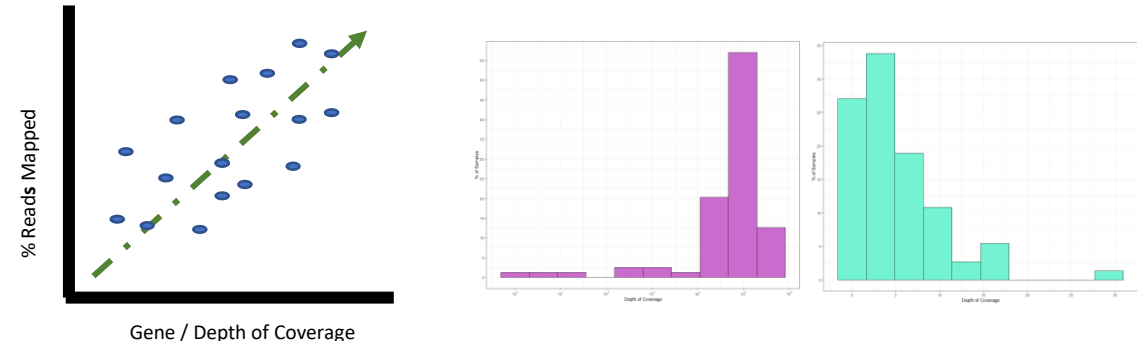
## Target Amplicon VS SMART-9N WGS

- Both methods are strongly influenced by % reads mapped (viral signal : host noise), especially SMART-9N
- LN34 Ct is NOT suitable for determining sequence success
- Targeted Amplicon more reliably sequenced RABV than SMART-9N



## Theiagen Genomics RefNAAP-WDL

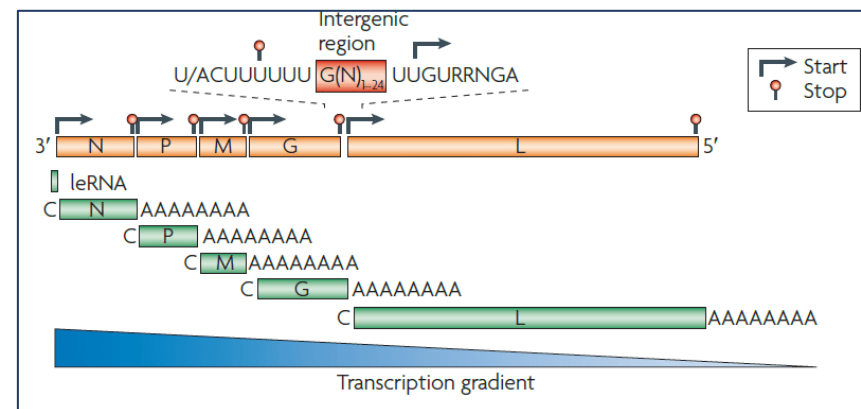
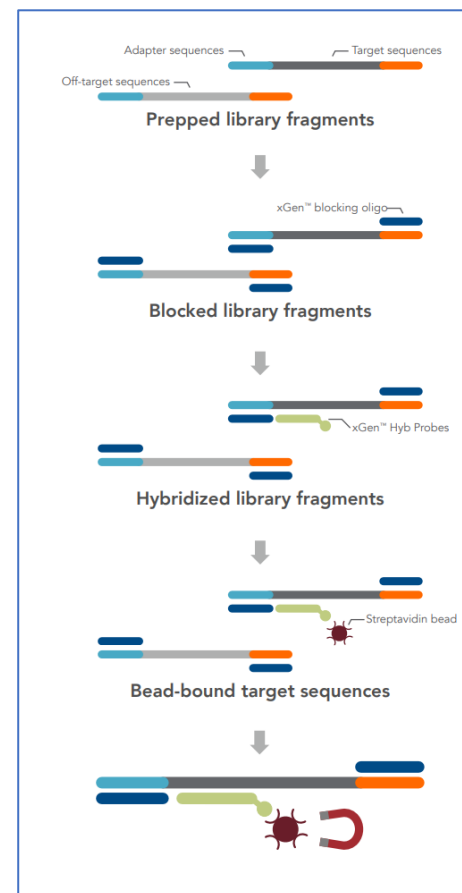
- Fully automated bioinformatics workflow for rabies sequencing
- Both Targeted Amplicon & Metagenomic WGS data
- Provides QC, Major & Minor clades, Closest NCBI full genome accession, CDS Coverage



# Next Steps

1. Re-sequence Mismatched Samples
  - Evidence of Cross contamination
  - Eliminate Analyst Error
2. Increase percent reads mapped for SMART-9N WGS method
  - Increase RABV RNA signal – Decrease Host RNA noise
  - Hybrid Capture (Target Enrichment)
  - Target polyadenylated transcripts using SMART Primers

## Hybrid Capture



# Acknowledgments

- Drew Francis, M(ASCP) – ASPHL
- Crystal Gigante, PhD - CDC
- Inês Mendes, PhD – Theigen Genomics
- Michelle Scribner, PhD – Theigen Genomics





# Questions

# Cost Analysis

## Targeted Amplicon

Reagent	Cost / Unit	Rxn / Unit	Cost / Rxn
ONT Ligation Sequencing Kit	3,920	576	6.81
PCR Barcoding Kit	311.64	72	4.33
NEB Companion Module	985.99	288	3.42
SSIV Master Mix	1,150.56	250	4.60
TaKaRa LA Taq	1,605.75	200	8.03
Ampure XP Beads	1,821.25	360	5.06
Qubit dsDNA HS	404	500	0.81
<b><u>Flow Cells</u></b>	<b><u>4,900</u></b>	<b><u>96</u></b>	<b><u>51.04</u></b>
Primers	191.78	1000	0.19
<b>Total Cost per Sample</b>			<b>\$84.29</b>

*Representative of 12 samples / library.  
96 sample / library options available.*

## SMART-9N WGS

Reagent	Cost / Unit	Rxn / Unit	Cost / Rxn
Turbo Dnase	155.04	500	.31
RNA Clean XP	1106.33	440	2.52
dNTPs	685.92	1000	.66
SSIV RT	1762.56	200	8.81
Rnase OUT	230.05	125	1.84
TaKaRa LA Taq	1605.75	400	4.01
AMPure XP Beads	1821.25	1200	1.52
Qubit dsDNA HS	404	500	0.81
<b><u>Flow Cells</u></b>	<b><u>4900</u></b>	<b><u>96</u></b>	<b><u>51.04</u></b>
Rapid PCR Barcoding Kit	685.02	144	4.76
Primers	80.36	1000	0.08
<b>Total Cost per Sample</b>			<b>\$76.36</b>