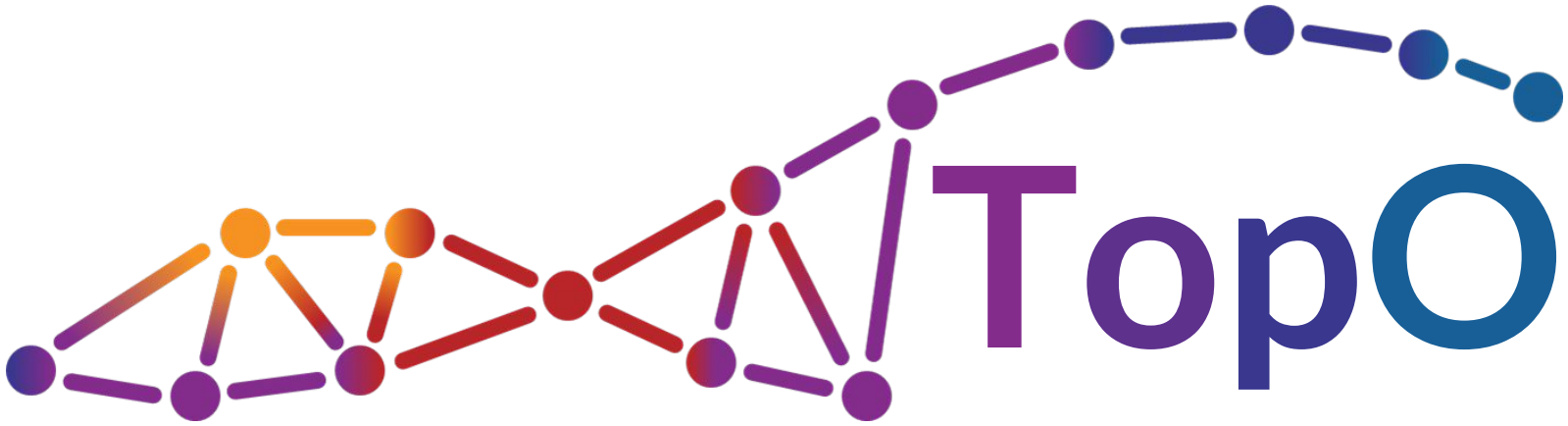


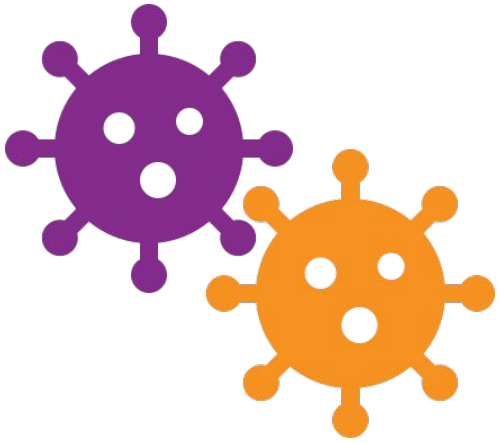
Topological Omics for Non-tree Evolution

Claire Angelie Gabisay and Alfonso Miguel Cruz

Mentored by: CJ Palpal-latoc

16 August 2024

Viral Recombination



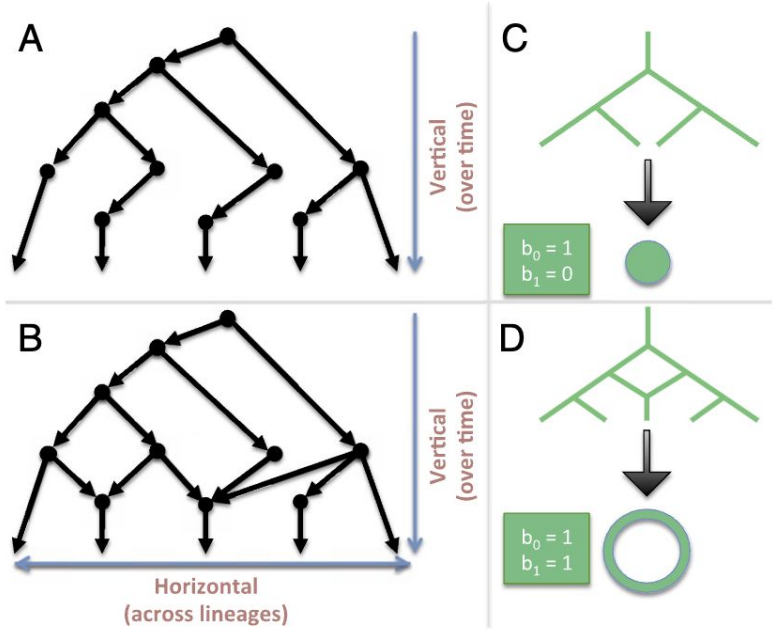
Viral lineages infect the same cell

Combine their genetic material

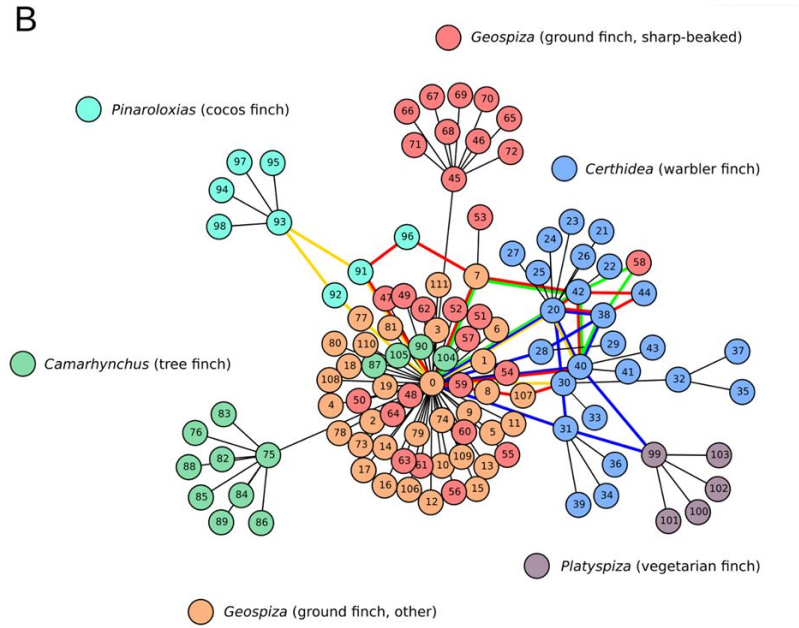
And form a new hybrid lineage

But they are difficult to detect

Existing Work



Recombinants form loops in a phylogenetic network

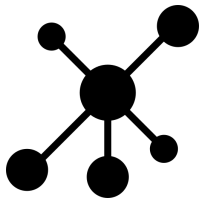


A constructed ancestral recombination graph

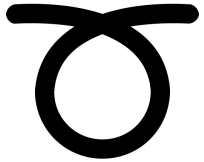
Left: Chan (2013) Right: Cámara (2016)

Topology: The Study of Holes

Types of Holes include:

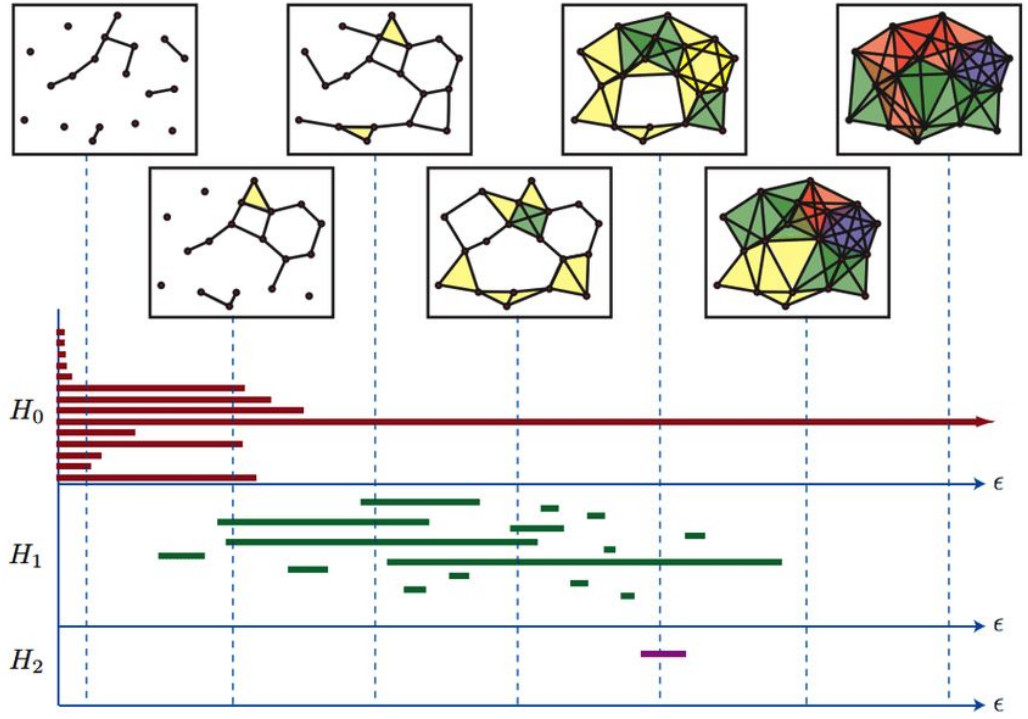


1D holes (connected components)
H_0 - set of 1D holes with β_0 elements



2D holes (loops)
H_1 - set of 2D holes with β_1 elements

3D holes (voids)
H_2 - set of 3D holes with β_2 elements



Ghrist (2008)

where H_k - k^{th} homology group (set of (k-1) dimensional holes),
β_k - Betti number of H_k (number of elements in H_k)

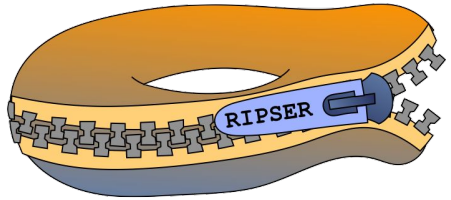
Hole Genome Sequencing

S_1 = TCGAAAGGTTAG
S_2 = TCGATAGGTTGG
S_3 = TCGCTAGGAACC

Sequence Data

	S_1	S_2	S_3
S_1	0	2	6
S_2	2	0	5
S_3	6	5	0

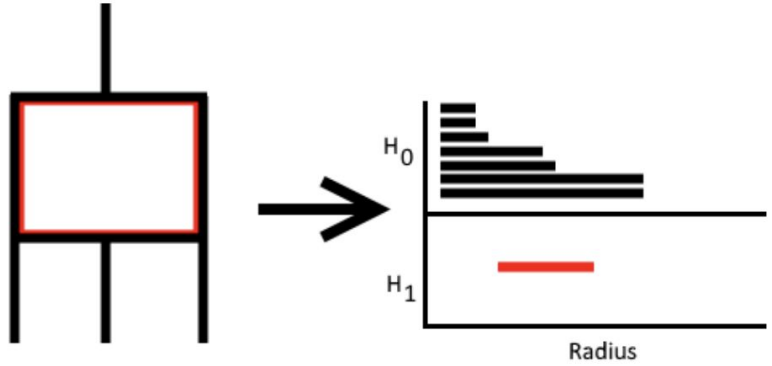
Hamming Distances



Persistent Homology

H_0	(0, 1)
	(0, 3)
	(0, 5)
H_1	(10, 15)

Birth and Death Times



Analysis

Humphreys (2019)

Our Goals

Our main goal:
Determine whether topology can detect recombination

?

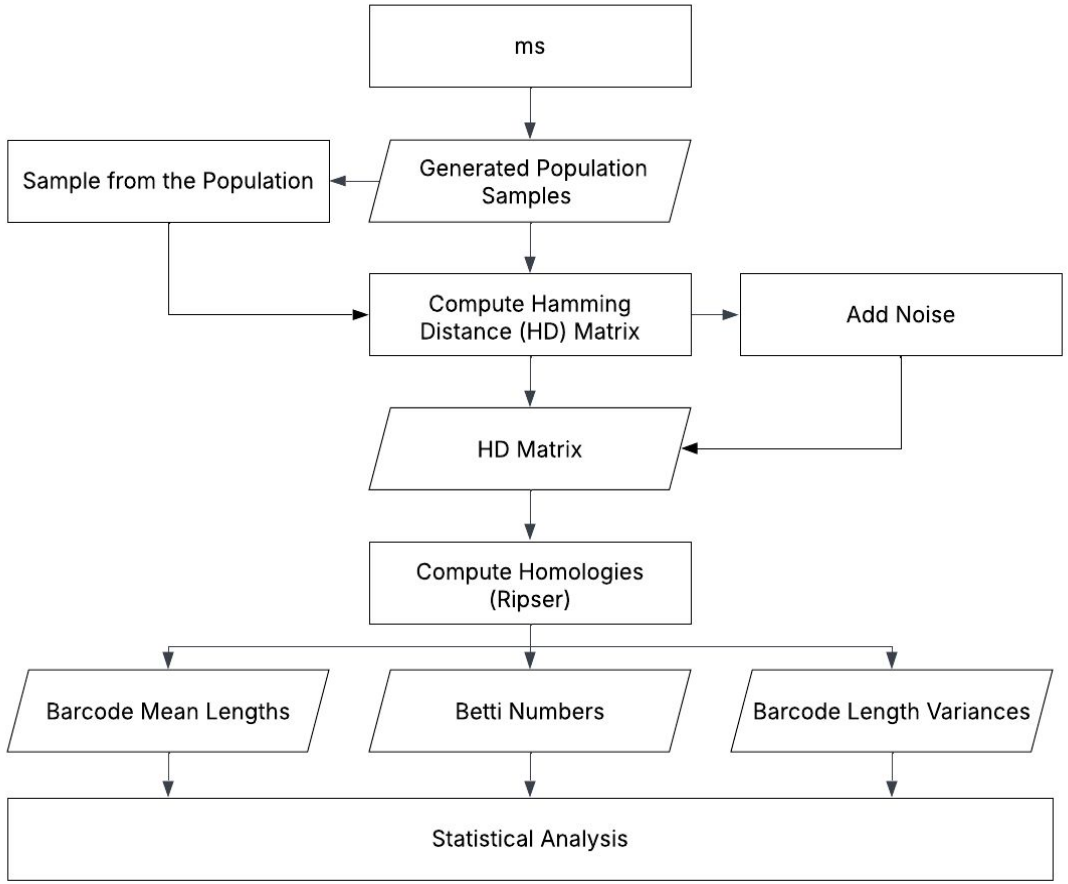
Is topology robust to noise and sparse genomic samples?

?

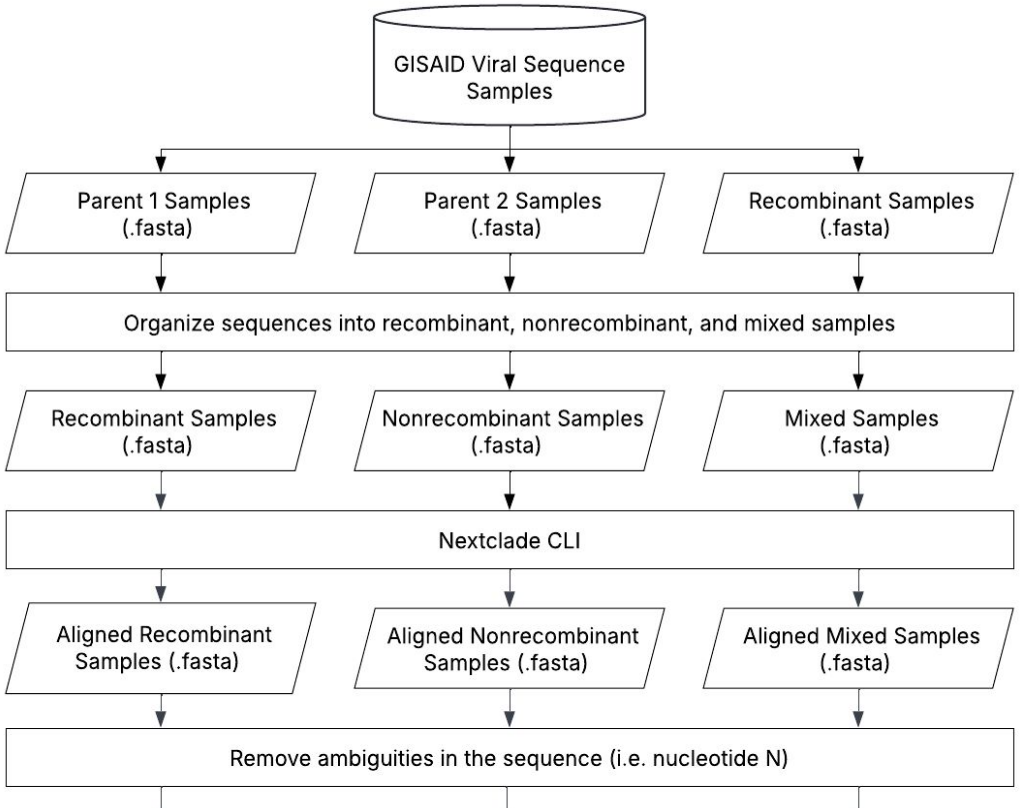
Does recombination change the topology of genomic samples?

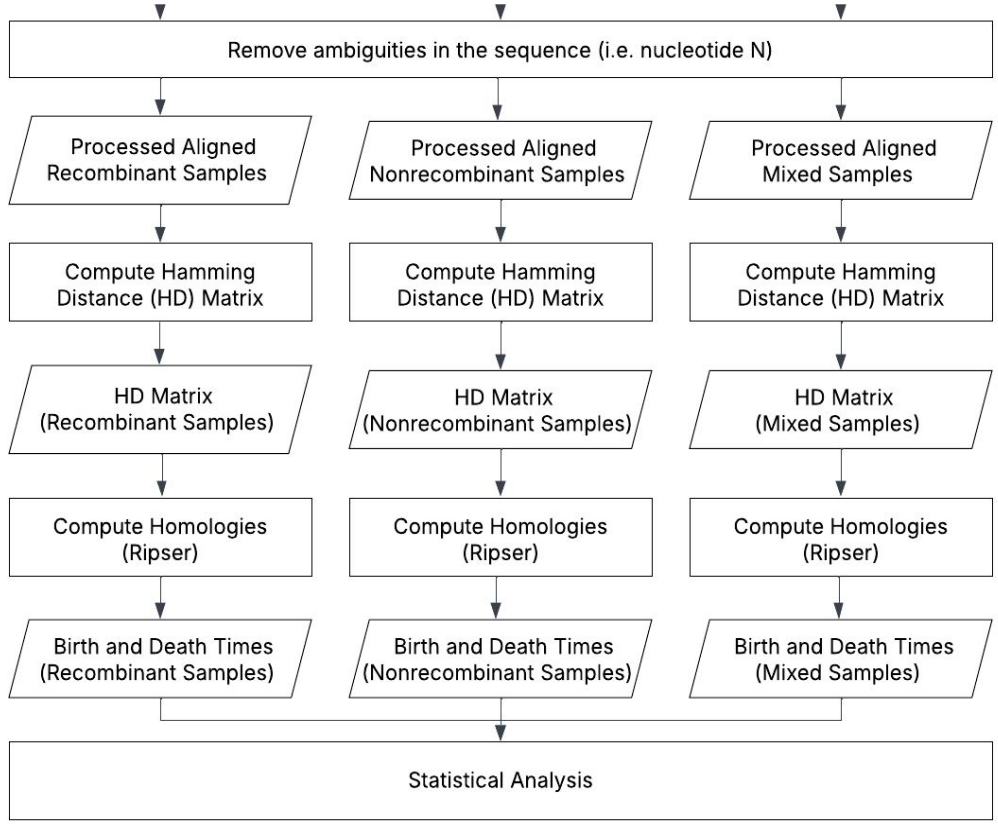
Methodology

Goal 1 Pipeline



Goal 2 Pipeline





3 Recombinant Lineages and their Parent Lineages were taken from GISAID (XBC.1, XBE, XBZ) and were used to get the birth and death time pairs

TopONE

Results

TopONE

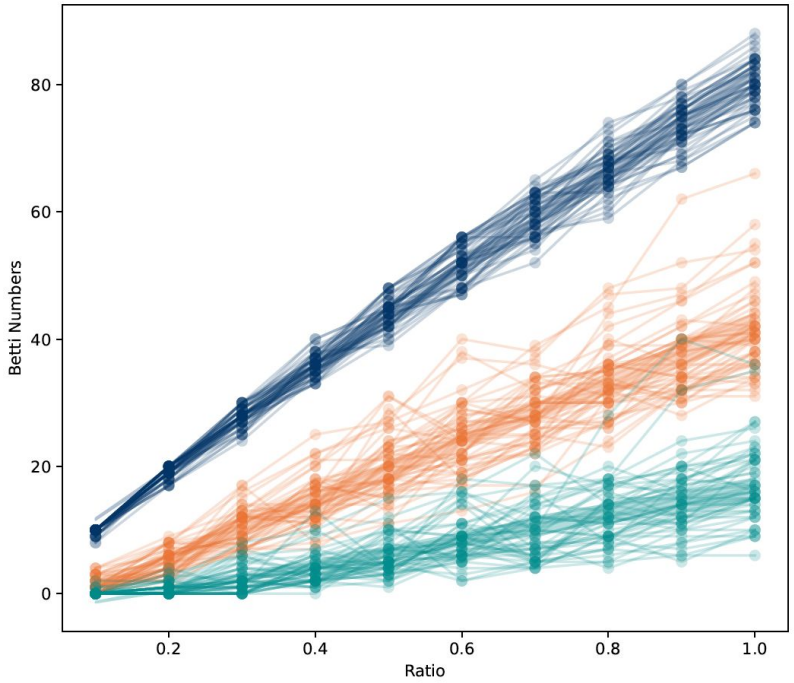
Population Genetics Simulations

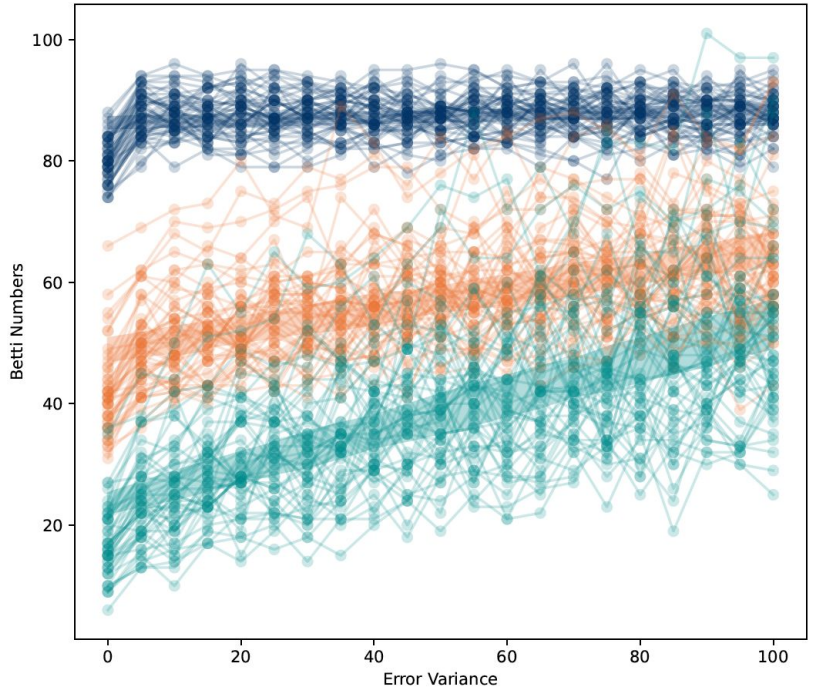
Sensitivity Analysis

TopONE

Betti numbers vs increasing noise and varying sampling sparsity

connected components
loops
voids





Increased sampling → increased data
→ more representative structures

Increased noise distorts actual data
→ creates new loops and voids

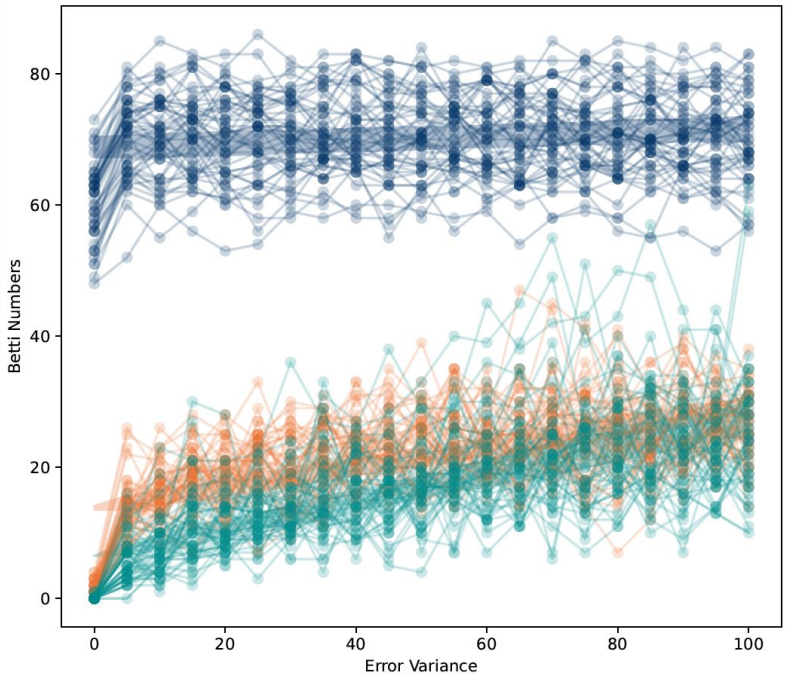
Increased sampling ratio and noise creates new structures

Noise and the mutation rate θ

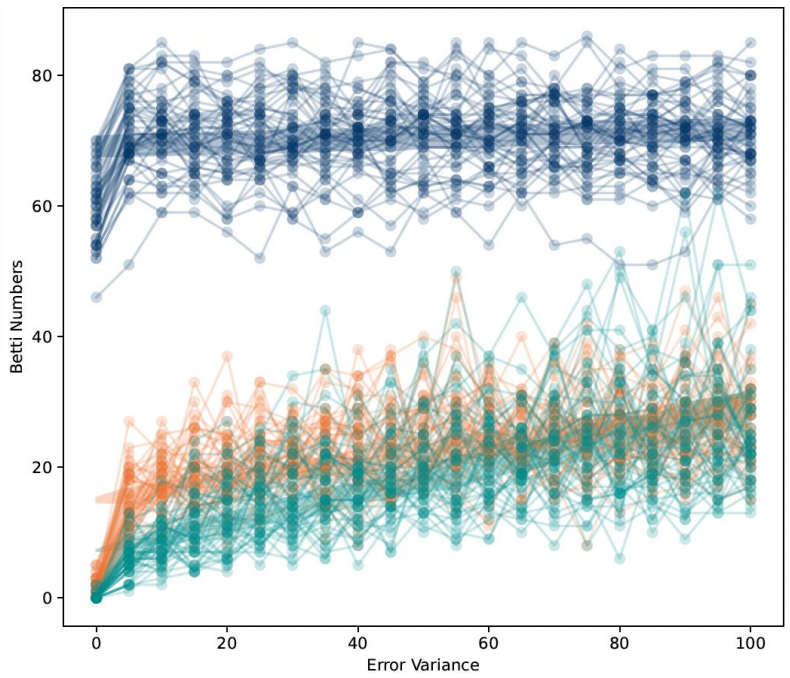
TopONE

connected components
loops
voids

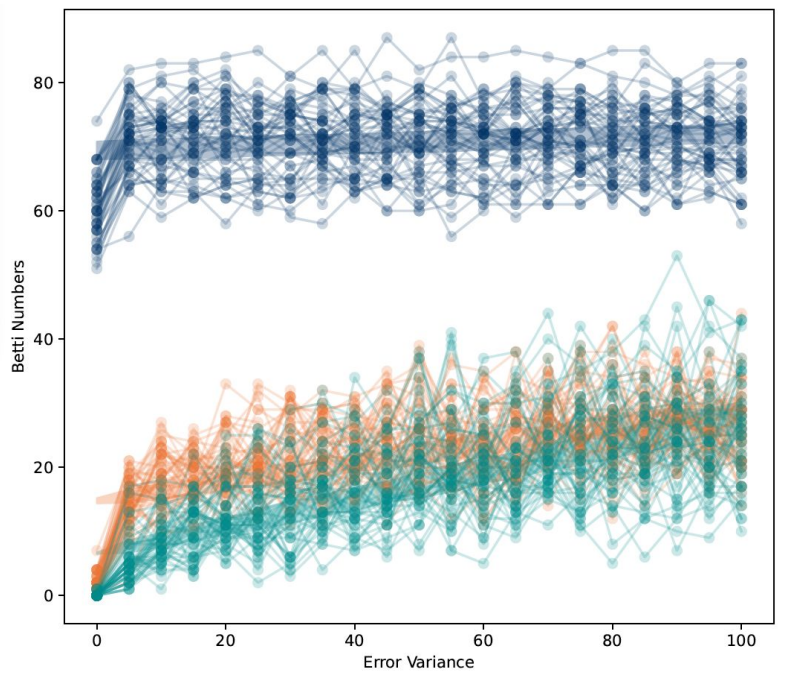
Betti numbers vs increasing noise at varying mutation rates



$\theta = 50$



$\theta = 500$

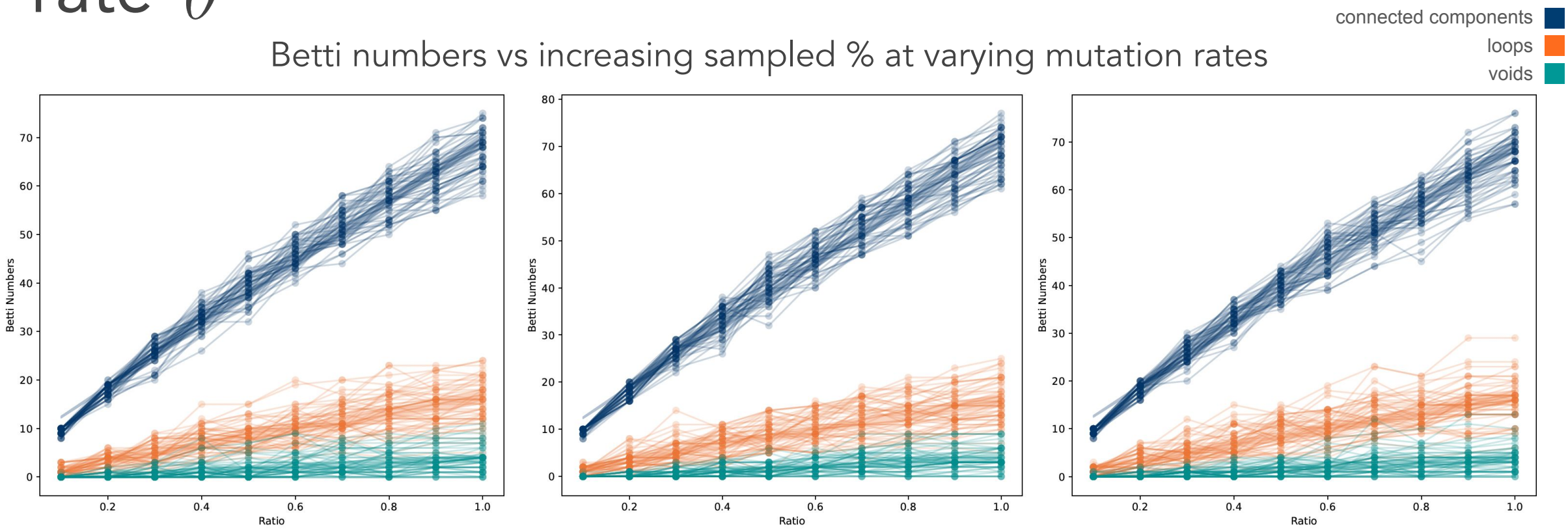


$\theta = 5000$

The mutation rate does not drive the changes in topology.

Sampling sparsity and the mutation rate θ

Betti numbers vs increasing sampled % at varying mutation rates



$\theta = 50$ $\theta = 500$ $\theta = 5000$

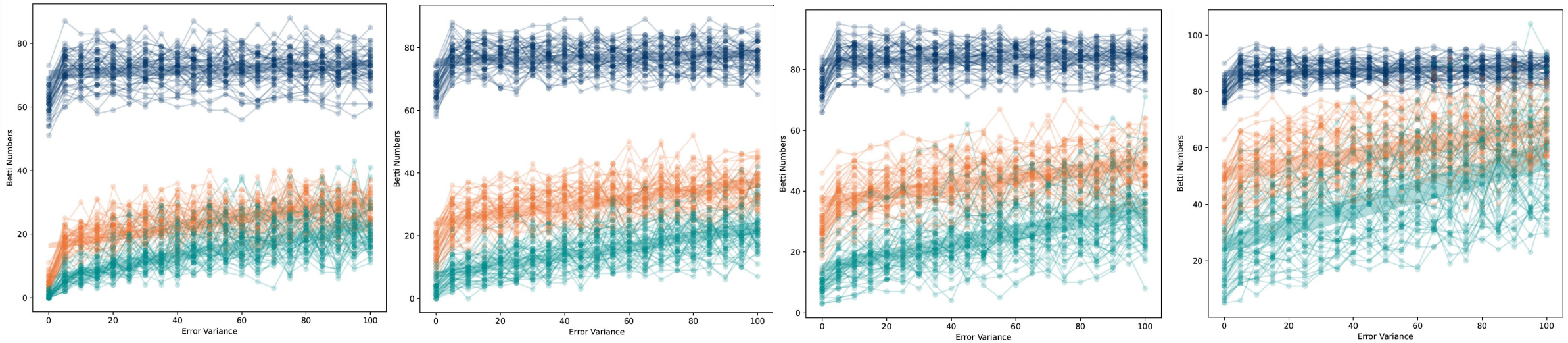
The mutation rate does not drive the changes in topology.

Noise and the recombination rate ρ

TopONE

connected components
loops
voids

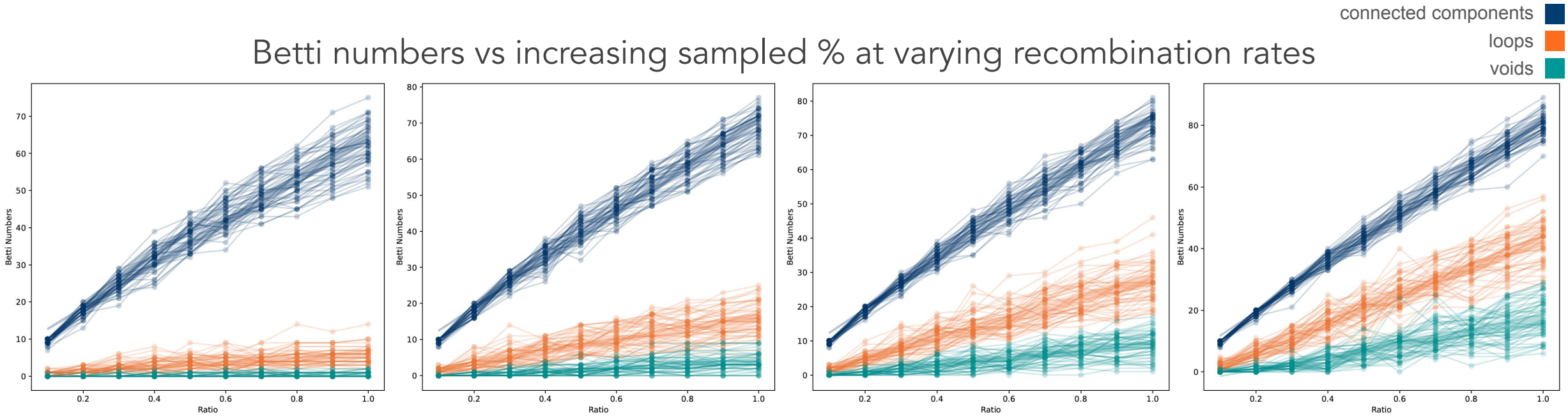
Betti numbers vs increasing noise at varying recombination rates



$\rho = 12$ $\rho = 36$ $\rho = 72$ $\rho = 144$

Increased recombinations in data drives the change in topology.

Sampling sparsity and the recombination rate ρ



$\rho = 12$ $\rho = 36$ $\rho = 72$ $\rho = 144$

Increased recombinations in data drives the change in topology.

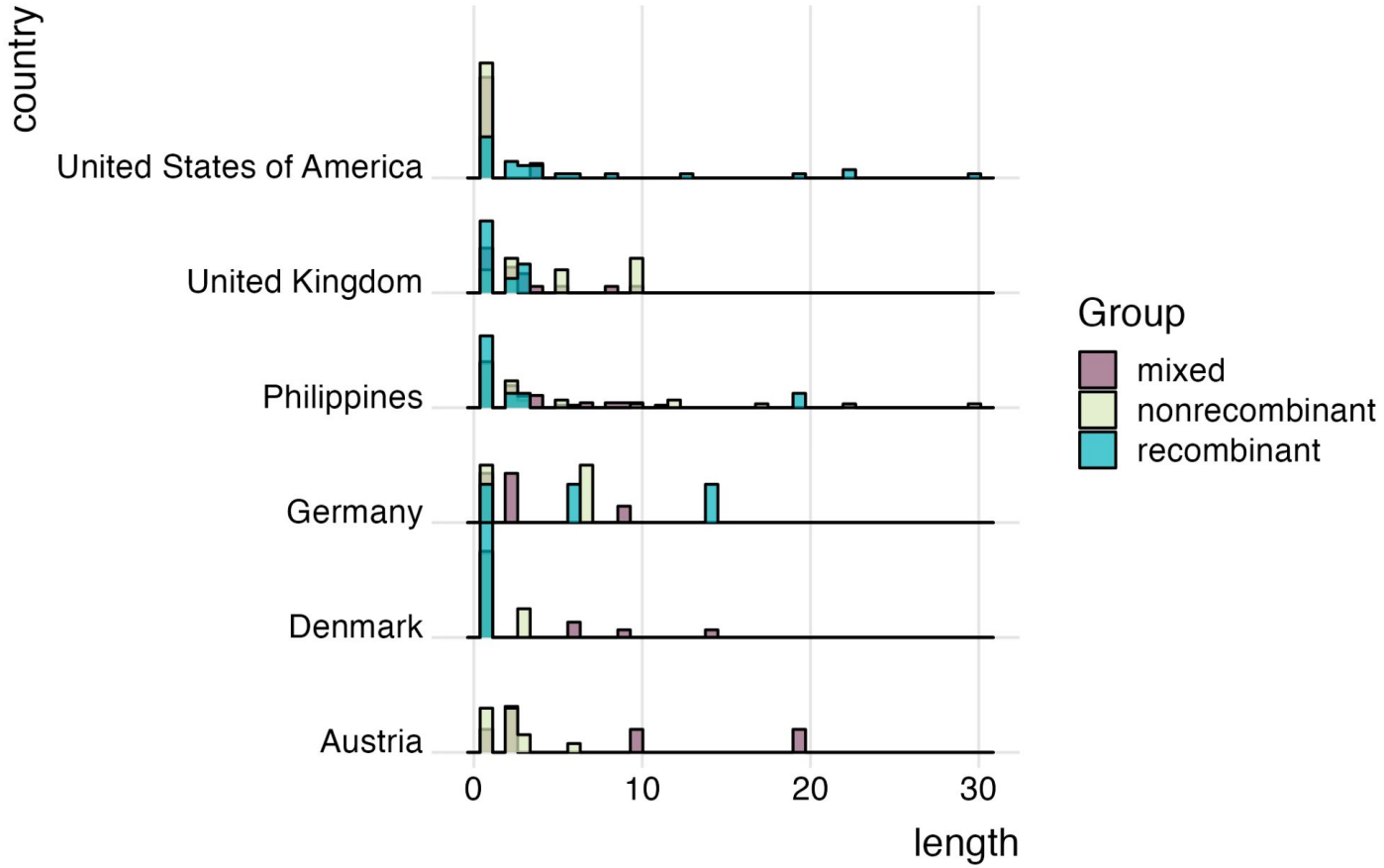
TopONE

Statistical Analysis

SARS-CoV-2 GISAID Samples

Persistence Distributions

Barcode Lengths of Persistence Diagrams



High-recombinant yield samples

Hypothesis Testing

Is there a significant difference in the birth and death times distributions between different group types in each country?

	XBC.1	XBE		XBZ
***p*-values**	Philippines	UK	USA	Denmark
Birth Times	0.003	0.014	0.019	0.321
Death Times	0.003	0.015	0.023	0.298

Results from Kruskal-Wallis Test and Independent Mann-Whitney U Test at 5% level of significance

Significant

Not significant

High-recombinant yield samples

Post-hoc Analysis

Which group type pairs are significantly different?

	XBC.1		XBE				XBZ	
***p*-values**	Philippines		UK		USA		Denmark	
	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths
Mixed vs Non-recombinant	0.002	0.013	0.059	0.057	1.000	1.000	1.000	0.937
Mixed vs Recombinant	1.000	1.000	1.000	1.000	0.017	0.020	0.491	0.472
Recombinant vs Non-recombinant	0.484	0.478	0.019	0.021	1.000	1.000	1.000	1.000

Significant

Not significant

Results taken at 5% level of significance

Low-recombinant yield samples

Hypothesis Testing

Is there a significant difference in the birth and death times distributions between different group types in each country?

	XBC.1				XBZ	
***p*-values**	China	Singapore	S. Korea	USA	Austria	Germany
Birth Times	0.969	0.300	0.008	0.361	0.921	0.297
Death Times	0.958	0.398	0.023	0.407	0.729	0.110

Results from Kruskal-Wallis Test and Independent Mann-Whitney U Test at 5% level of significance

Significant

Not significant

Low-recombinant yield samples

Post-hoc Analysis

Which gene type pairs are significantly different?

	XBC.1								XBZ			
***p*-values**	China		Singapore		S. Korea		USA		Austria		Germany	
	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths
Mixed vs Non-recombinant	0.969	0.958	0.300	0.398			0.361	0.407	0.921	0.729	0.724	0.199
Mixed vs Recombinant					0.008	0.023					0.578	0.433
Recombinant vs Non-recombinant											1.000	1.000

Significant

Not significant

Results taken at 5% level of significance

Summary

Our research questions

TopONE

? Is the data topology robust to noise and sparse genomic samples?

- Increasing noise creates new loops
- Increased sampling builds more representative data

? Does recombination change the data topology of genomic samples?

- Yes but subject to further investigation

Can topology detect recombination?

Potentially so. Further analysis is recommended for more real-world samples.

TopONE

Thank you!