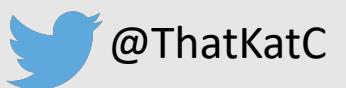


Making Genomic Surveillance Deliver

A Lineage Classification and Nomenclature System to Inform Rabies Elimination

■ Kathryn Campbell

k.campbell.1@research.gla.ac.uk

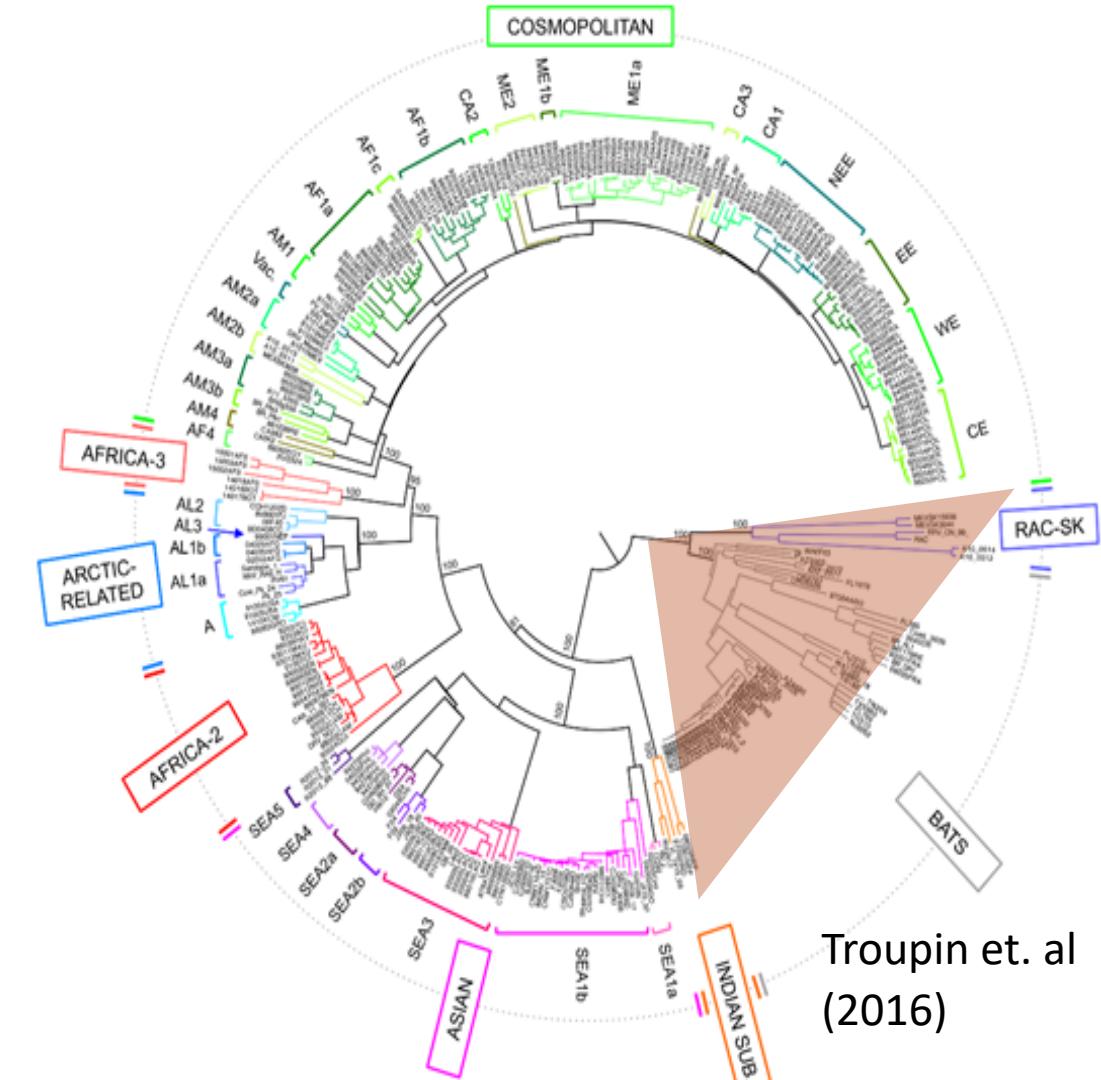


University
of Glasgow



The global rabies phylogeny reveals large-scale geographic structure

- 2 major phylogenetic groups; bat related and dog related (5-7 clades)
- Same geographical area group closely together
 - Host species also important but less so
 - Major clades are segregated by barriers
 - Oceans, mountain ranges, human factors



Greater phylogenetic resolution is needed to guide control programmes

- Progress towards elimination, should see reduction in viral diversity
- Need to see if any lineages have become extinct, or if any are persisting
 - Use this to direct elimination efforts
 - Monitor how we are progressing towards elimination

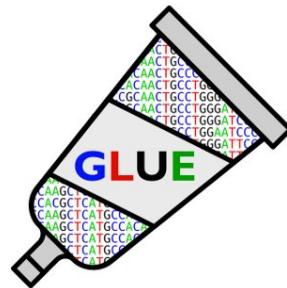
BUT

- Needs a high-definition classification system
 - Current system is very broad – often only 1 subclade circulating
 - Can't track changes
 - Needs to be universal
 - Lots of studies, but all define and name lineages differently
 - Can't compare and communicate these

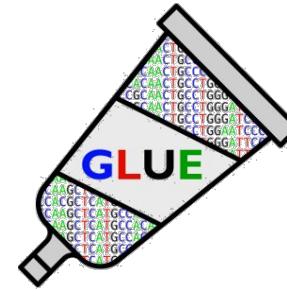
Need one, high definition system to classify all rabies globally



SCAN ME



RABV-GLUE

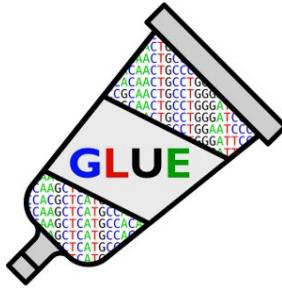


Find Relevant Sequences and Studies

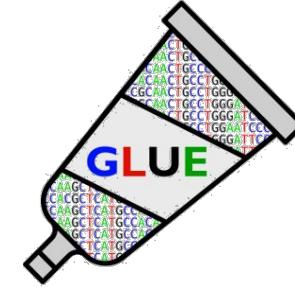
Analyse Your Own Sequences



SCAN ME



RABV-GLUE



Find Relevant Sequences and Studies

Analyse Your Own Sequences

Filter for only sequences of interest

Coverage of Genome Region ▾ Whole genome ▾ is at least 90 %

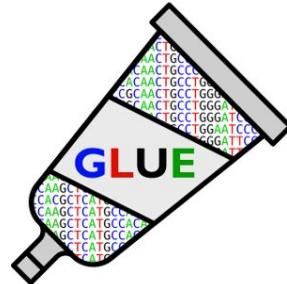
Country of Origin ▾ contains Tanzania

See and download all the relevant sequences and information about them

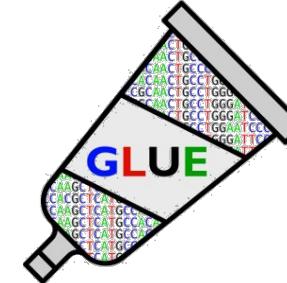
NCBI Nucleotide ID	Major / Minor Clade	NCBI Entry Creation Date	NCBI Last Update Date	Sequence Length	Isolate ID	Country of Origin	Host Species	Collection Year Range	Reference
KF155002	Cosmopolitan AF1b	13-Jul-2013	13-Jul-2013	11923	RV2772	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 23822119 ♂
KR534217	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2490	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534218	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2491	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534219	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2492	Tanzania (TZA)	<i>Canis familiaris</i>	2007	PubMed 27774283 ♂
KR534220	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2493	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534228	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11914	RV2788	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 27774283 ♂
KR534229	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11849	RV2789	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 27774283 ♂



SCAN ME



RABV-GLUE



Find Relevant Sequences and Studies

Filter for only sequences of interest

Coverage of Genome Region	Whole genome	is at least	90	%
Country of Origin	contains	Tanzania		

See and download all the relevant sequences and information about them

NCBI Nucleotide ID	Major / Minor Clade	NCBI Entry Creation Date	NCBI Last Update Date	Sequence Length	Isolate ID	Country of Origin	Host Species	Collection Year Range	Reference
KF155002	Cosmopolitan AF1b	13-Jul-2013	13-Jul-2013	11923	RV2772	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 23822119 ♂
KR534217	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2490	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534218	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2491	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534219	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2492	Tanzania (TZA)	<i>Canis familiaris</i>	2007	PubMed 27774283 ♂
KR534220	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11923	RV2493	Tanzania (TZA)	<i>Canis familiaris</i>	2008	PubMed 27774283 ♂
KR534228	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11914	RV2788	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 27774283 ♂
KR534229	Cosmopolitan AF1b	06-Nov-2015	01-Nov-2016	11849	RV2789	Tanzania (TZA)	<i>Canis familiaris</i>	2010	PubMed 27774283 ♂

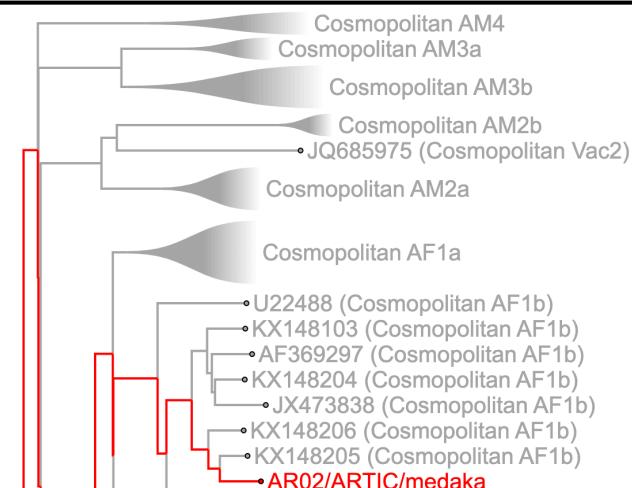
Analyse Your Own Sequences

Relevant information about the sequences

Sequence	Identified as RABV?	Major clade	Minor clade	Closest full genome reference sequence	N	P	M	G	L
AR02/ARTIC/medaka	Yes	Cosmopolitan	Cosmopolitan AF1b	KX148205	99.33%	100.00%	100.00%	100.00%	94.68%
AR03/ARTIC/medaka	Yes	Cosmopolitan	Cosmopolitan AF1b	KX148205	99.33%	100.00%	100.00%	100.00%	94.68%

See the closest relatives

Are there any mutations?



4	5	6	7
D	K	I	V
G A C A A G A T T G T A			
8			9

Data Curation

- Over 26,000 publicly available RABV sequences
- c10,000 entries missing essential metadata
 - Collection year, host, location
- Manual searching in publications
- <http://rabv-glue.cvr.gla.ac.uk/#/home>

RABV-GLUE

A Sequence Data Resource for Rabies Virus

Rabies virus (RABV) is a neglected zoonotic disease that causes around 59,000 human deaths each year, with a near 100% mortality rate after the onset of symptoms. The virus is a member of the Lyssavirus genus, within the Rhabdoviridae family, which is characterised by a single stranded, negative-sense RNA genome.

Infection with RABV can occur in all species of mammal, but up to 99% of human rabies cases arise from bites from infected domestic dogs. Vaccinating dogs to interrupt transmission is therefore paramount, and a major focus of the Zero by 30' global strategy to eliminate human deaths from dog-mediated rabies by 2030.

RABV-GLUE is a data-centric bioinformatics resource which organises RABV genome sequence data along evolutionary lines. RABV-GLUE aims to leverage new and existing RABV sequences in order to improve our understanding of the epidemiology and pathology of RABV.

The web version of RABV-GLUE can be used for basic analysis. An offline version of the resource can be used for more advanced work.

NOTE: we do not store any sequences submitted to RABV-GLUE!

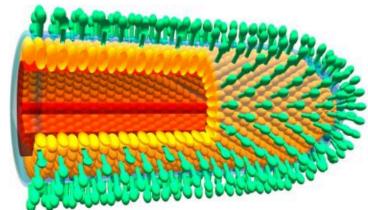
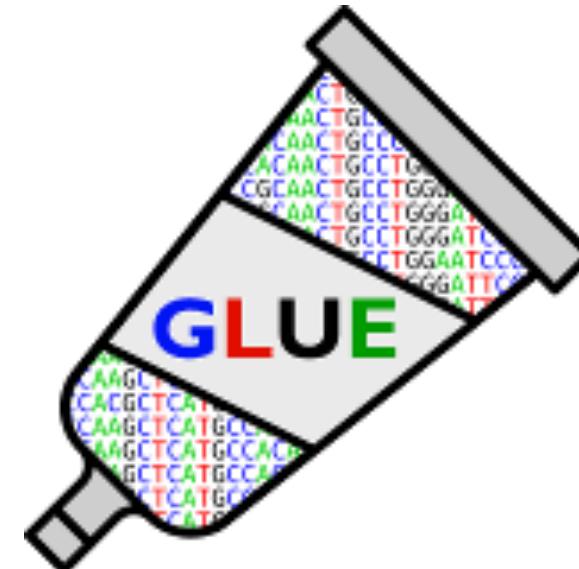


Image copyright www.cronodon.com, used with permission





SCAN ME

MAD DOG

Method for Assignment, Definition and Designation Of Global Lineages

Increase the resolution

See the patterns

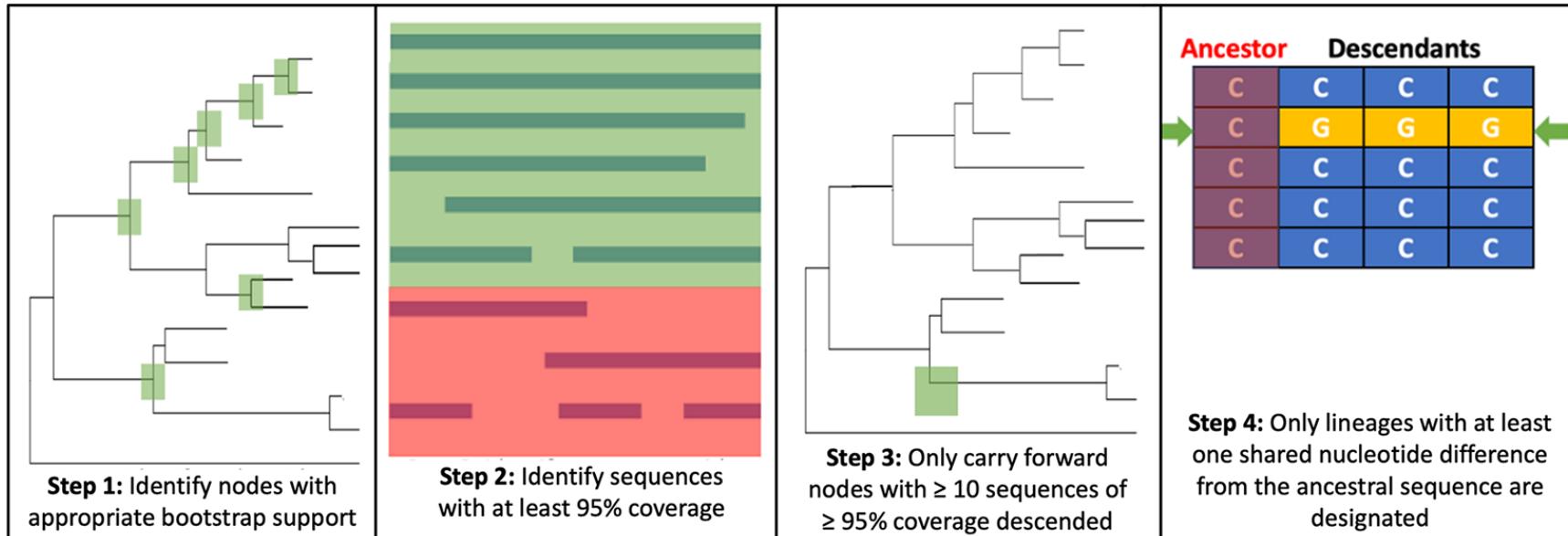
Stop the Spread

METHODS

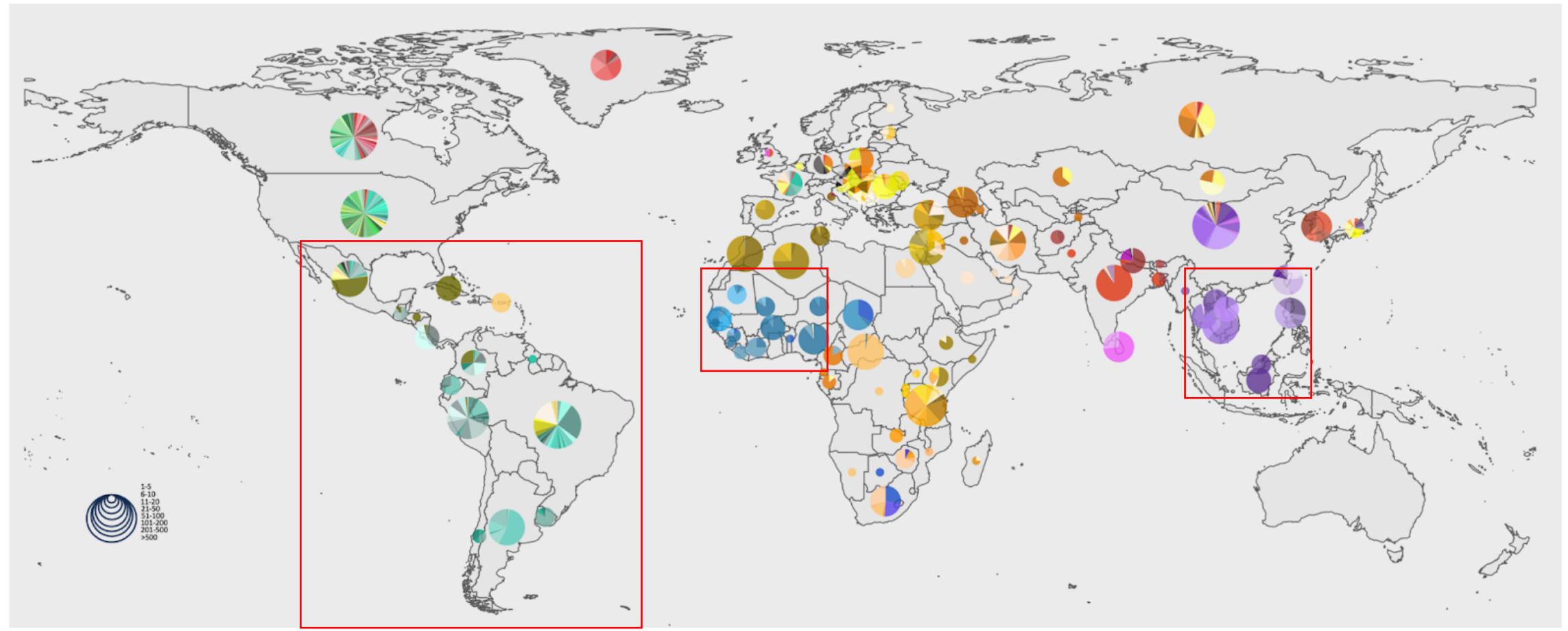
Data Processing



Lineage Designation

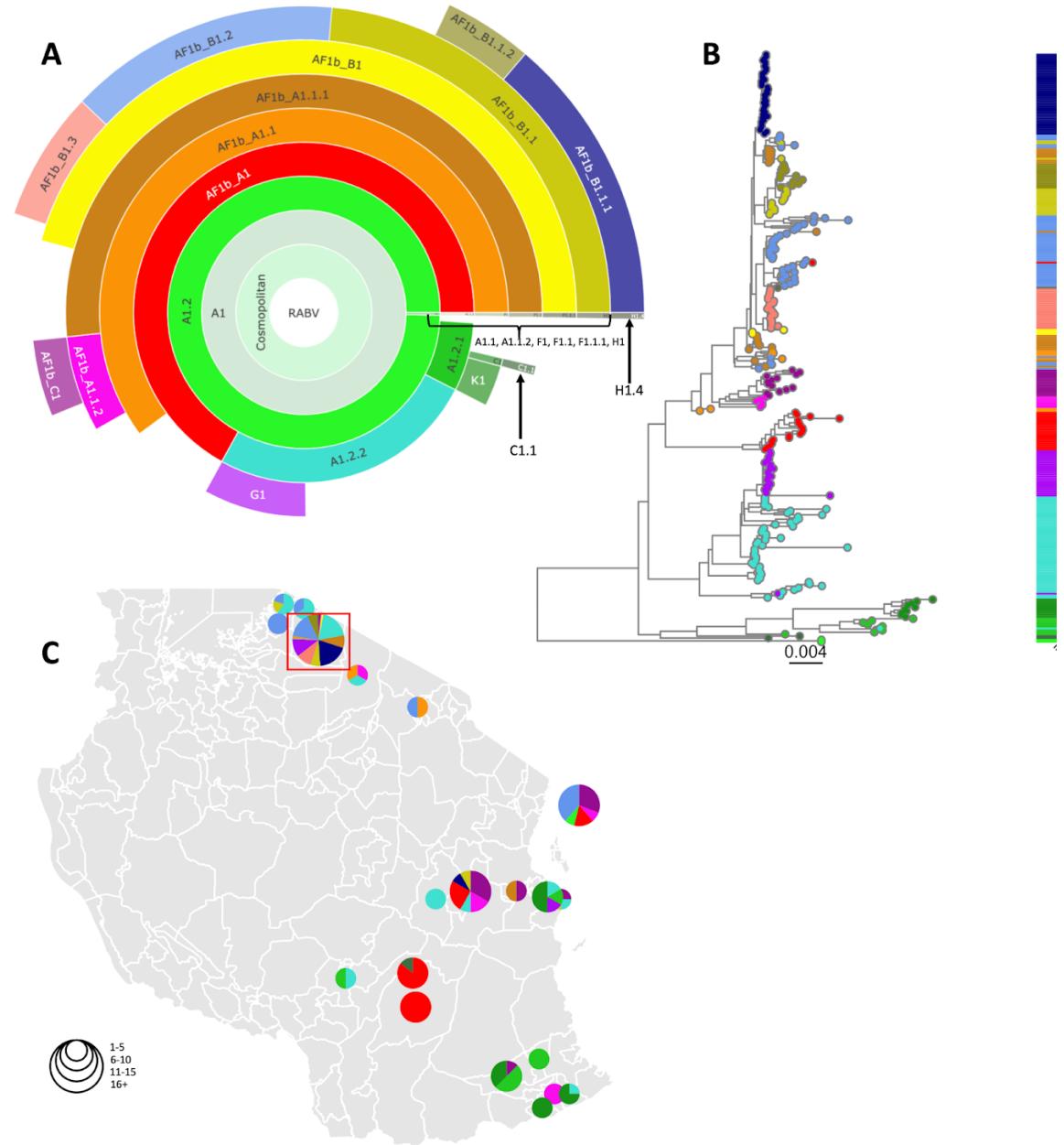


- Modelled after SARS-CoV-2 lineage designation
- Different parameters to consider for rabies!
 - Time scale
 - Existing classifications
 - Tracked since emergence vs decades of endemic circulation



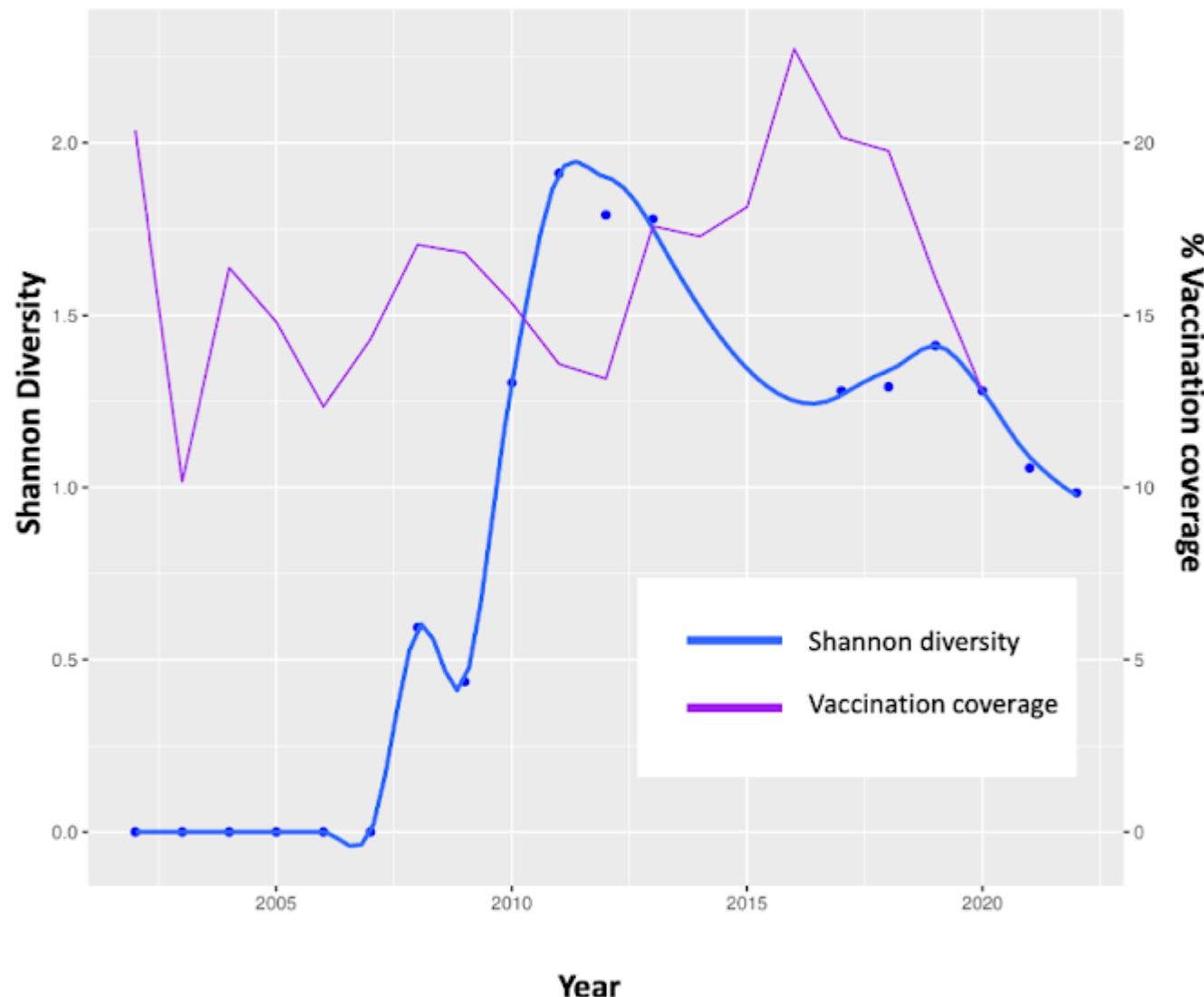
Lineage Overview

- 216 WGS from Serengeti (+90 from Tanzania)
 - 14 lineages (+4 from Tanzania)
 - 4 Serengeti only, 7 also seen in other countries across Africa



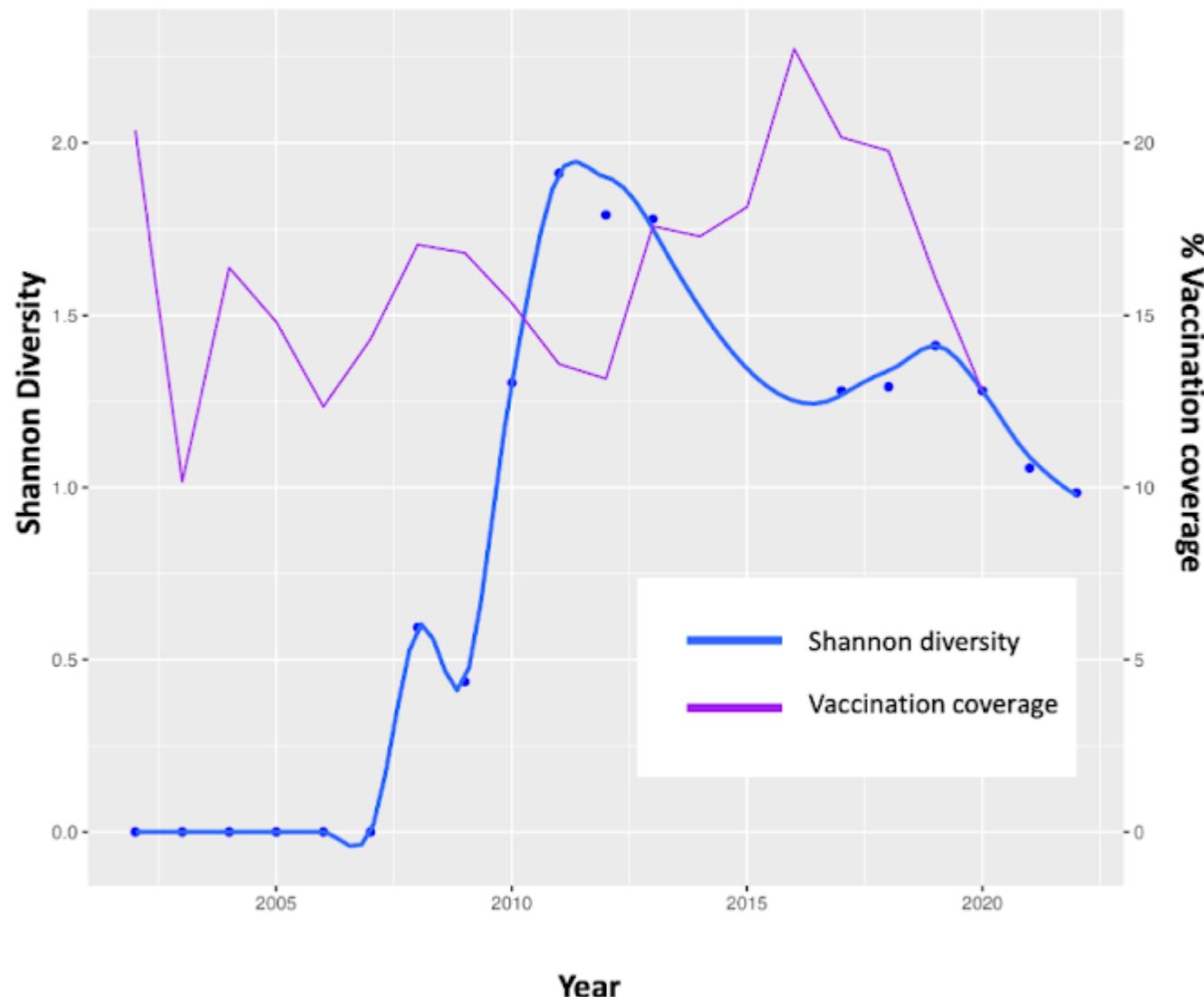
Changes over time

- Steadily reducing viral diversity
- Some lineages persist
 - Some emerge and then go extinct
 - 9 not seen in 2 years, 6 not seen in 5 years



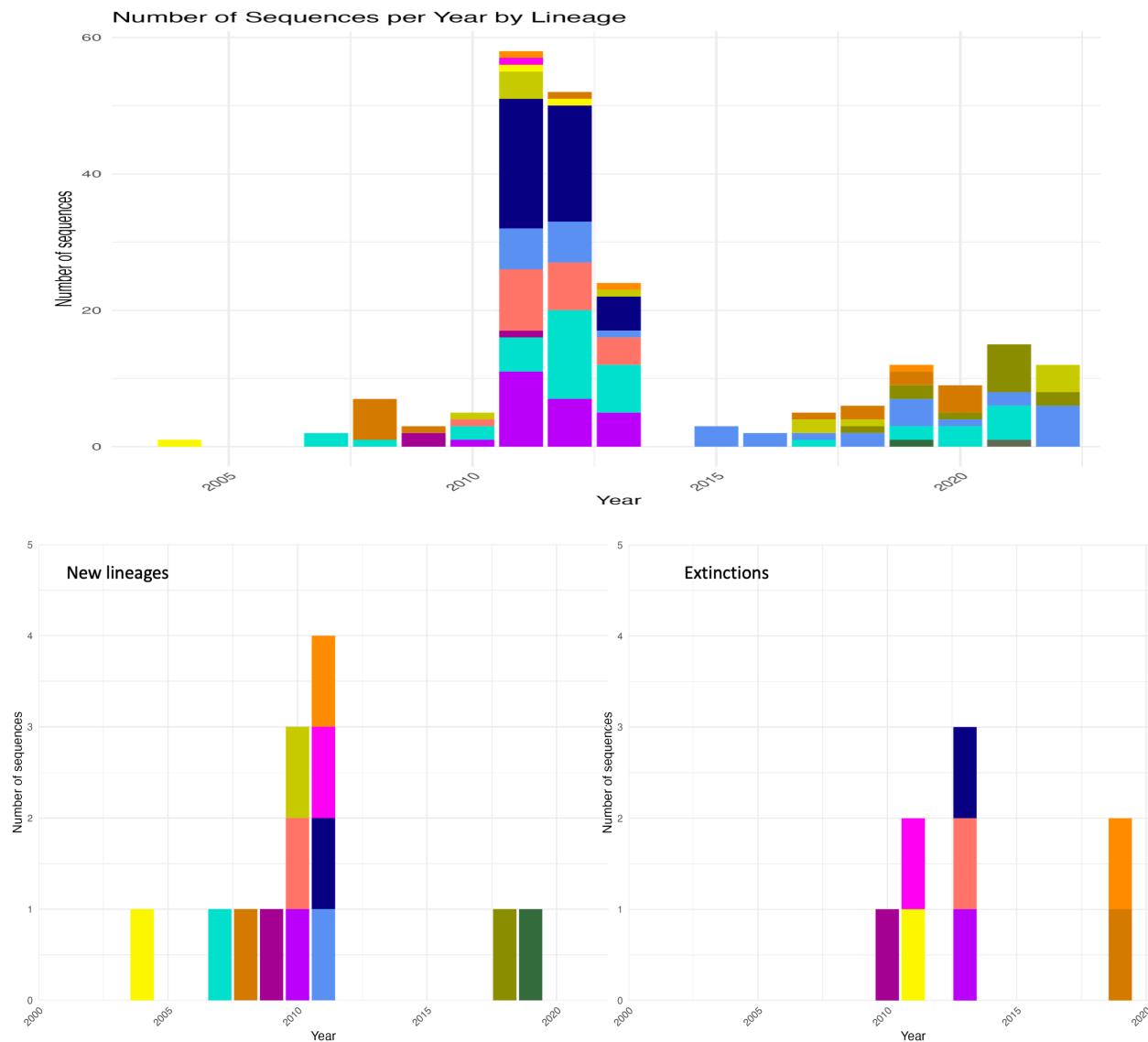
Changes over time

- Steadily reducing viral diversity
- Some lineages persist
 - Some emerge and then go extinct
 - 9 not seen in 2 years, 6 not seen in 5 years



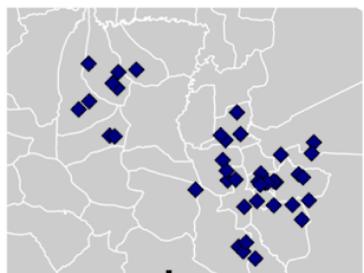
Changes over time

- Steadily reducing viral diversity
- Some lineages persist
 - Some emerge and then go extinct
 - 9 not seen in 2 years, 6 not seen in 5 years

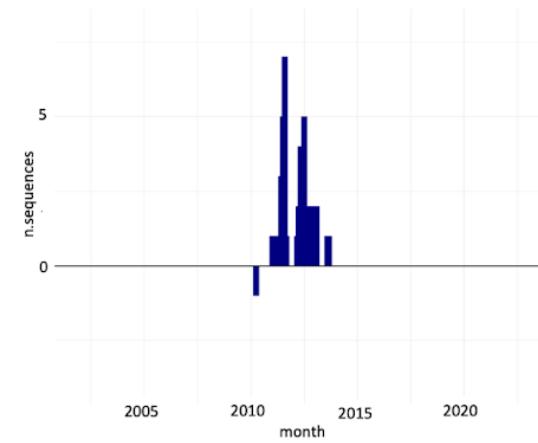


Metapopulation dynamics

- Cases can be introduced or reintroduced from wider Tanzania
- Some establish, others are single incursions
- Vaccination coverage needs to be maintained!

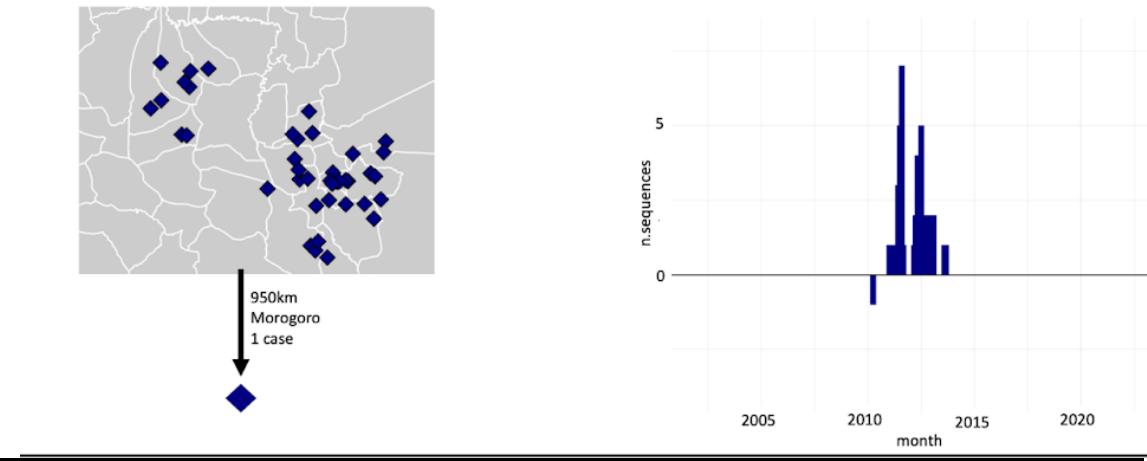
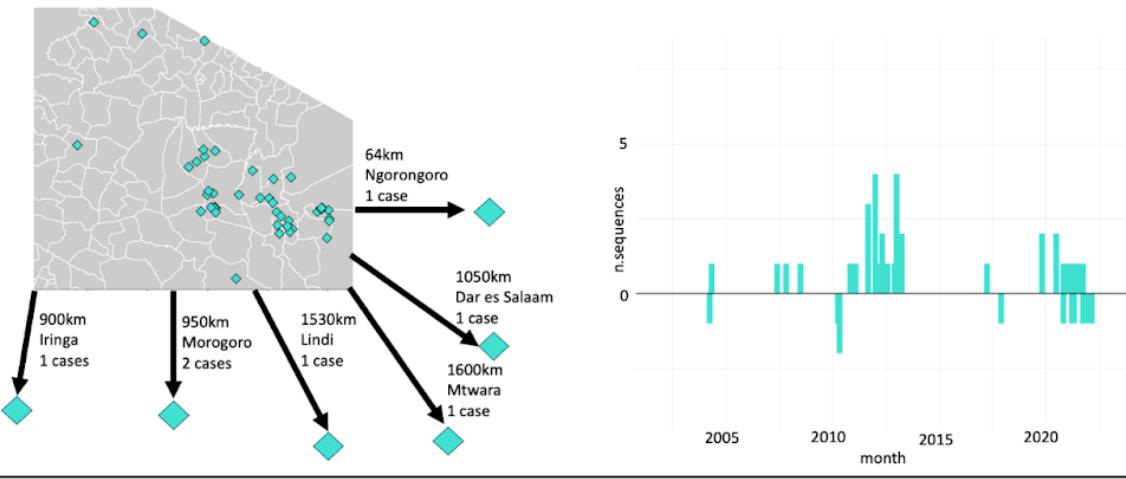


950km
Morogoro
1 case



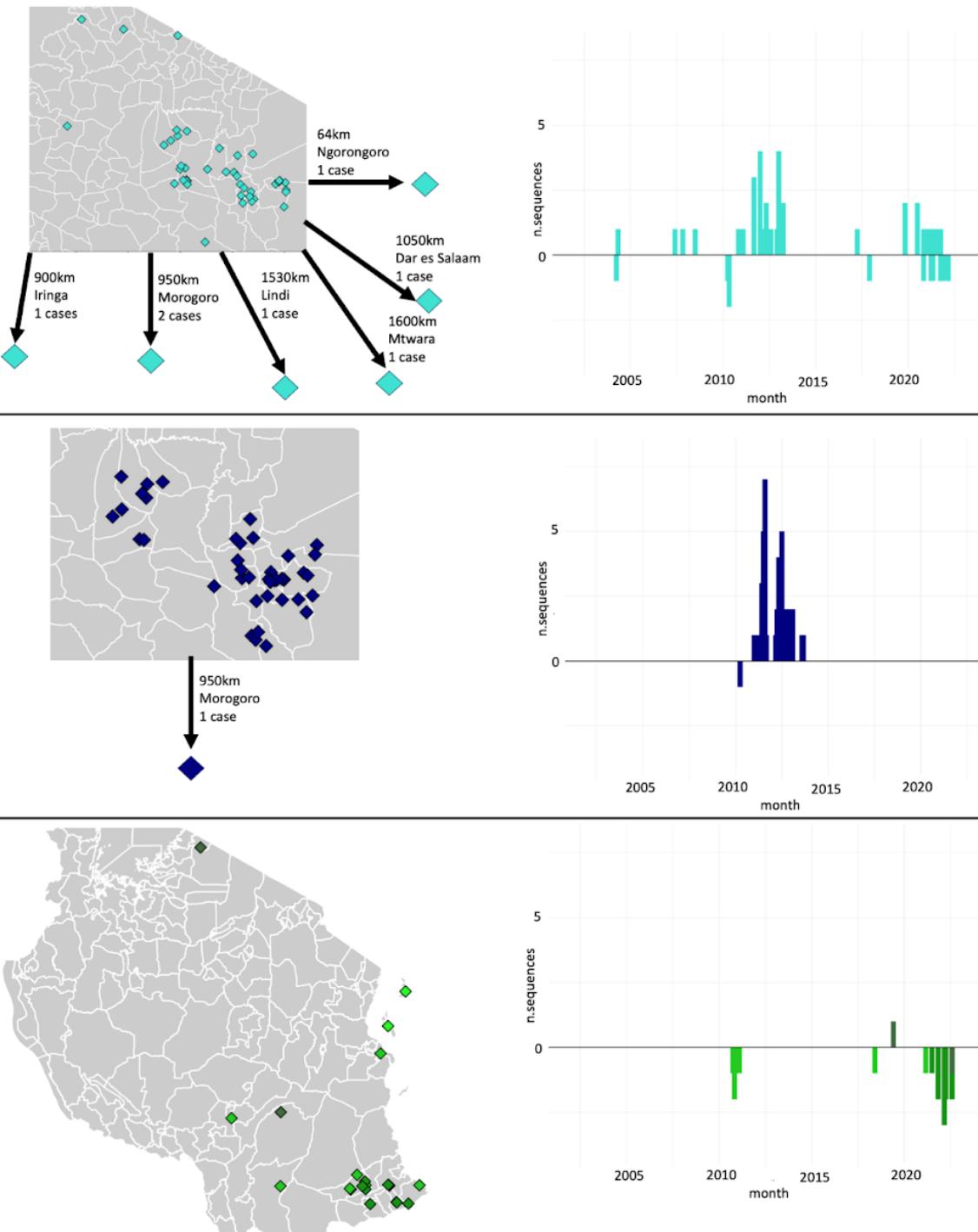
Metapopulation dynamics

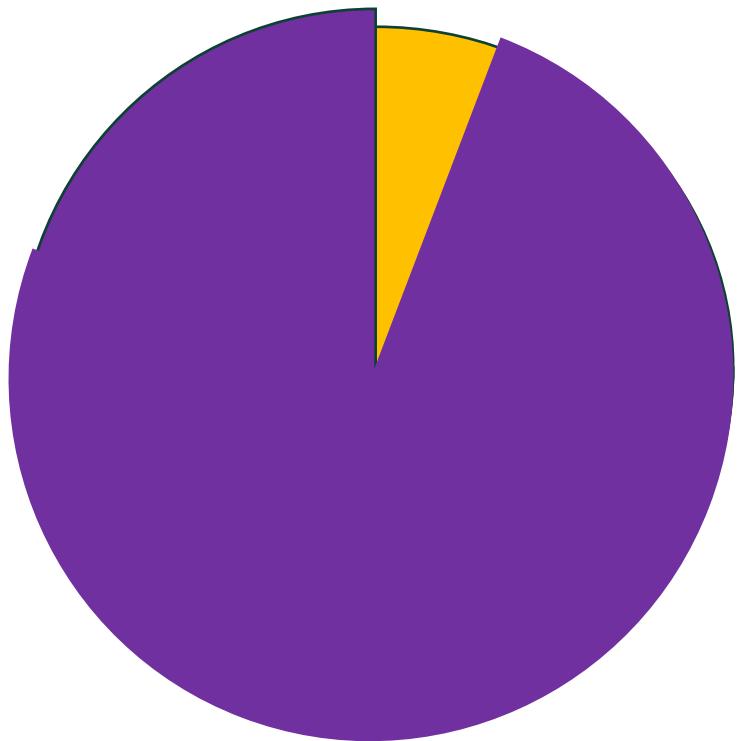
- Cases can be introduced or reintroduced from wider Tanzania
- Some establish, others are single incursions
- Vaccination coverage needs to be maintained!

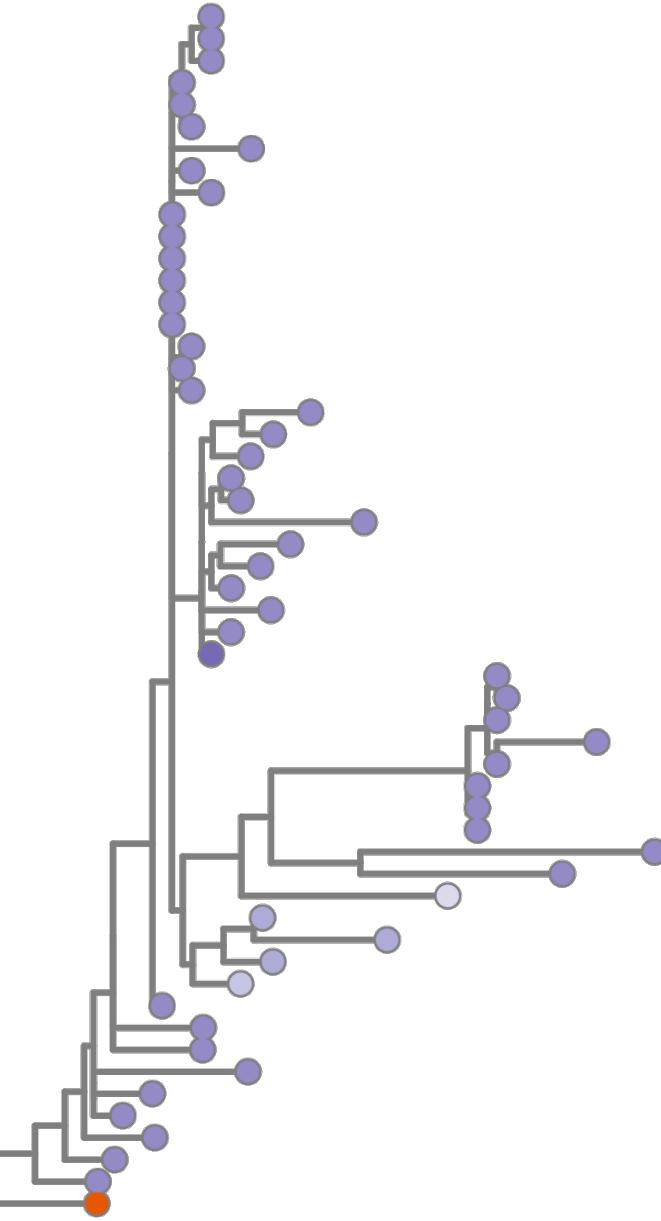
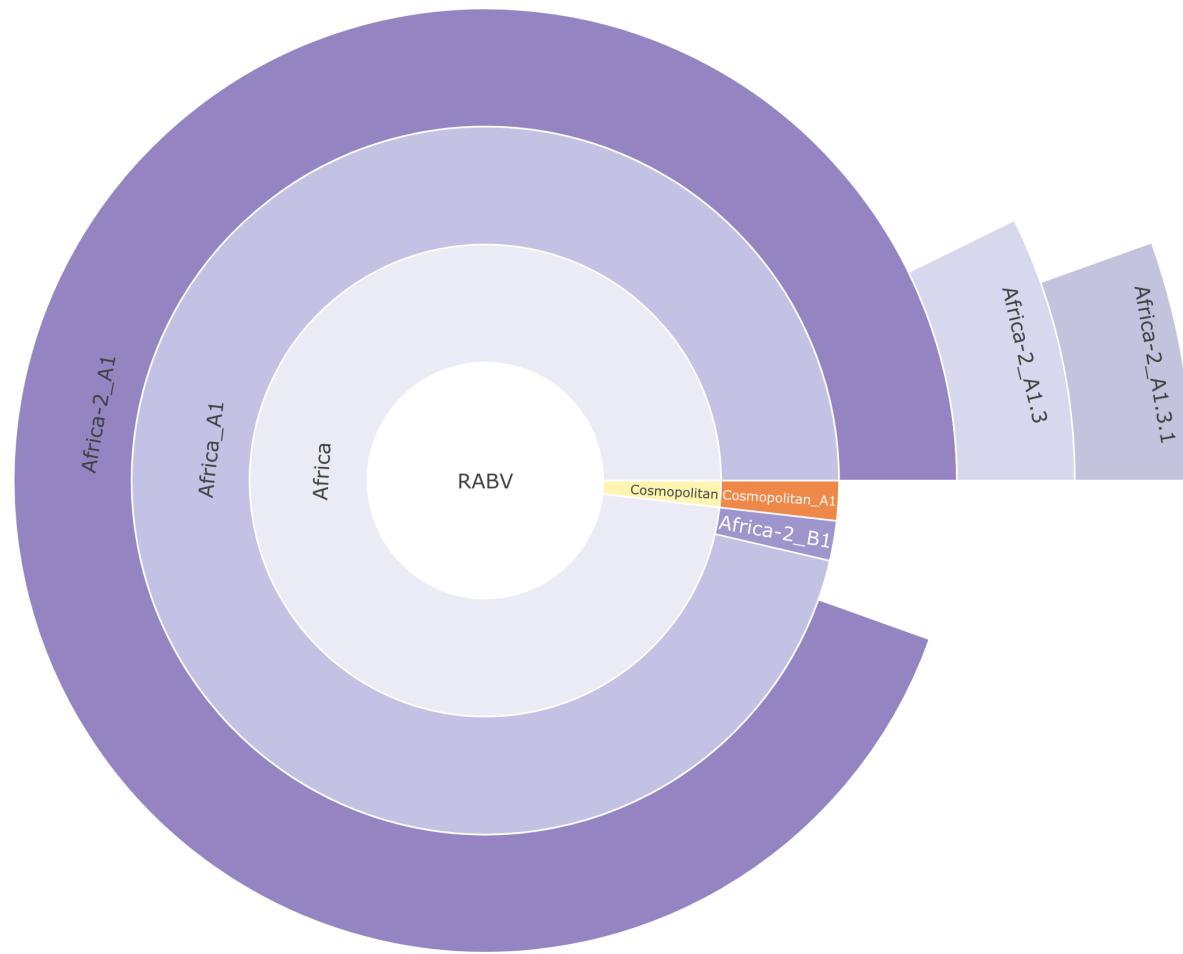


Metapopulation dynamics

- Cases can be introduced or reintroduced from wider Tanzania
- Some establish, others are single incursions
- Vaccination coverage needs to be maintained!







0.01

MAD DOG

Method for Assignment, Definition and Designation Of Global Lineages

INPUT

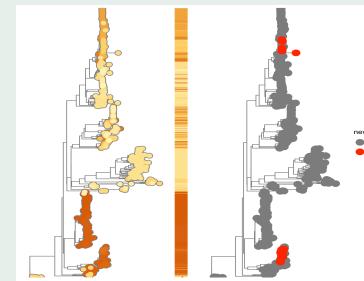
Sequences to be tested and corresponding metadata

OUTPUTS

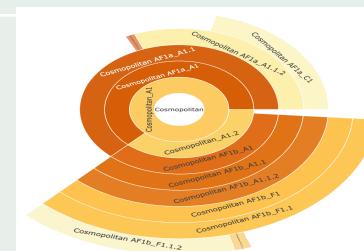
Lineage information

lineage	country	year_first	year_last	n_seqs	parent
Cosmopolitan AF1a_A1.1	c("Ethiopia", "Kenya", "Somalia")	1988	2018	25	Cosmopolitan AF1a_A1

Creation of and placement in a relevant phylogeny



Hierachal figure of relevant lineages



Emerging and undersampled areas

lineage	tips	distance	country	year_first	year_last
Cosmopolitan AF1b_F1.1_E1	5	0.07246488	Kenya	2021	2022

Singletons of Interest

lineage	n_singletons	singleton_countries	singleton_years
Cosmopolitan AF1b_D1.1	1	Tanzania	2015

Assignment or Designation?

- **Assignment** is an initial, quick, less computationally intensive step for finding out roughly which lineages sequences are from
 - Sequences are compared against a reference set for each lineage, and are assigned the lineage of the closest match in the reference
- **Designation** is a slow, thorough, computationally intensive process for finding new lineages and accurately assigning existing lineages
 - This involves comparison with all public whole genome and N gene sequences, and a tree built from your sequences and relevant public ones.
 - The position sequences fall in the tree compared to existing sequences tells us the lineage. We then check to see if all of the rules are met on any new groups of sequences to determine new lineages

Usage

- MADDODG is really easy to use!
- 2 steps to set up:
 - Download the MADDODG repository from GitHub
 - Set up the MADDODG conda environment
- 2 data requirements:
 - A fasta file with all the sequences to be tested
 - A csv metadata file with sequence IDs exactly matching the fasta file, and the year and country of collection for each sequence
- 4 steps to run:
 - Pull the GitHub repository to make sure any changes are integrated
 - Enter the MADDODG conda environment
 - Navigate to the MADDODG folder in command line
 - Run the command **sh designation.sh** or **sh assignment.sh**

Usage

- MADDODG is really easy to use!
- 2 steps to set up:
 - Download the MADDODG repository from GitHub
 - Set up the MADDODG conda environment
- 2 data requirements:
 - A fasta file with all the sequences to be tested
 - A csv metadata file with sequence IDs exactly matching the fasta file, and the year and country of collection for each sequence
- 4 steps to run:
 - Pull the GitHub repository to make sure any changes are integrated
 - Enter the MADDODG conda environment
 - Navigate to the MADDODG folder in command line
 - Run the command **sh designation.sh** or **sh assignment.sh**

Practical 1

- Follow the instructions on GitHub in **day5/day5_GLUE_MADDOG.md**

Practical 2

- Follow the instructions in **day5_MADDOG_results.md**

Acknowledgements



- Katie Hampson
- Kirstyn Brunker
- Rob Gifford
- Roman Biek
- Rowan Durrant



THE UNIVERSITY
of EDINBURGH

- Andrew Rambaut
- Aine O'Toole
- Verity Hill



UNIVERSITY OF NAIROBI

- Gurdeep Jaswant



- Criselda Bautista

