

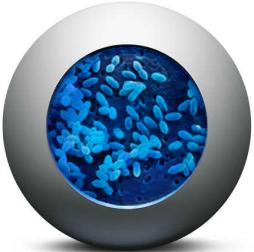
Spatiotemporal dynamics of *Streptococcus pneumoniae*

Sophie Belman

Supervisors: Stephen Bentley & Henrik Salje



Pathogen
Dynamics group
UNIVERSITY OF CAMBRIDGE



Presentation outline

1) Background of *S. pneumoniae*

- Life history
- Mortality
- Carriage and transmission
- Prevention

2) Aims of PhD

3) Spatiotemporal analysis results

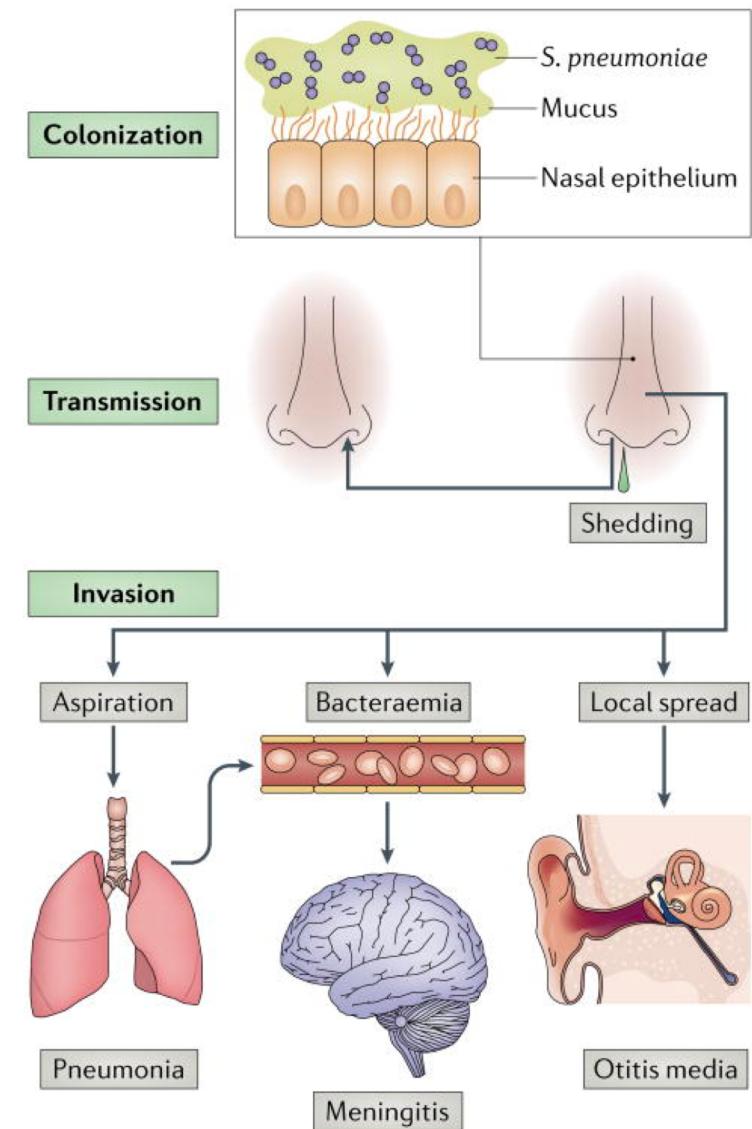


Streptococcus pneumoniae life history

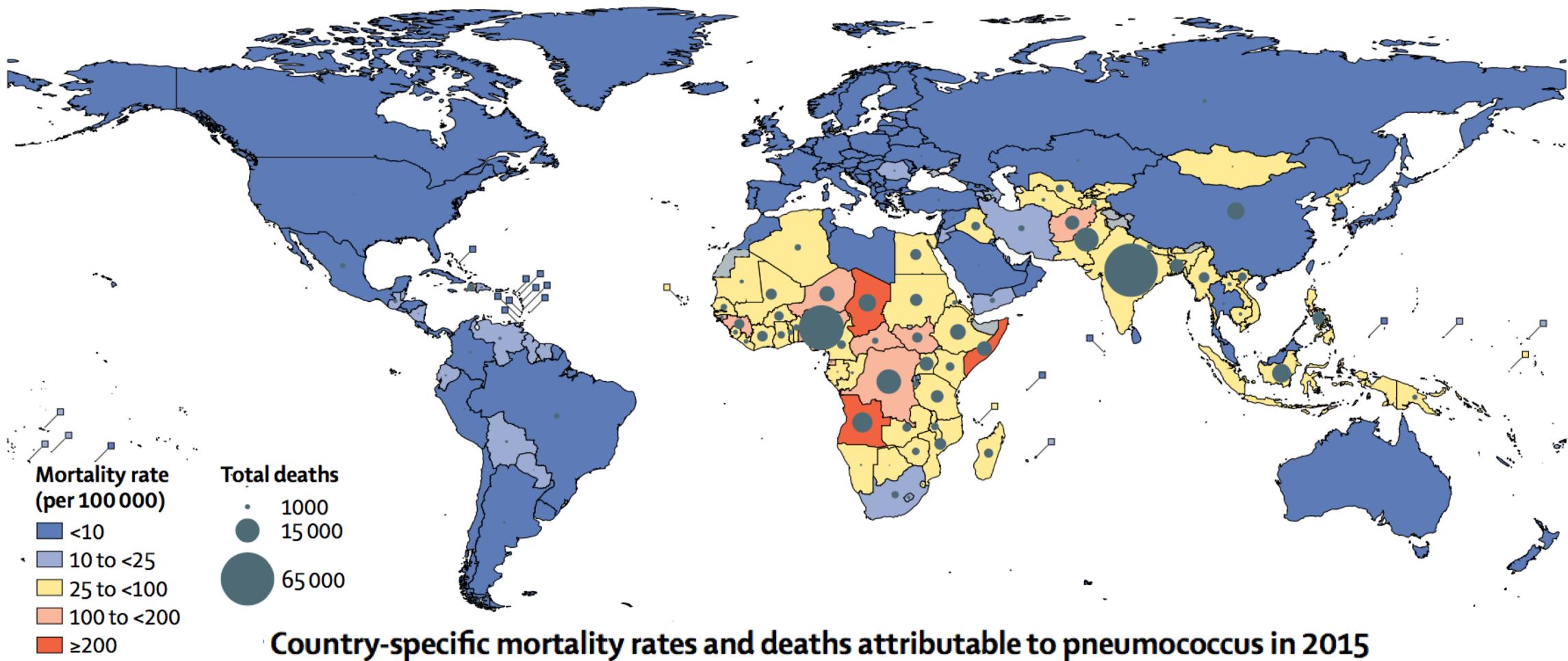
Niche: Opportunistic, human obligate bacteria. Inhabits nasopharyngeal passage of both children and adults.

Disease: Ranges from asymptomatic to severe.

Toll: Responsible for half a million deaths of children <5 every year.
Primary agent responsible for pneumonia deaths. 55.5% of deaths from LRTI attributed to *S. pneumoniae*.

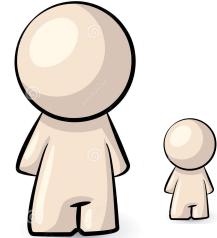


Global mortality corresponds with carriage rate distribution

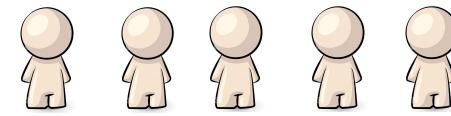


Transmission occurs via respiratory droplets

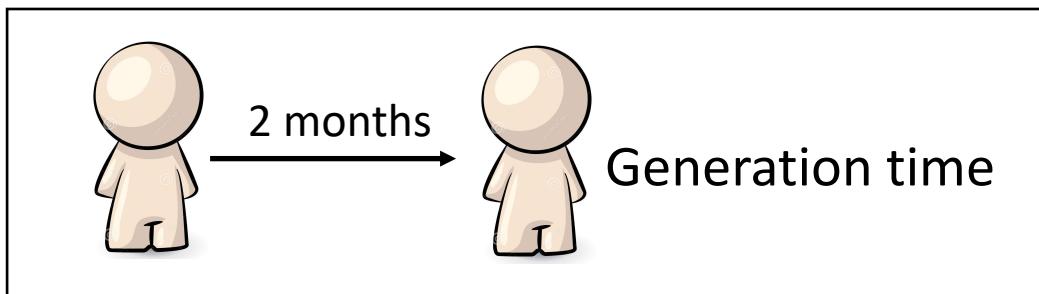
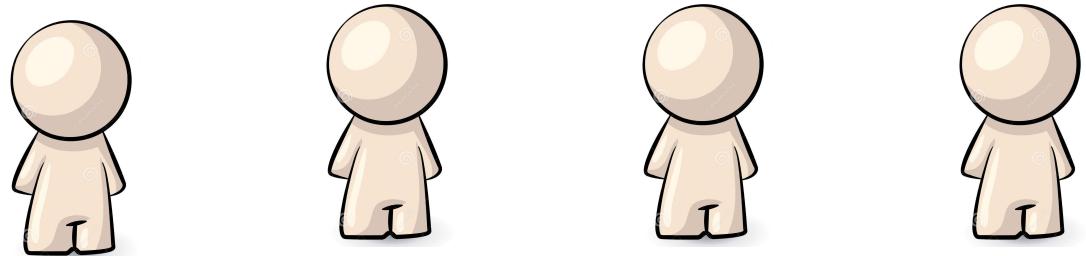
Mother to child transmission



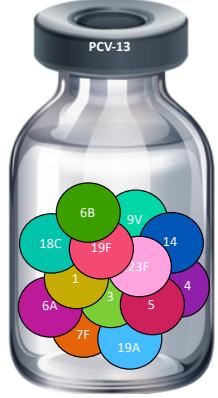
Child-to-child transmission



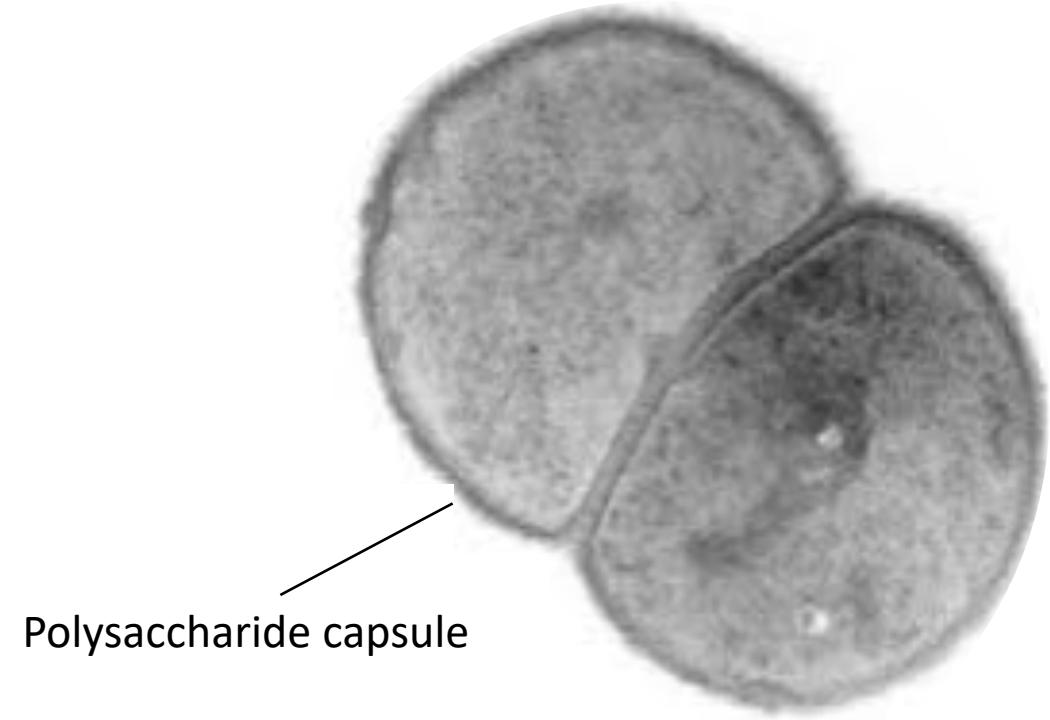
Adult role in transmission (?)



Pneumococcal conjugate vaccine protects against 13 serotypes

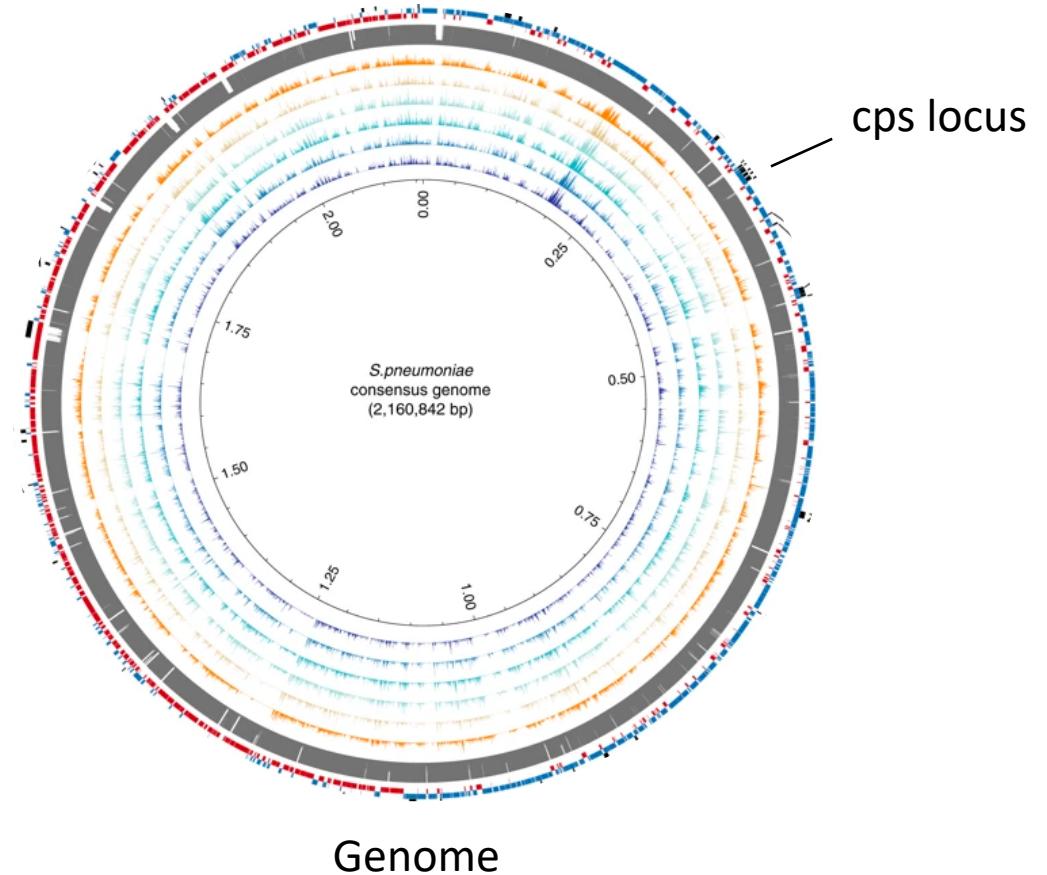


- *Streptococcus pneumoniae* comprises 100 serotypes (cps)



Polysaccharide capsule

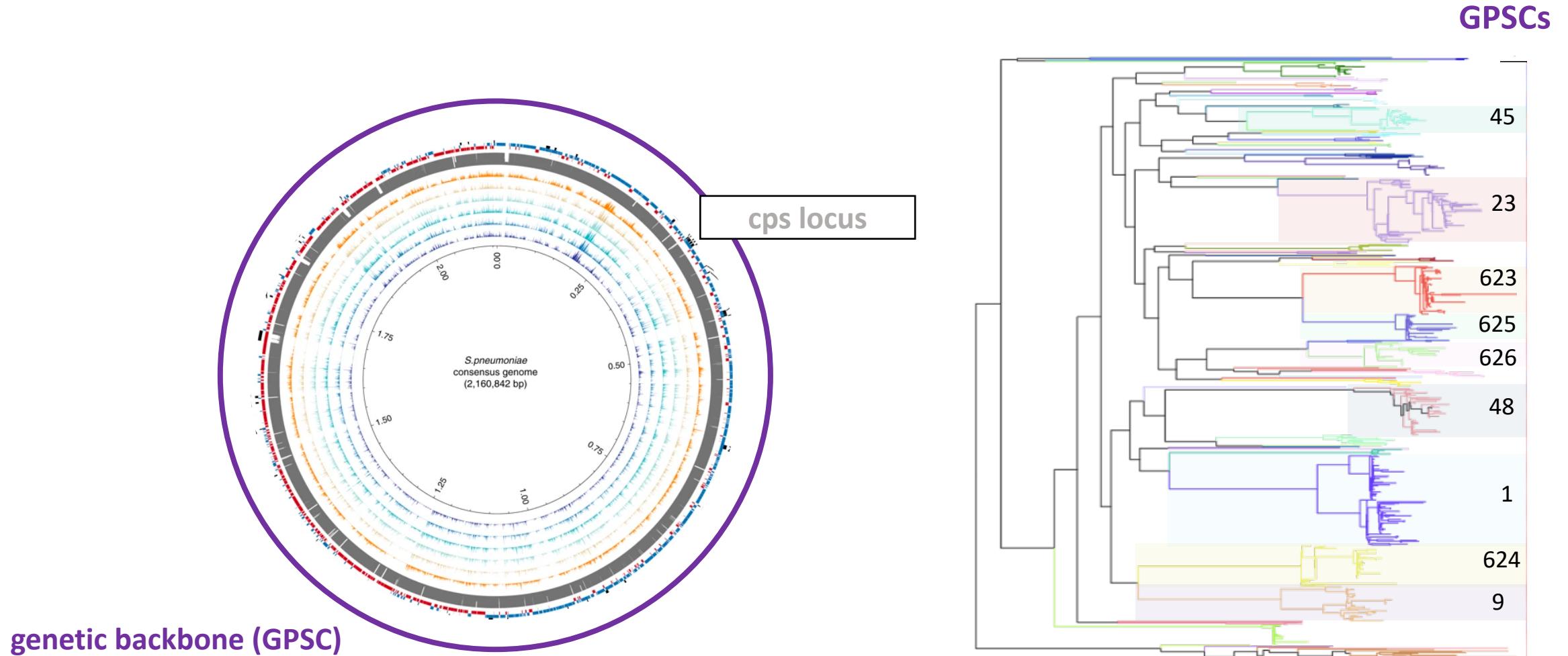
Pair of diplococci



Genome

Highly diverse – clustered into over 800 different sequence clusters

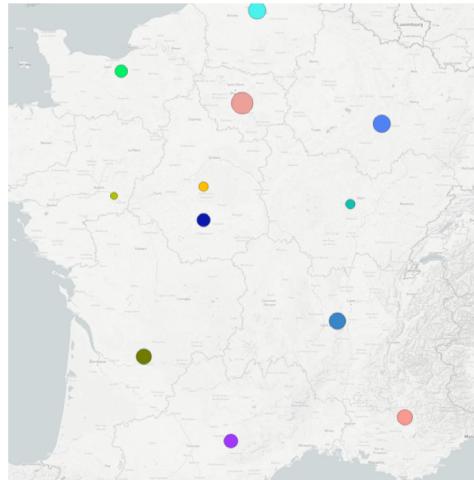
Serotypes (*cps*) are on over 800 different genetic backbones (GPSCs)



The extreme diversity has masked the dynamics of geographic spread

- What about migration? – How quickly? Which routes?

Within Region – Between Region



France
1000km



South Africa
2000km

Between Countries



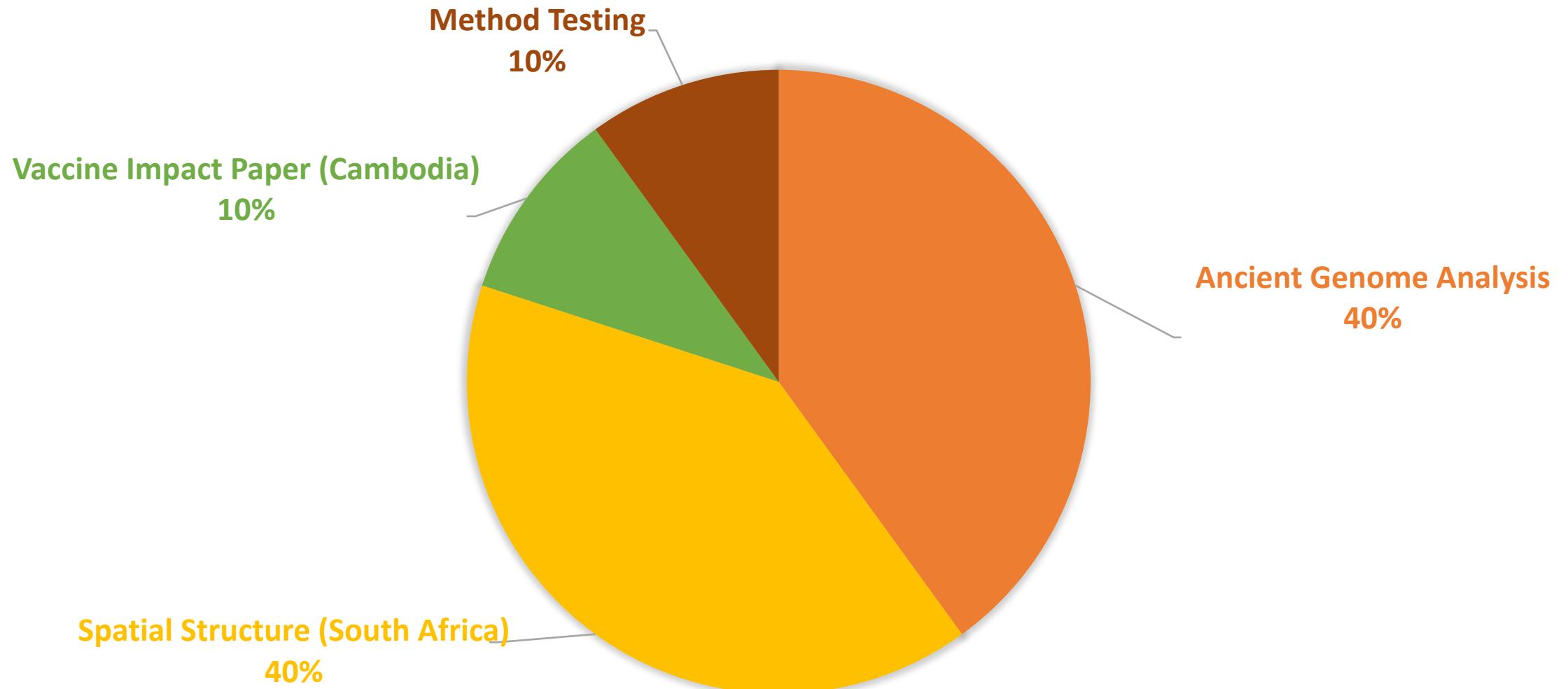
Aims of my PhD

- Describing the spatial structure and rate of pneumococcal spread.
- Obtain a mechanistic understanding of pneumococcal migration.
- Reconstructing previous migration based on extant diversity.

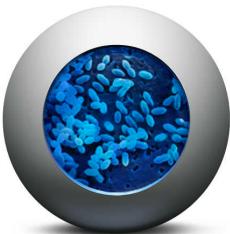
Aims of my first year

- Describing the spatial structure and rate of pneumococcal spread.
- Obtain a mechanistic understanding of pneumococcal migration.
- Reconstructing previous migration based on extant diversity.

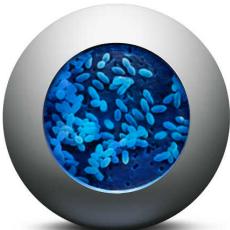
What I've actually done this year



Global Pneumococcal Sequencing project (GPS)



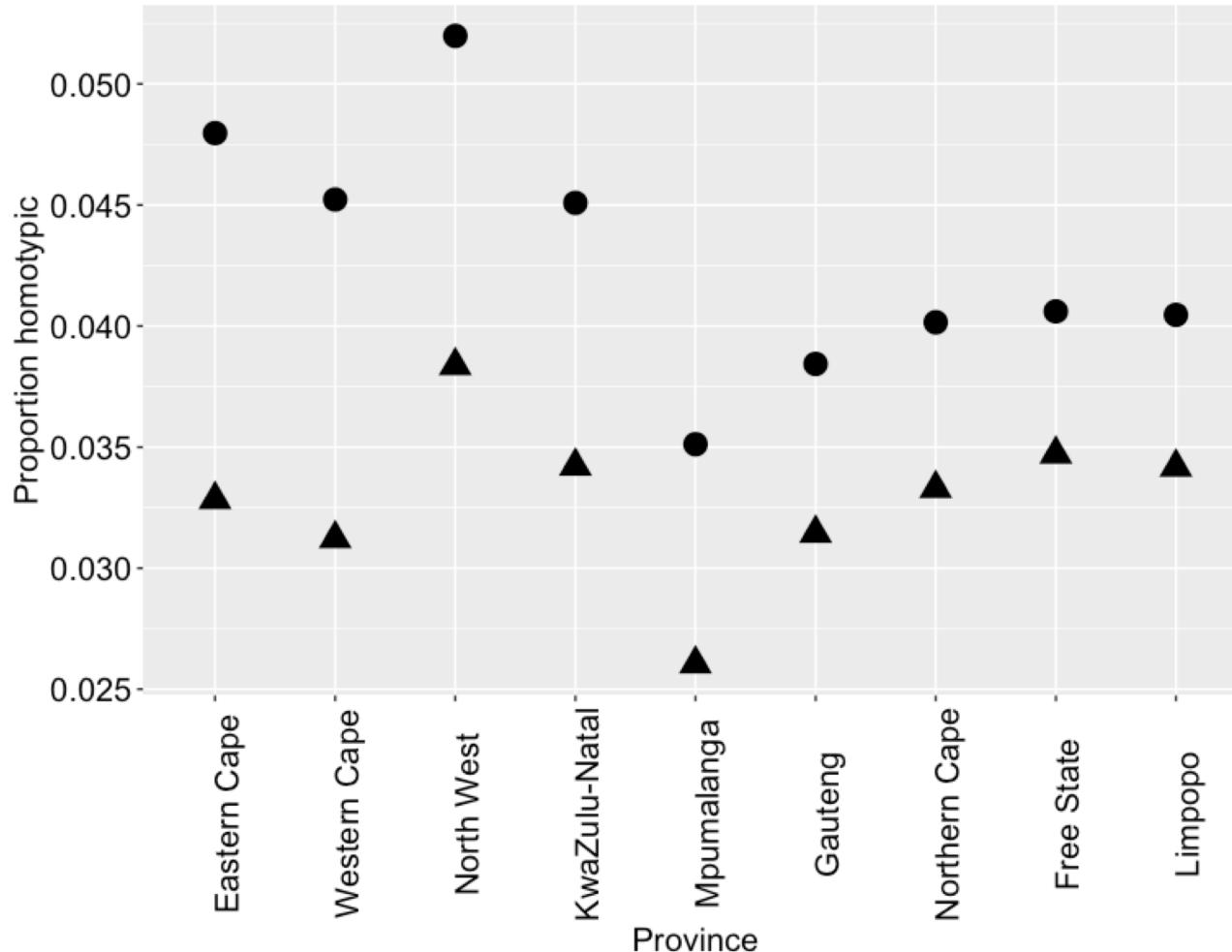
6920 pneumococcal genomes collected from South Africa (GPS)



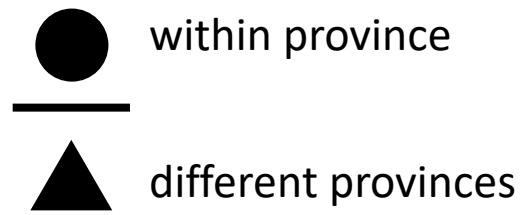
Province	Capitals (other population hubs)	N	GPSCs (N)
Gauteng	Johannesburg (Soweto, Pretoria)	3171	157
Mpumalanga	Nelspruit (Agincourt)	1308	115
Western Cape	Cape Town	880	91
KwaZulu-Natal	Pietermaritzburg (Durban)	649	81
Free State	Bloemfontein	347	66
Eastern Cape	Bhisho (Port Elizabeth)	300	61
North West	Mahikeng	116	42
Northern Cape	Kimberley	80	33
Limpopo	Polokwane	69	38

Is there spatial structure among GPSCs in South Africa?

More homotypic strains within provinces than between provinces

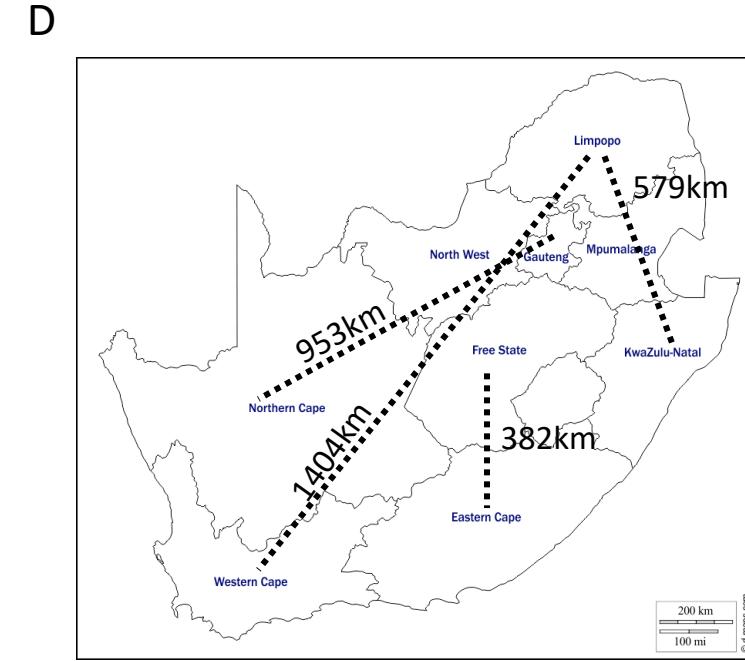
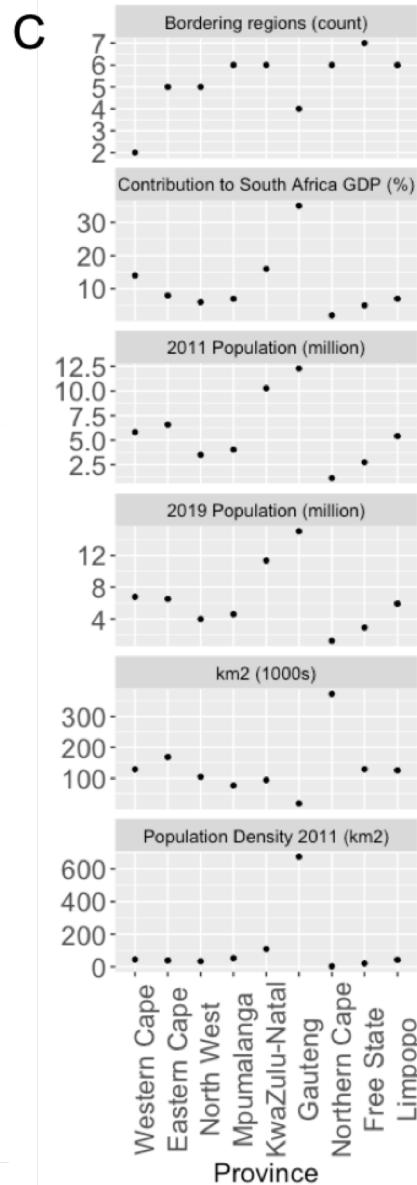
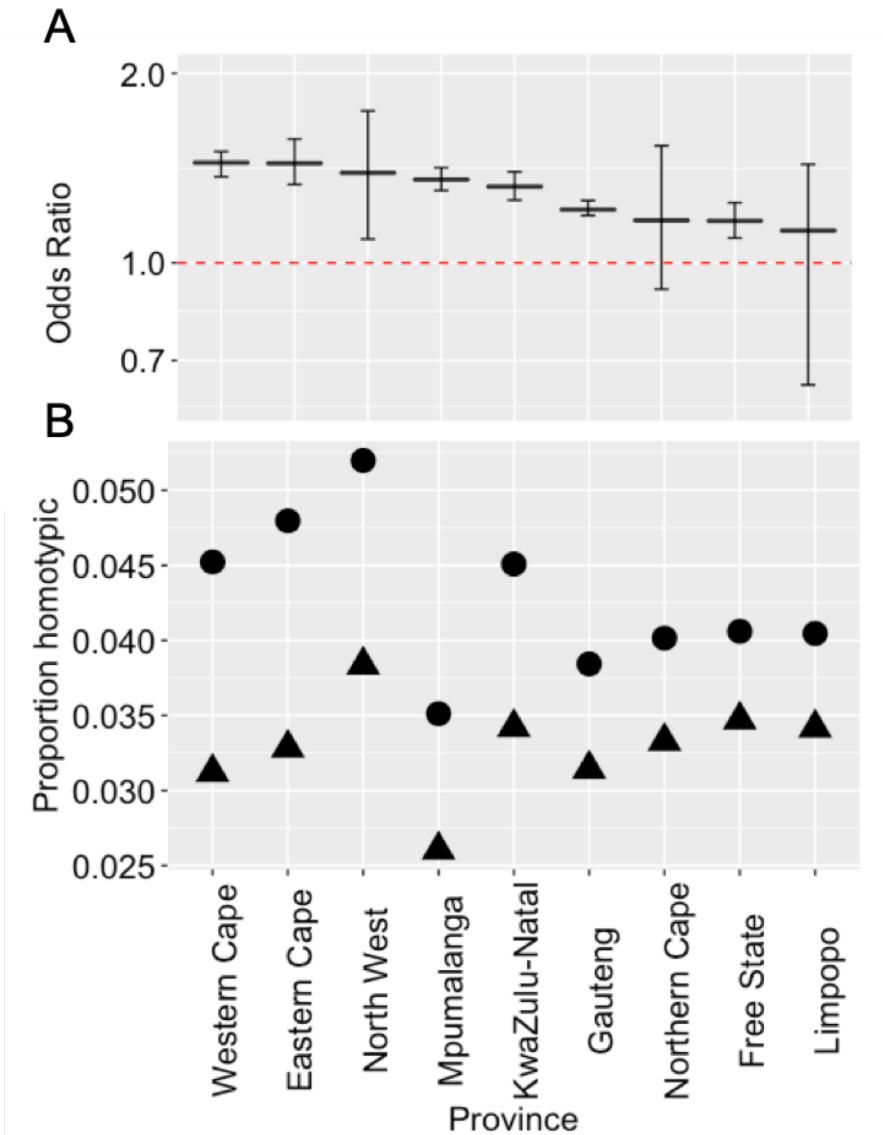


$$\frac{\frac{(r * t) * s}{r * t}}{\frac{((1 - r) * t) * s}{(1 - r) * t}}$$

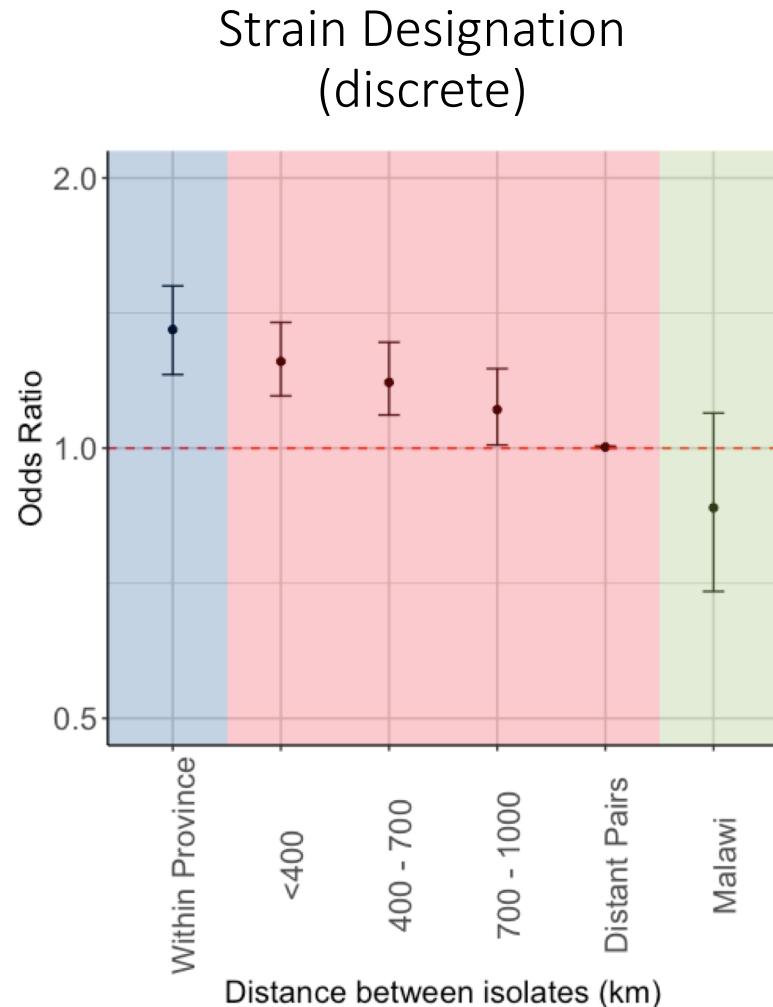


*Same Province = r
Same Time = t
Same Strain = s*

More homotypic strains within provinces than between provinces

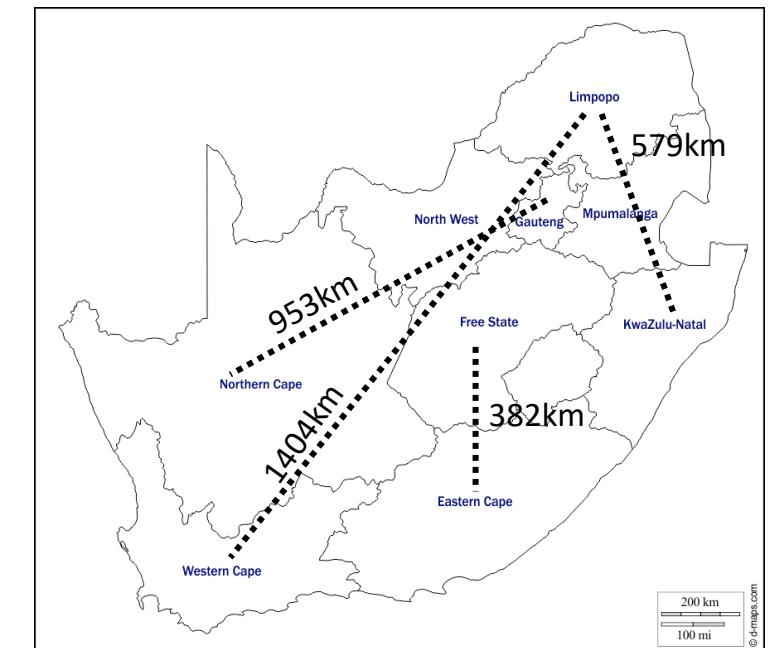
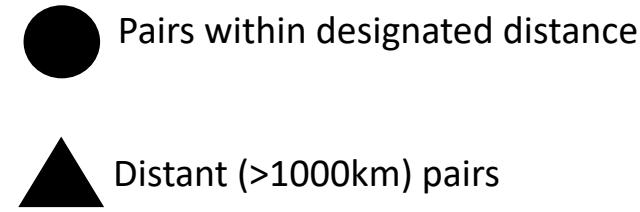


Odds of similarity across geographic distances compared with distant pairs

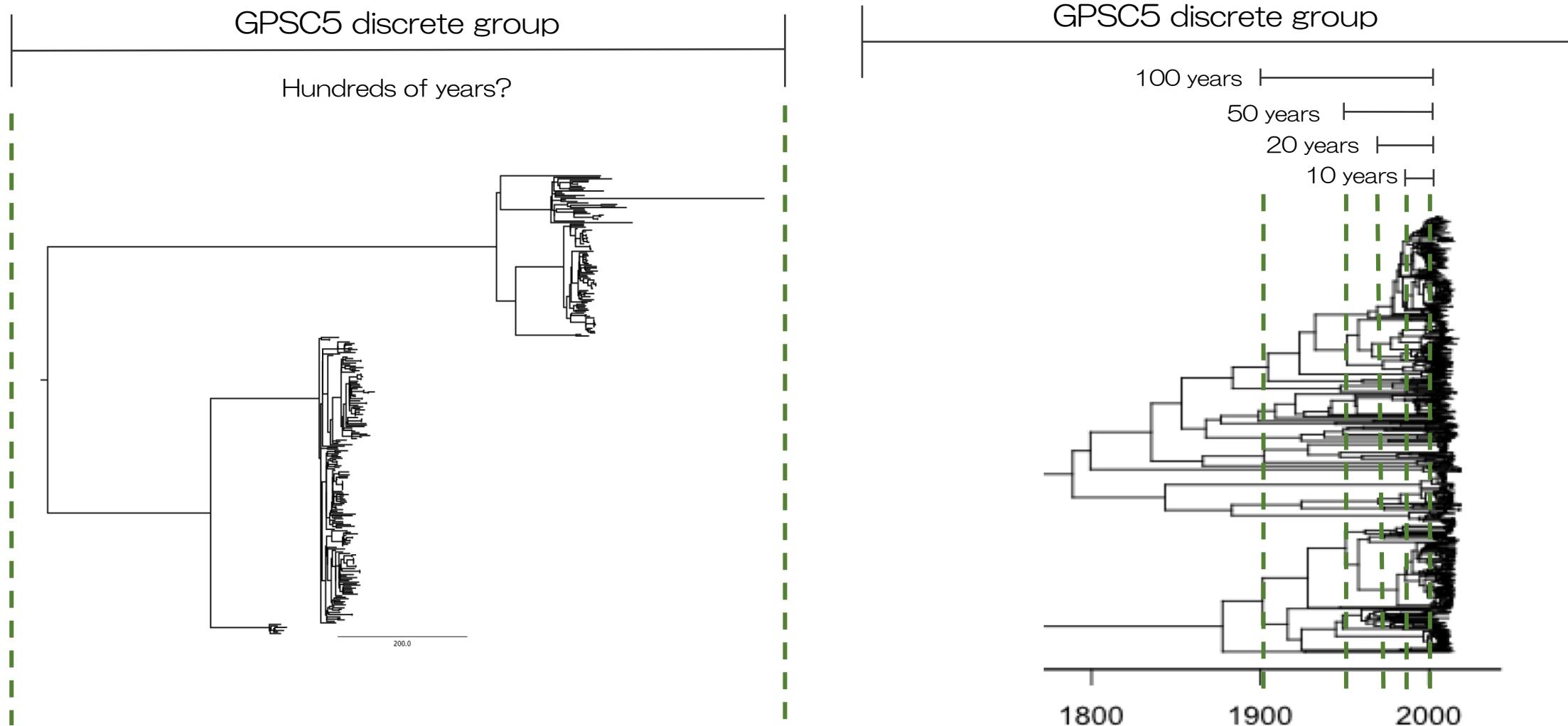


$$\frac{(r * t) * s}{r * t}$$
$$\frac{(r * t) * s}{r * t}$$

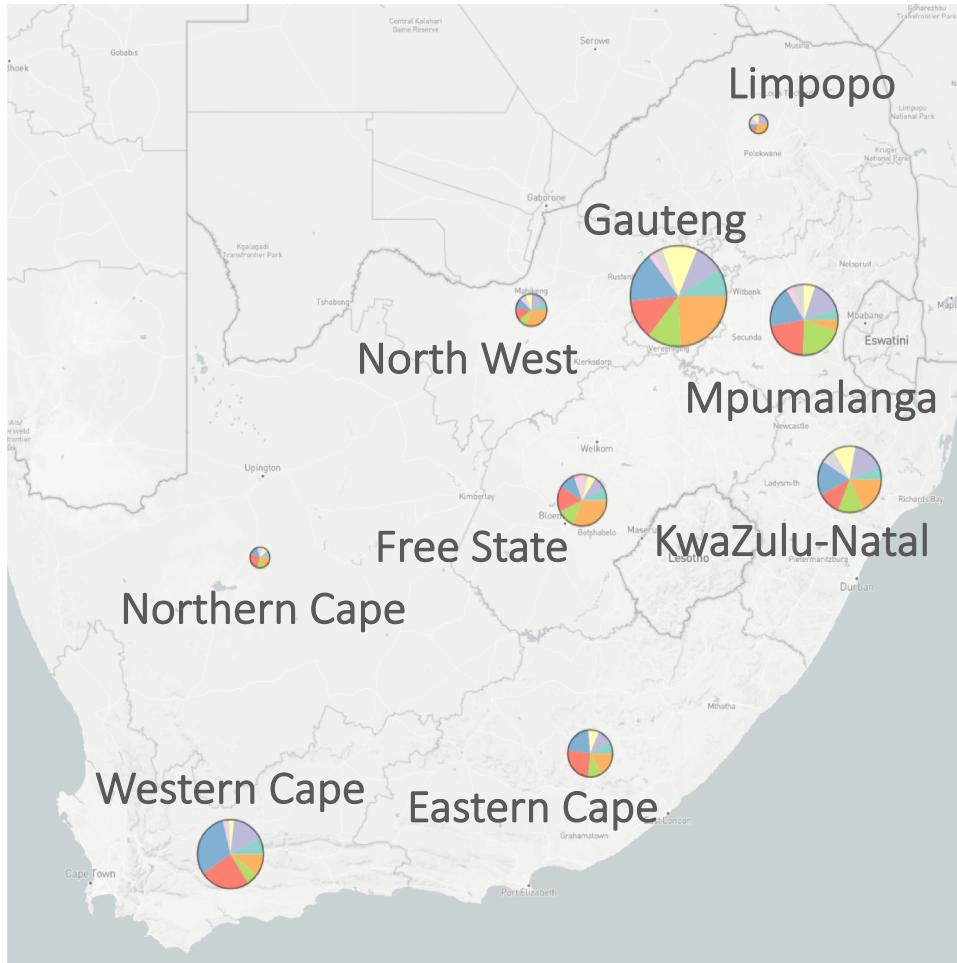
Distance = r
Same Time = t
Same Strain = s



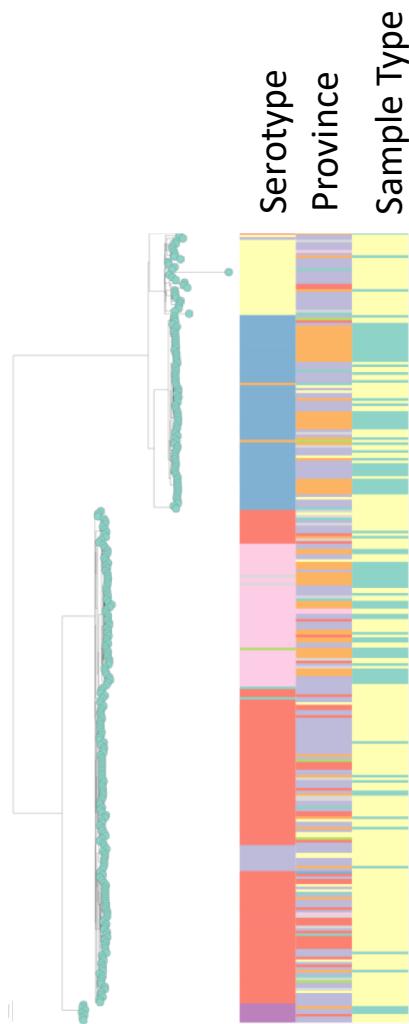
Time resolved phylogenies improve resolution



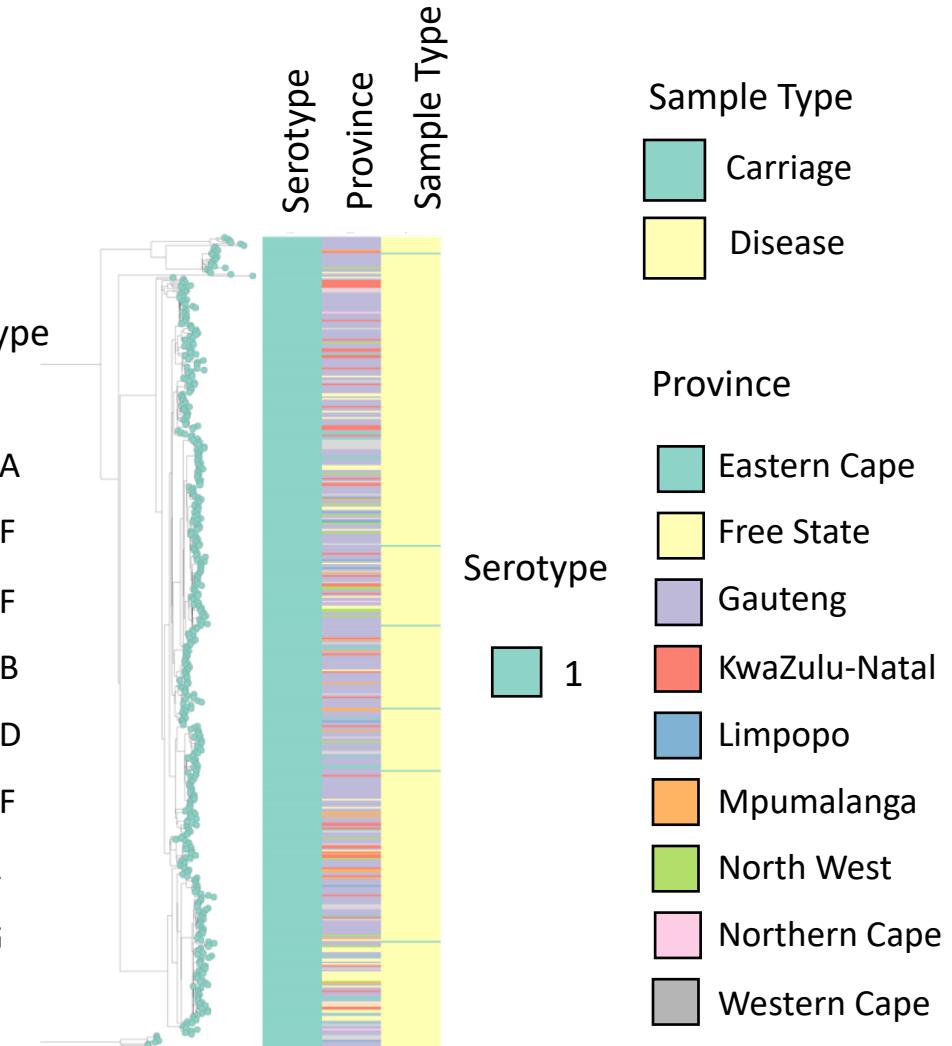
9 GPSCs from all 9 Provinces



GPSC5

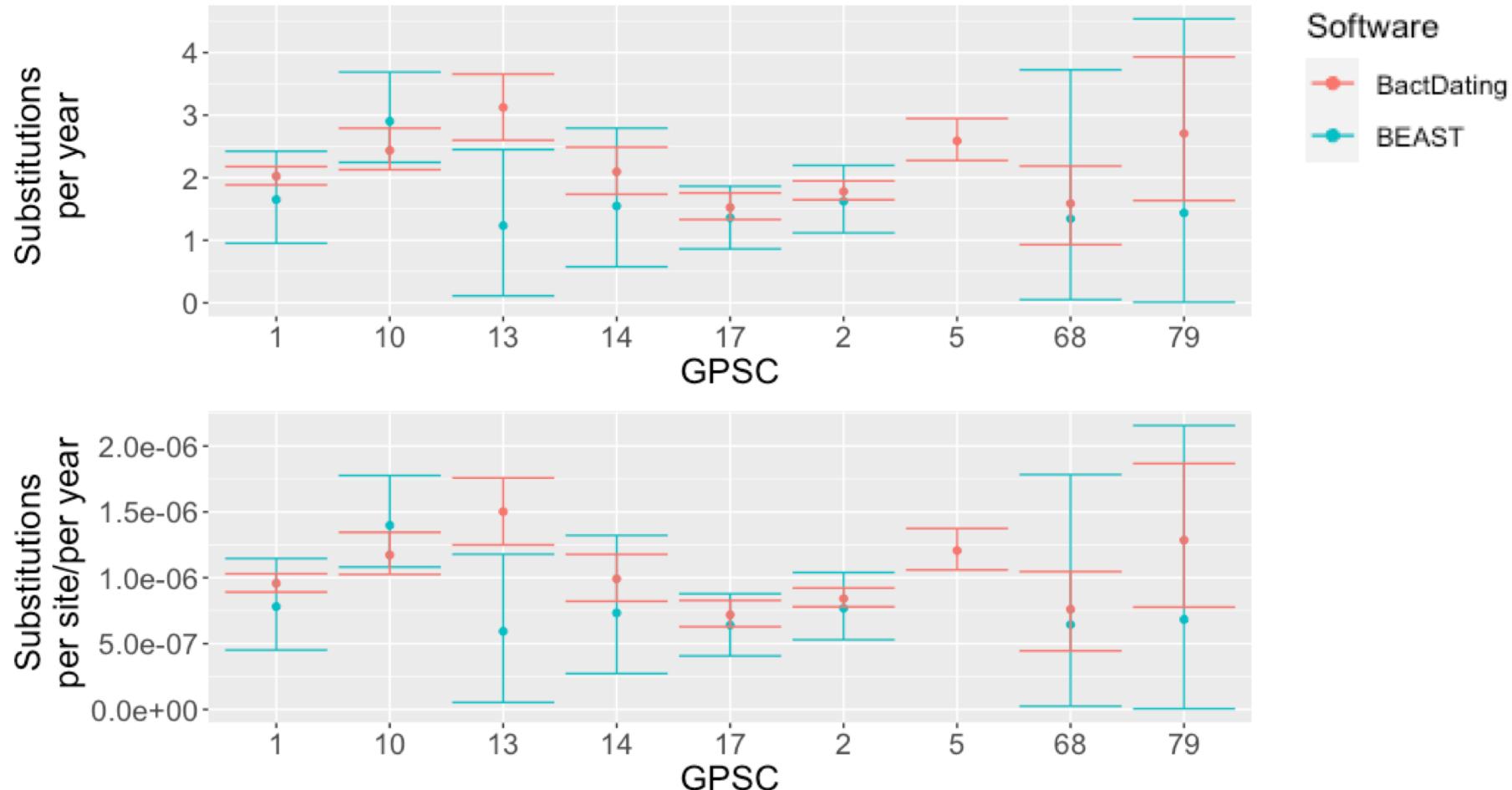


GPSC2

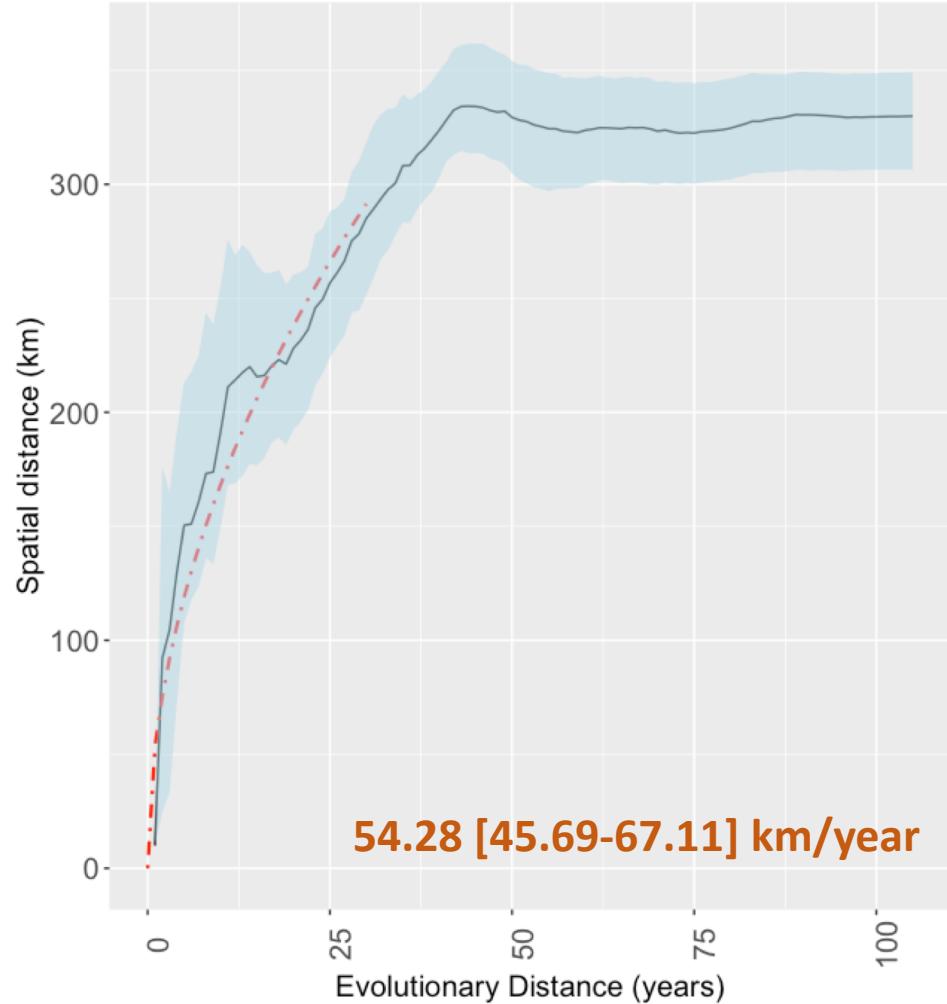


Are different methods concordant – can I trust my divergence times?

BactDating vs. BEAST



Rate of geographic spread (migration) in South Africa



Grey dashed line is the mean distance across South Africa.

Light blue band is 95% confidence intervals.

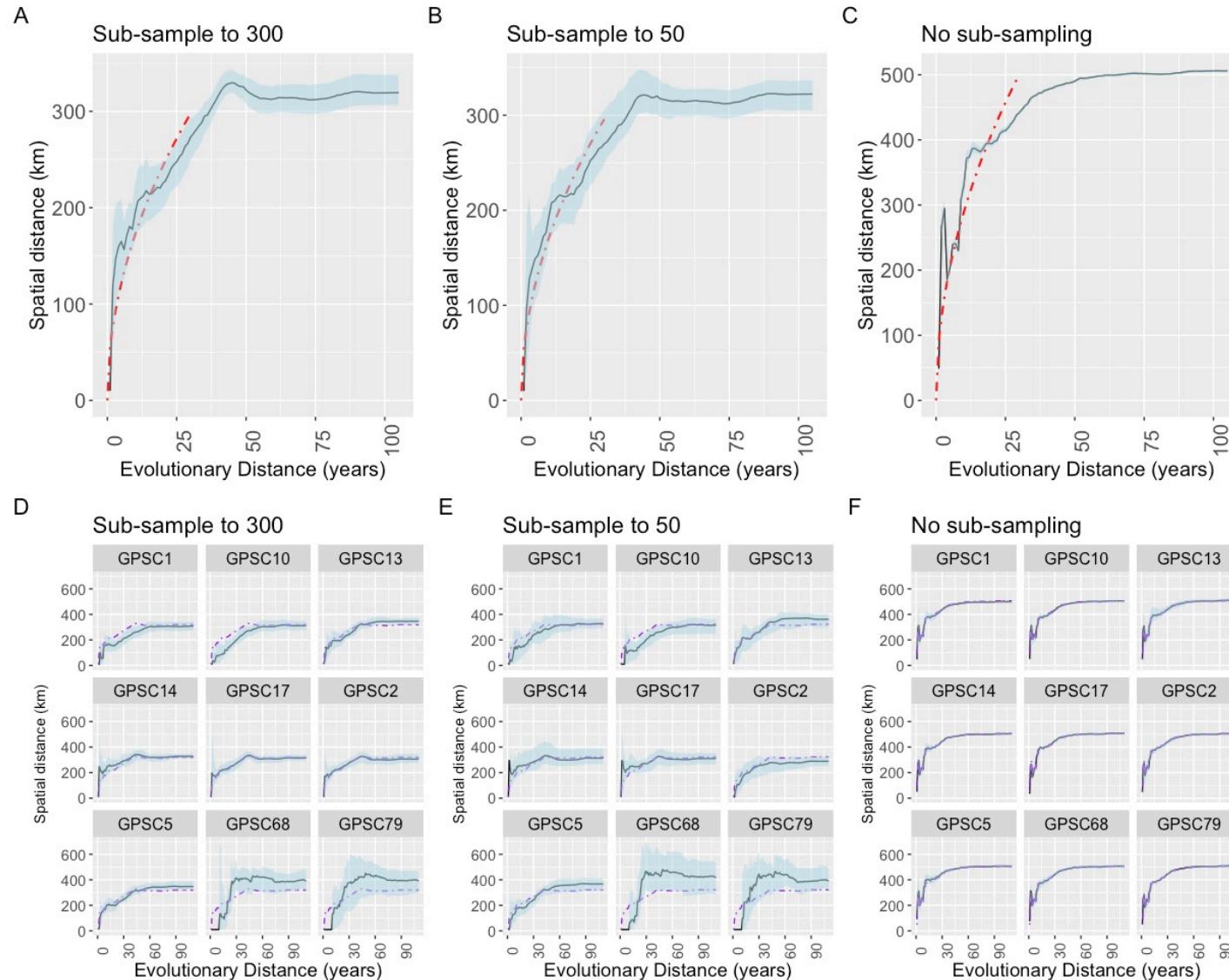
Orange annotation is mean (μ_k) rate of transmission across South Africa.

Red dot-dash line is the μ_a over evolutionary time truncated at a μ_k calculated within 10 years genetic distance.

Sub-sampling to 50 per GPSC in each region without replacement.

Cumulative.

Sub-sampling to compensate for sampling bias

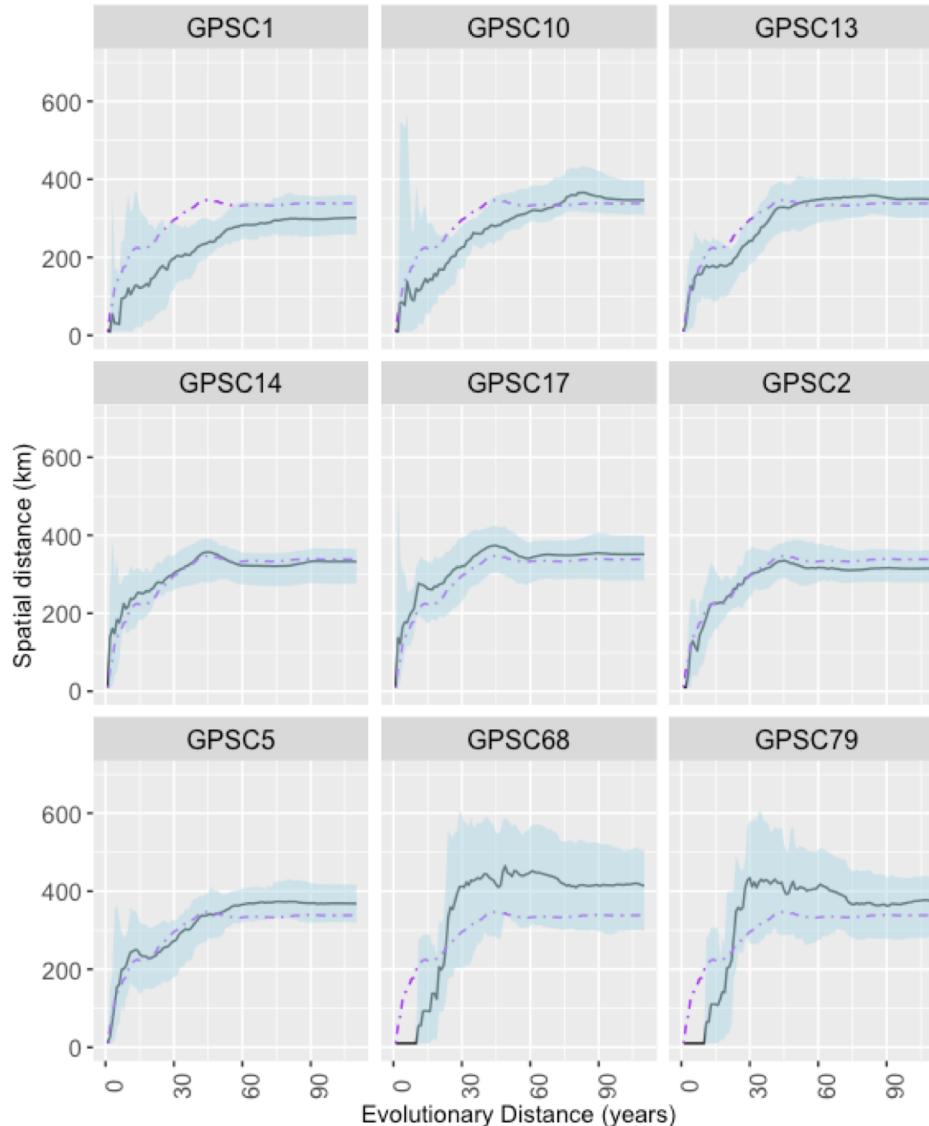


Rates for individual GPSCs are similar to overall

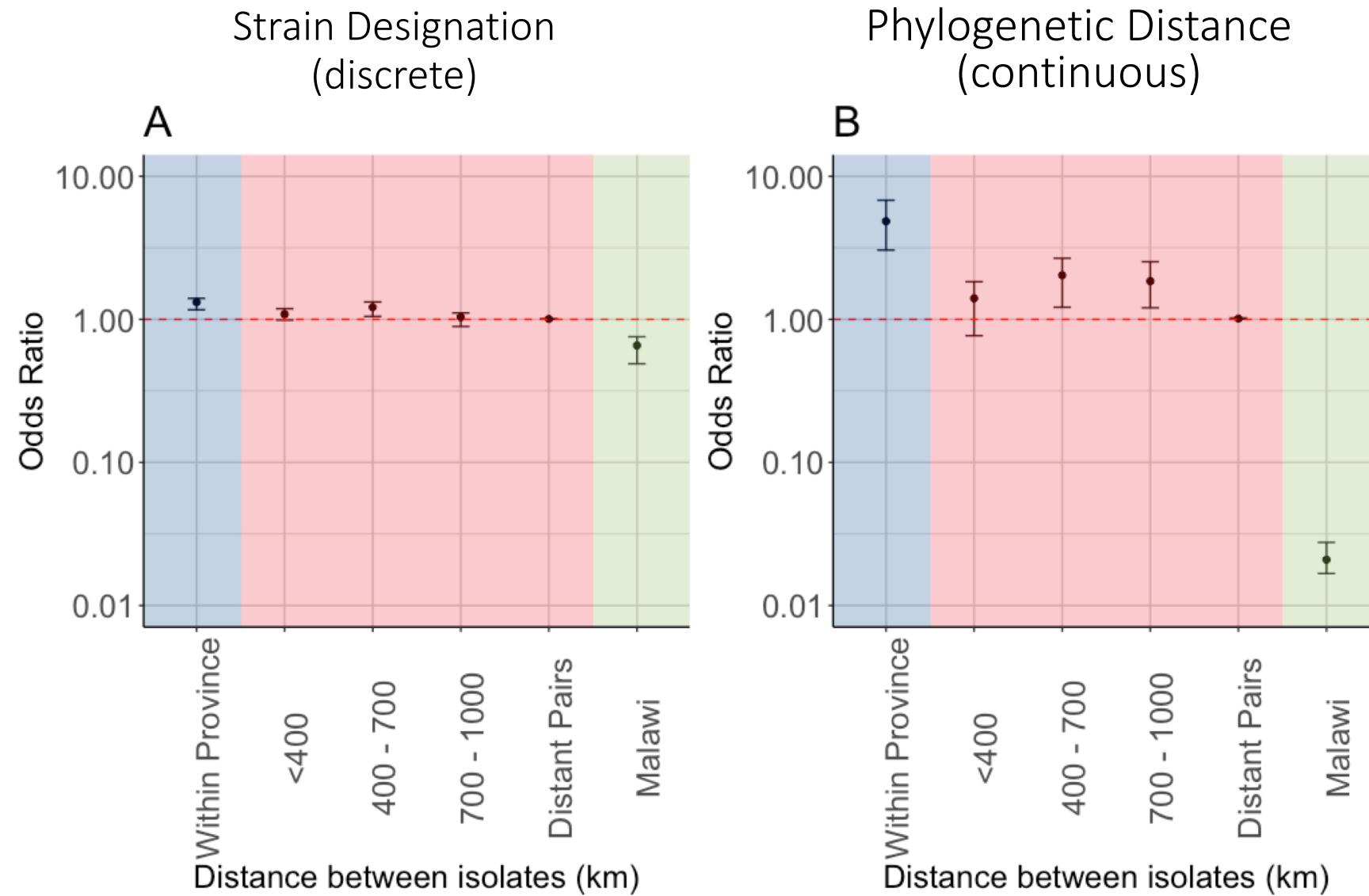
Purple dashed line is overall.

Light blue band is 95% confidence intervals.

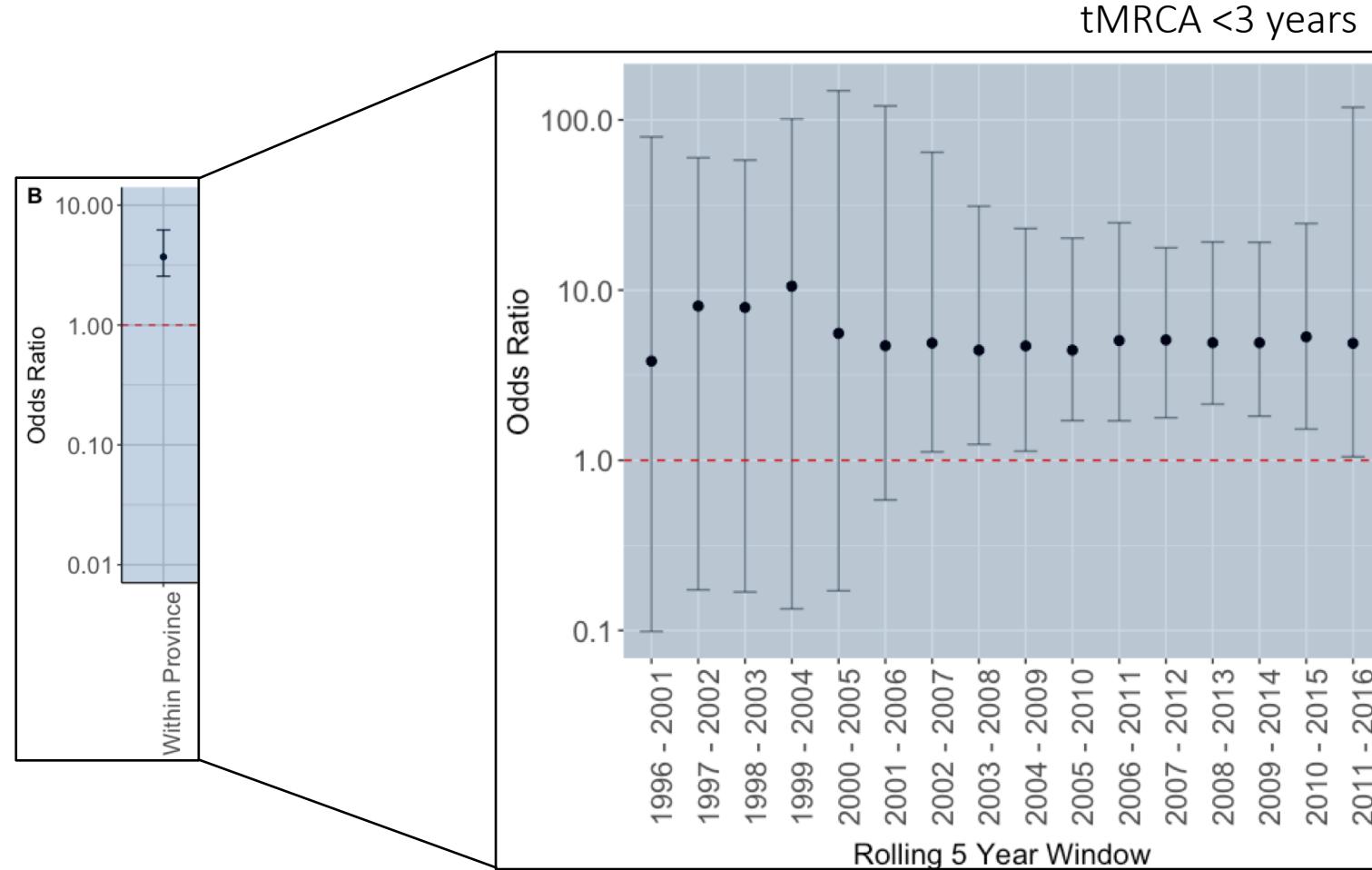
Sub-sampling to 50 per GPSC in each region without replacement. Cumulative over time.



Odds of similarity across geographic distances compared with distant pairs

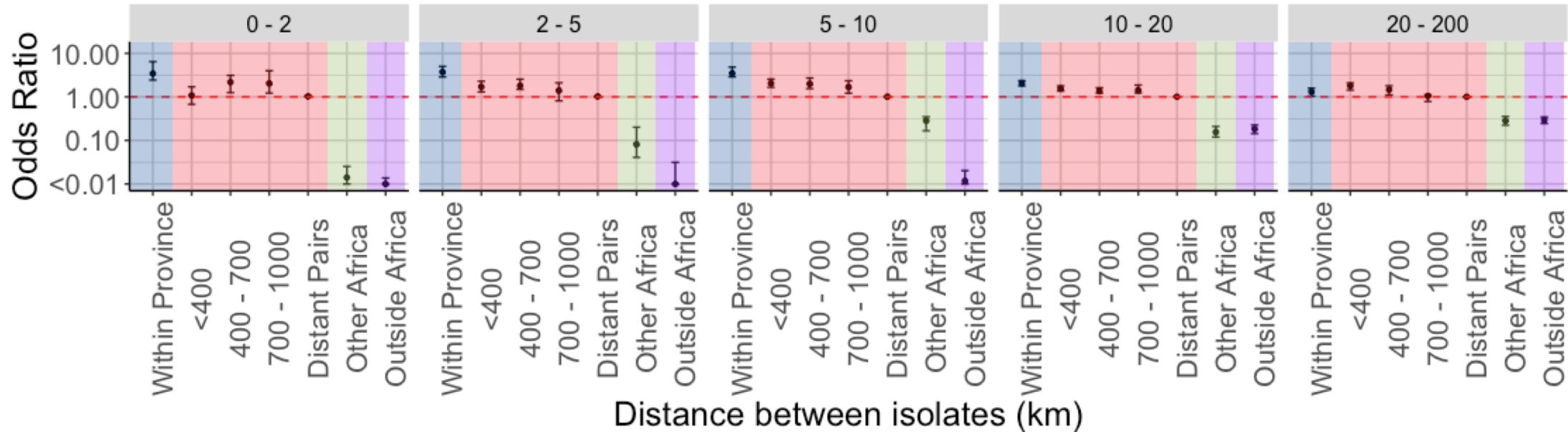


No change in spatial structure from 1996-2014

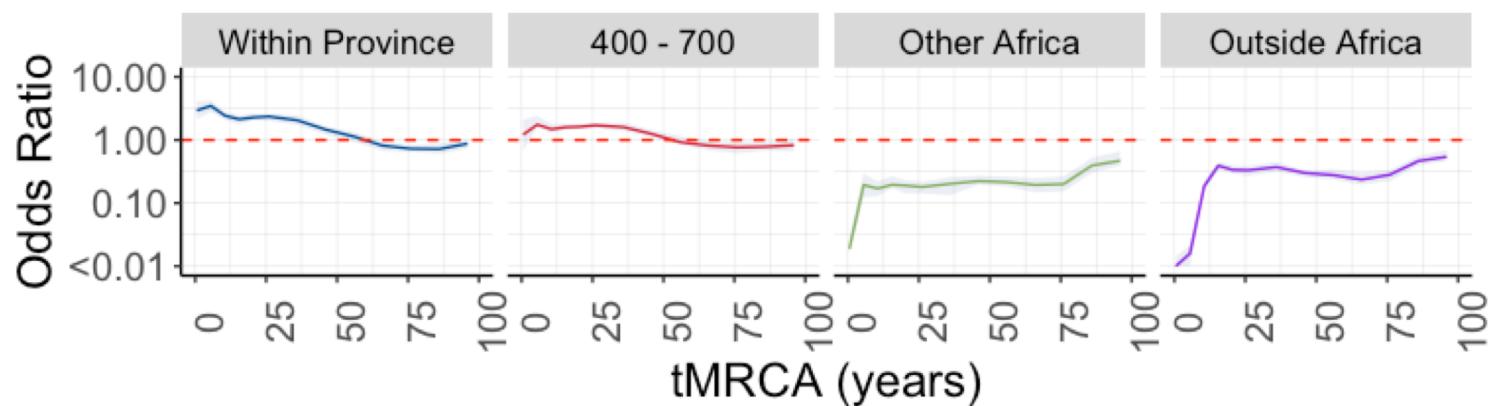


50 years for strains to become homogenous across South Africa

A



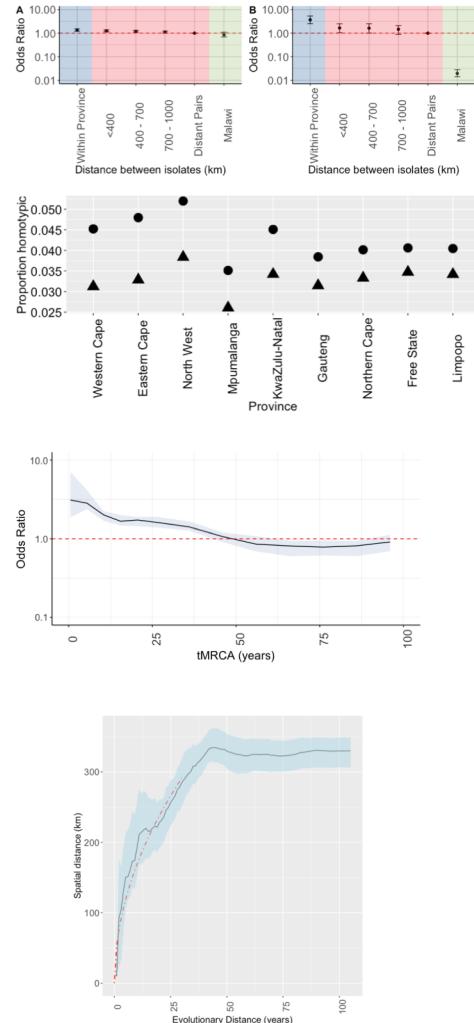
B



Conclusions

Spatial structure, with distinct populations circulating in different provinces, exists.

- Strain clustering (GPSCs) is able to elucidate spatial dynamics.
- Time resolved WGS phylogenies allow higher resolution investigation of spatial dynamics.



Streptococcus pneumoniae takes ~50 years to become homogenously mixed across South Africa.

A general estimate for the rate of geographic spread is ~50km/year utilizing a simple model.

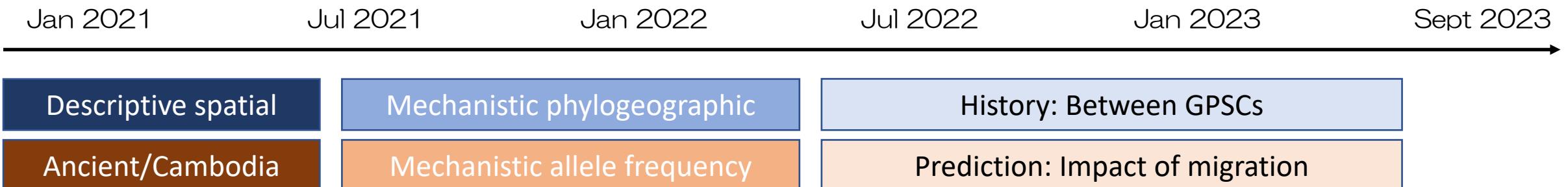
- There is no blatant difference between rate of geographic spread for different GPSCs.

Future directions

- Mechanistic explanations for movement
 - Facebook Data for Good to compensate for missing transmissions events.
 - Can we infer the true generation time – in two months where is the most probable location incorporating the movement data and the genetic data?
- Reconstructing previous migrations and their mechanisms
 - msprime and allele frequencies in different locations at different time points.
 - Python framework.

History: Reconstruct further history tMRCA between GPSCs – can I shift away from time resolution to determine this?

Prediction: Is it possible to predict the impact of pneumococcal migration from one population on a strains spread in another population? What about changes in human migration impact on pneumococcal migration?

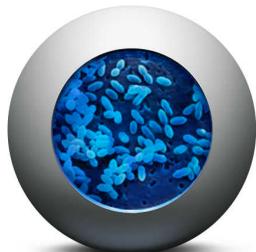


Acknowledgements

- Henrik Salje
 - Stephen Bentley
 - Pathogen Dynamics Group University of Cambridge
 - Team284 Wellcome Sanger Institute
 - Global Pneumococcal Sequencing Project
- ...and everyone else who's given input and helped me thus far.



Pathogen
Dynamics group
UNIVERSITY OF CAMBRIDGE



Which of these are newly expanded and which have long been endemic?

ARTICLE

<https://doi.org/10.1038/s41467-019-13549-9>

OPEN

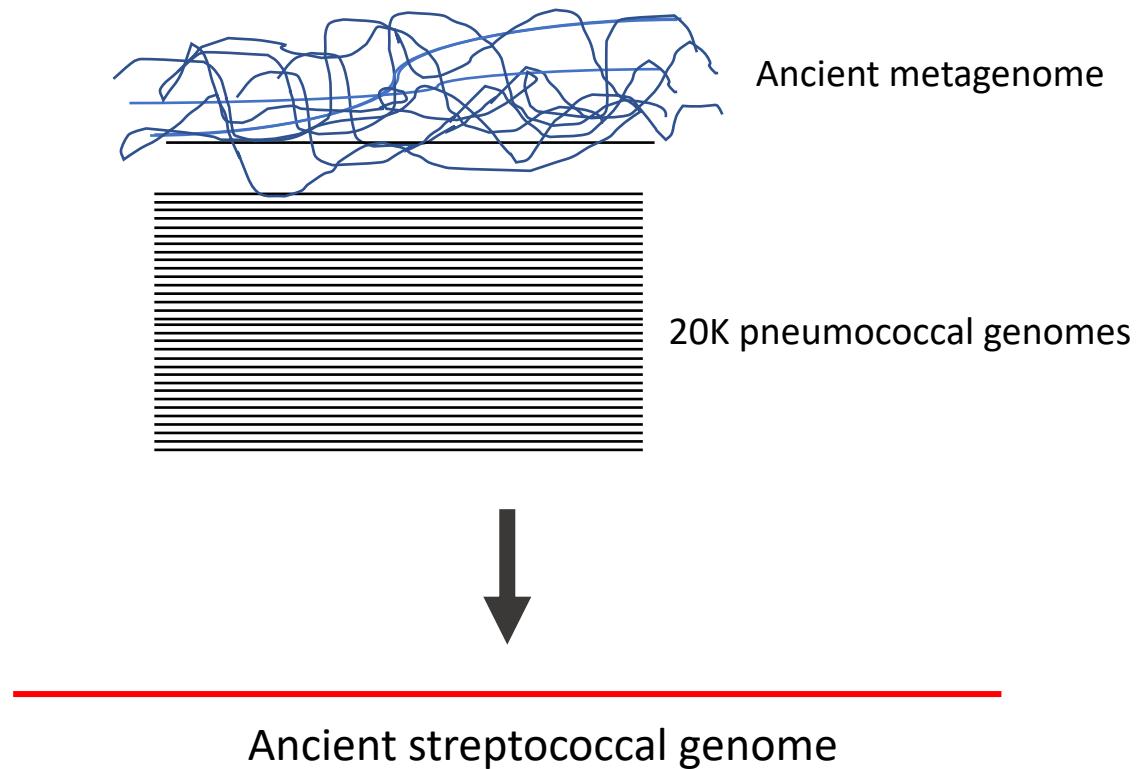
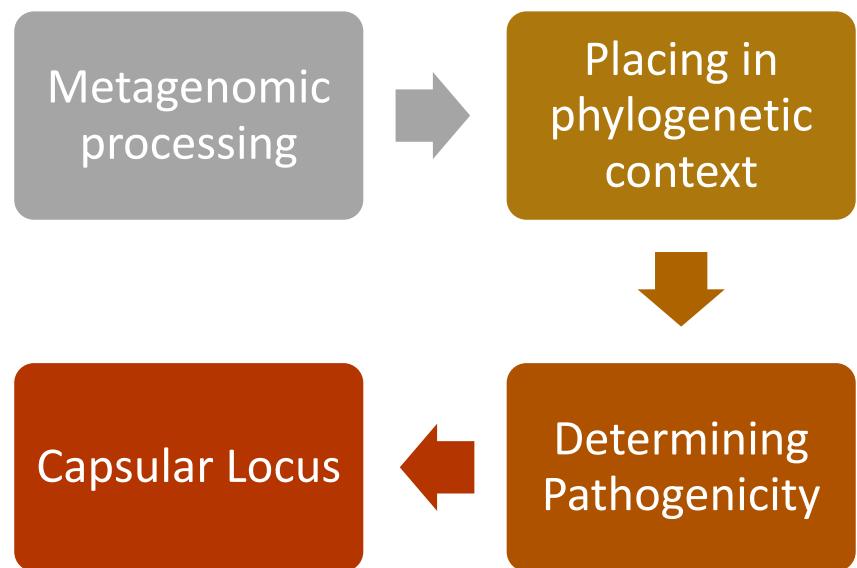
A 5700 year-old human genome and oral microbiome from chewed birch pitch

Theis Z.T. Jensen^{1,2,10}, Jonas Niemann^{1,2,10}, Katrine Højholt Iversen^{3,4,10}, Anna K. Fotakis¹, Shyam Gopalakrishnan¹, Ashild J. Vägène¹, Mikkel Winther Pedersen¹, Mikkel-Holger S. Sinding¹, Martin R. Ellegaard¹, Morten E. Allentoft¹, Liam T. Lanigan¹, Alberto J. Taurozzi¹, Sofie Holtsmark Nielsen¹, Michael W. Dee⁵, Martin N. Mortensen⁶, Mads C. Christensen⁶, Søren A. Sørensen⁷, Matthew J. Collins^{1,8}, M. Thomas P. Gilbert^{1,9}, Martin Sikora¹, Simon Rasmussen^{1,4} & Hannes Schroeder^{1*}

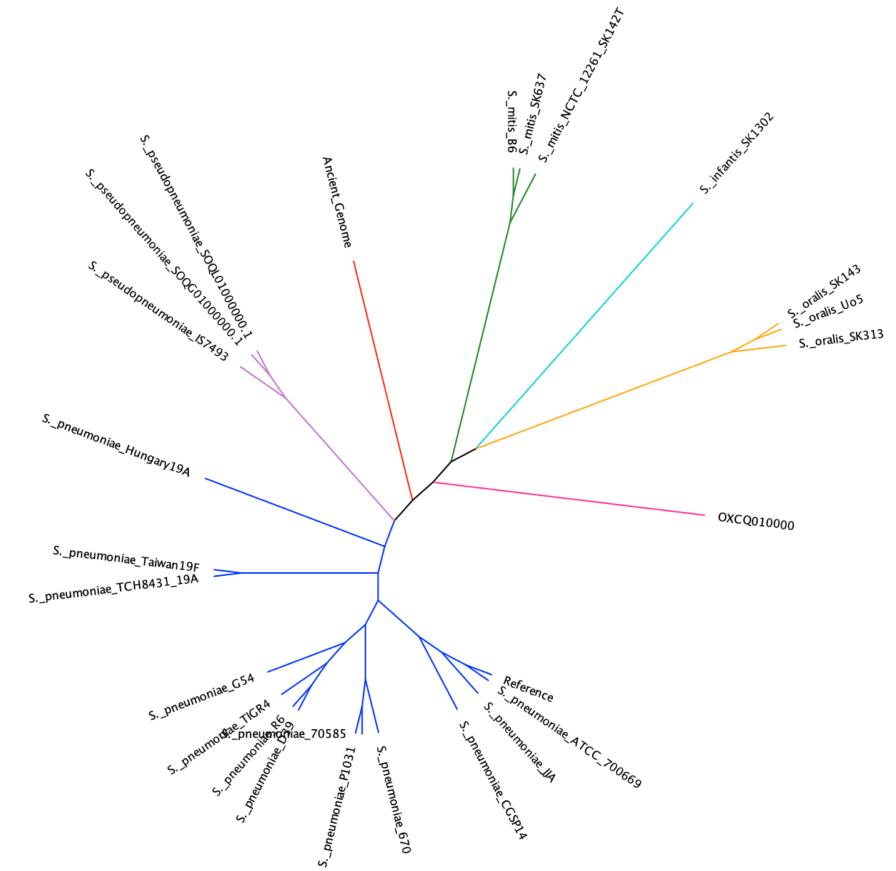
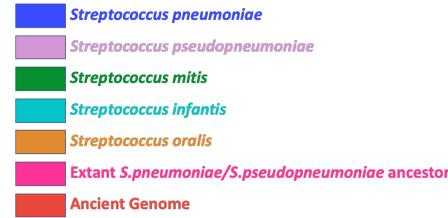
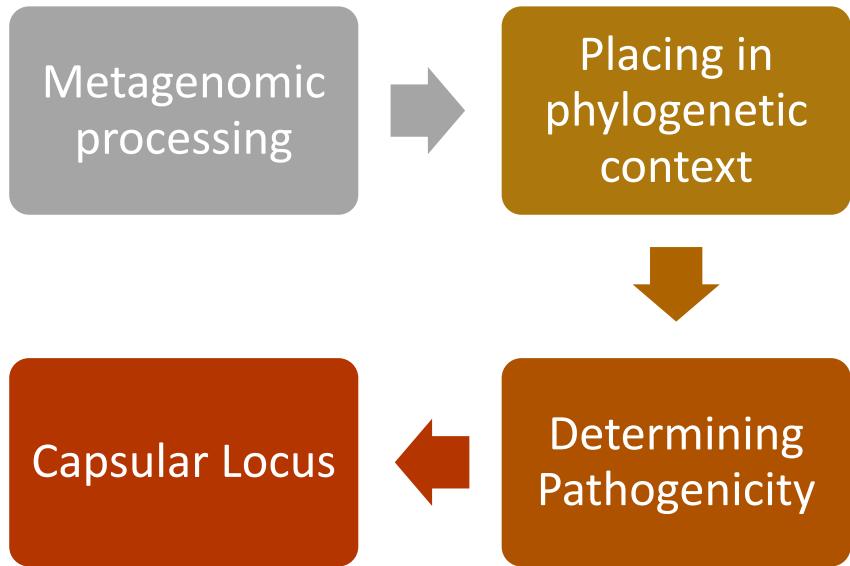
Species	Reads	Fragment length (bp)	DoC
Bacteria			
<i>Neisseria subflava</i>	308,732	56	7.5
<i>Rothia mucilaginosa</i>	296,610	52	6.9
<i>Streptococcus pneumoniae</i>	176,782	57	4.7
<i>Neisseria cinerea</i>	153,683	58	4.9
<i>Lautropia mirabilis</i>	117,040	53	2.0
<i>Neisseria meningitidis</i>	100,540	51	2.3
<i>Aggregatibacter segnis</i>	95,670	58	2.8
<i>Neisseria elongata</i>	68,407	54	1.6
<i>Prevotella intermedia</i>	65,324	56	1.2
<i>Streptococcus</i> sp. ChDC B345	52,614	61	1.6
<i>Streptococcus</i> sp. 431	43,787	59	1.2
<i>Aggregatibacter aphrophilus</i>	43,231	56	1.1
<i>Streptococcus pseudopneumoniae</i>	38,832	61	1.1



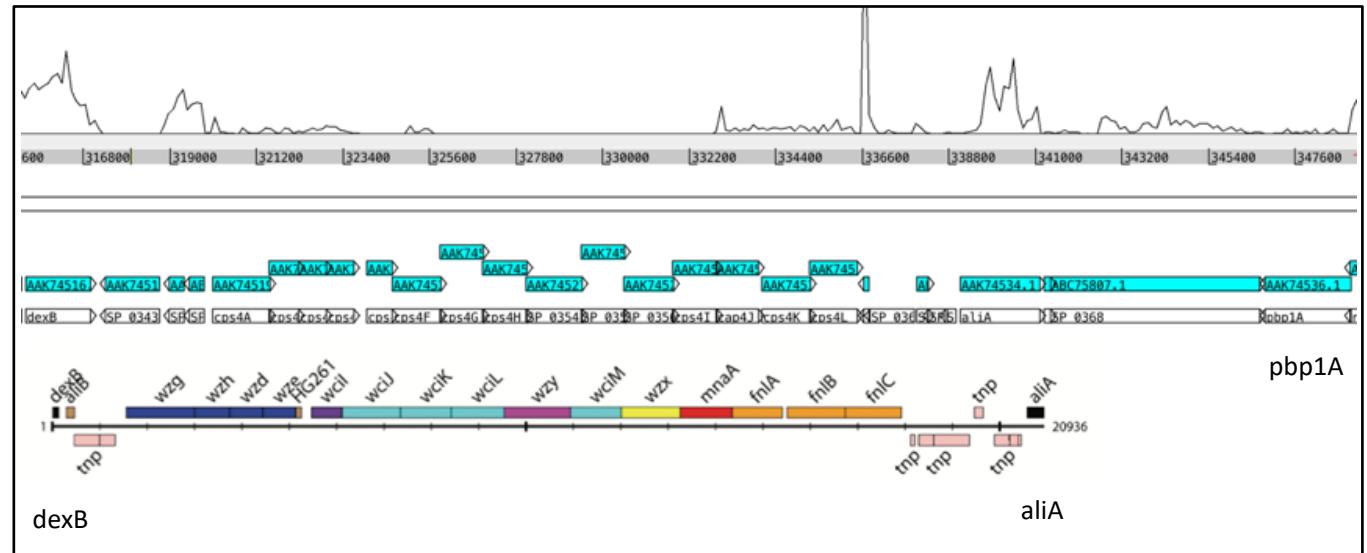
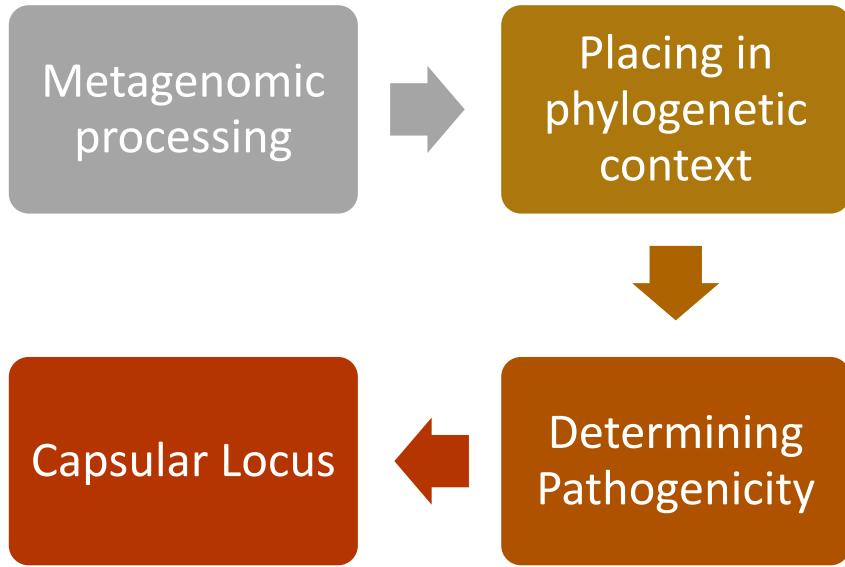
Ancient *Streptococcal* genome



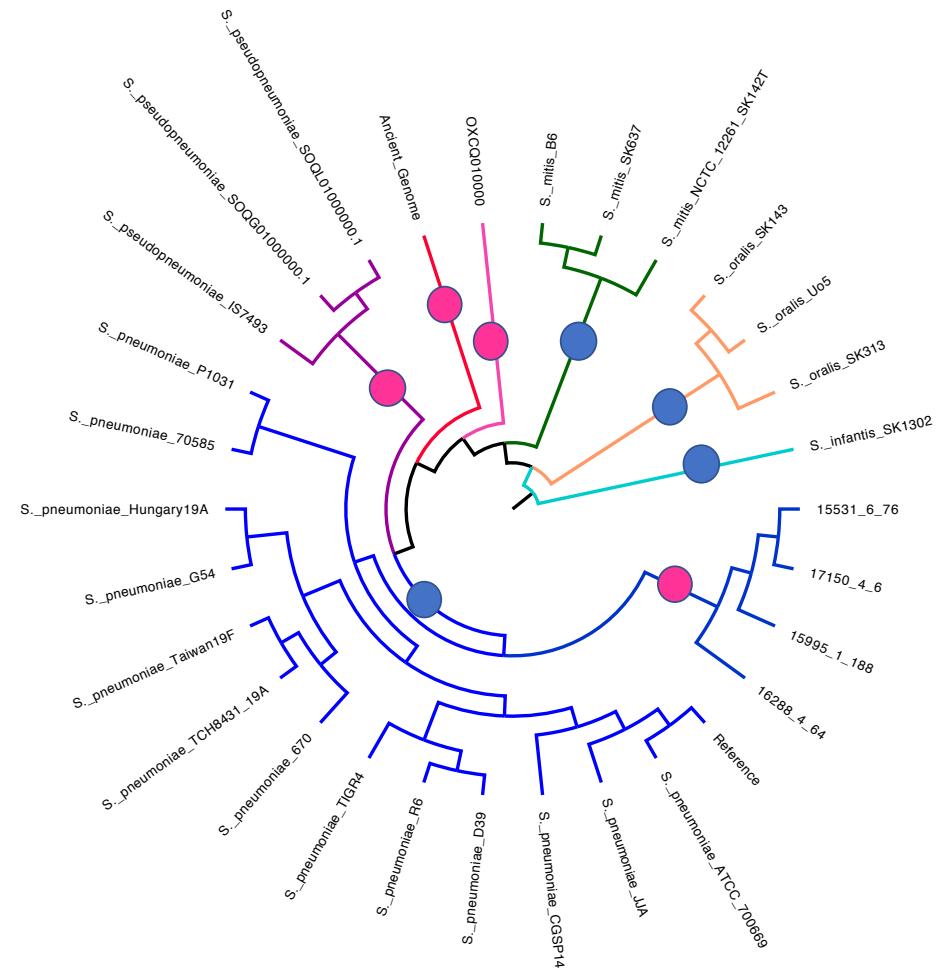
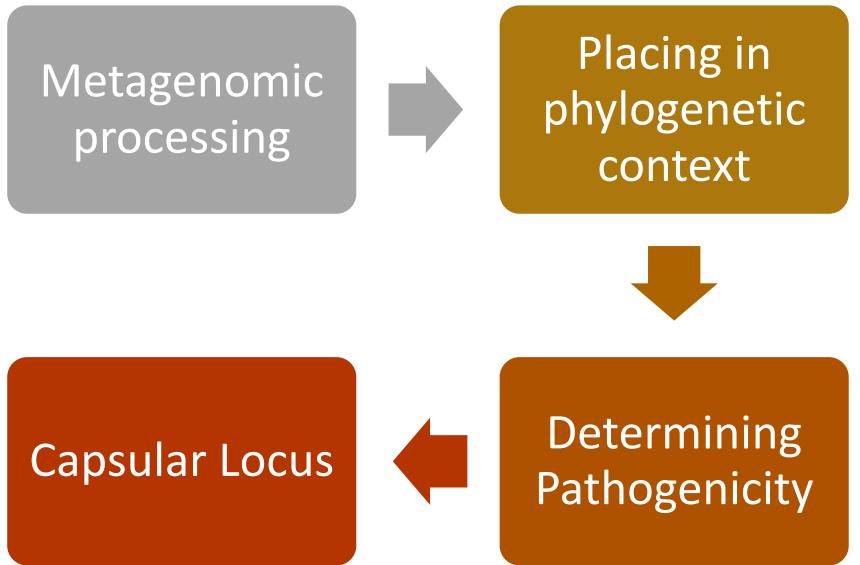
Ancient Streptococcal genome



Ancient Streptococcal genome



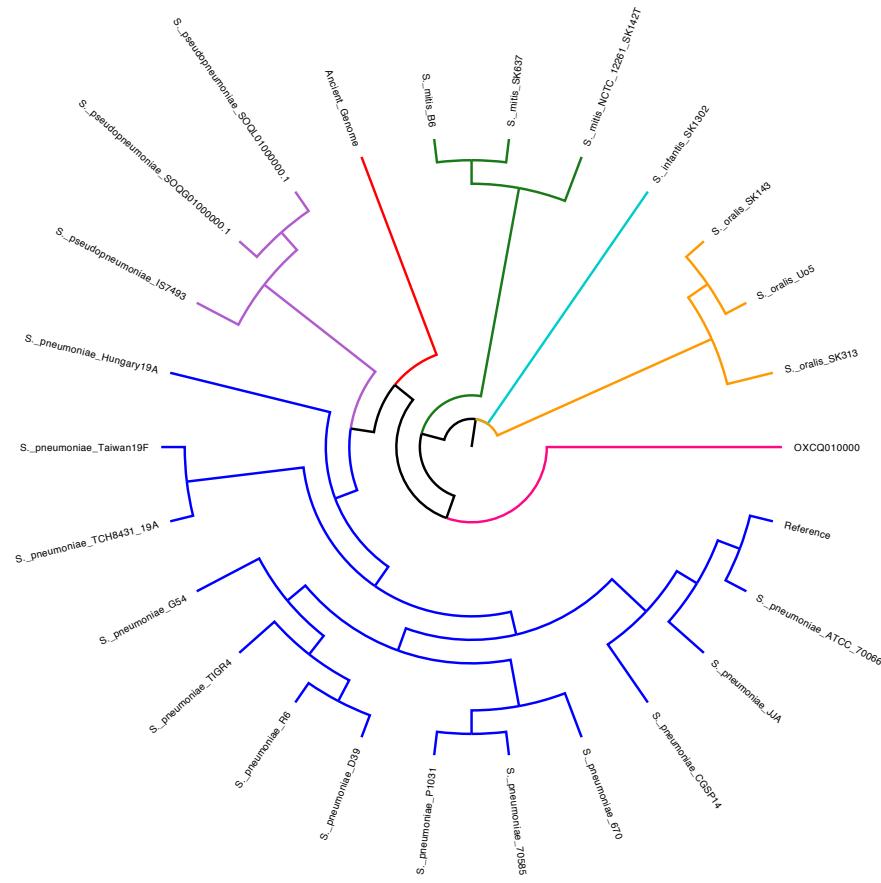
Ancient Streptococcal genome



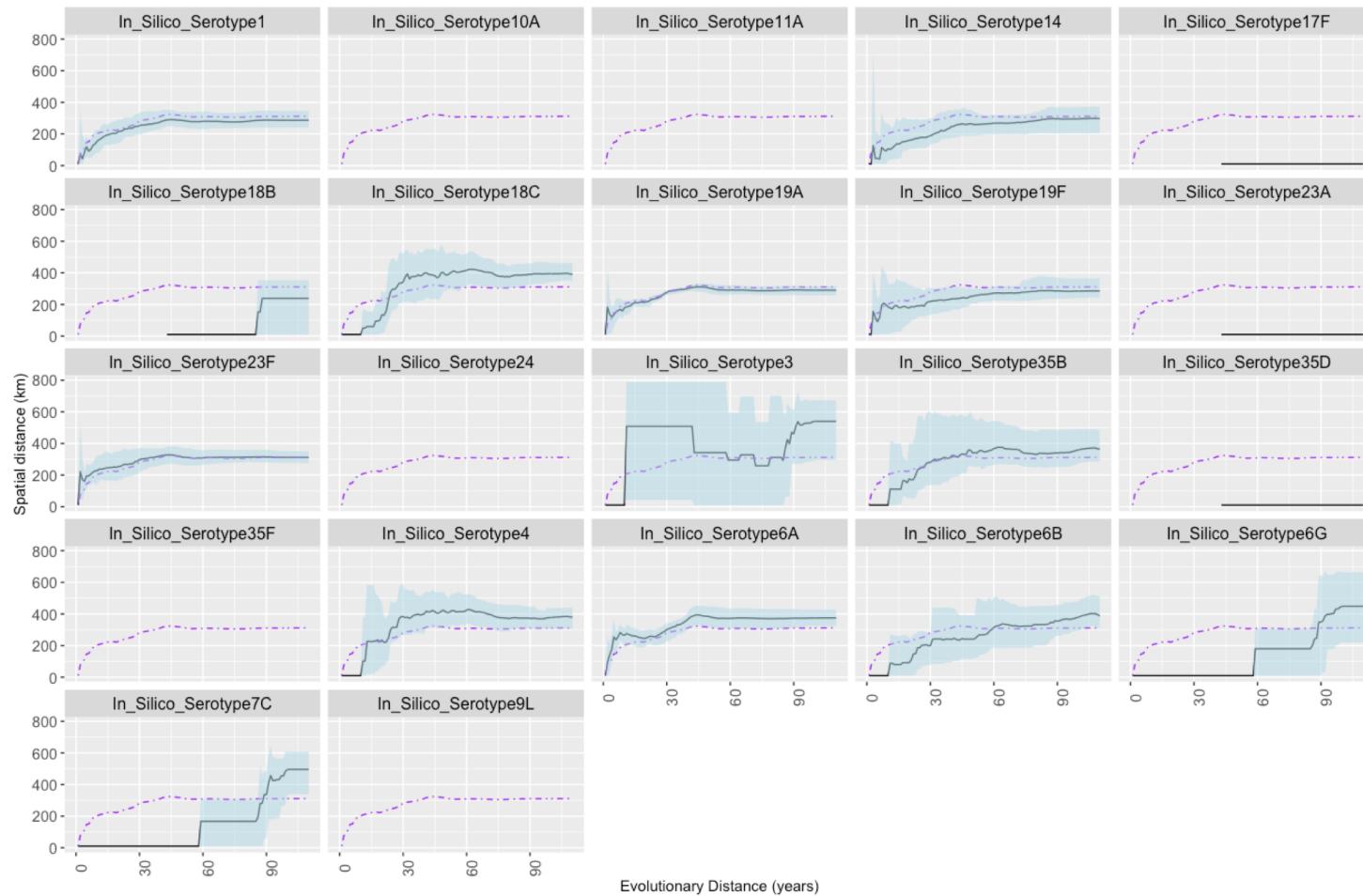
- *dexB-alternative ali/glf/pspK genes*
- *dexB-aliA typical cps*

Ancient “*Streptococcus syltholmis*” genome conclusions

- Novel extinct streptococcal species
- Non-encapsulated
- Possibly shorter carriage duration historically?
- Extant non-encapsulated *S. pneumo* may not be a result of a loss event

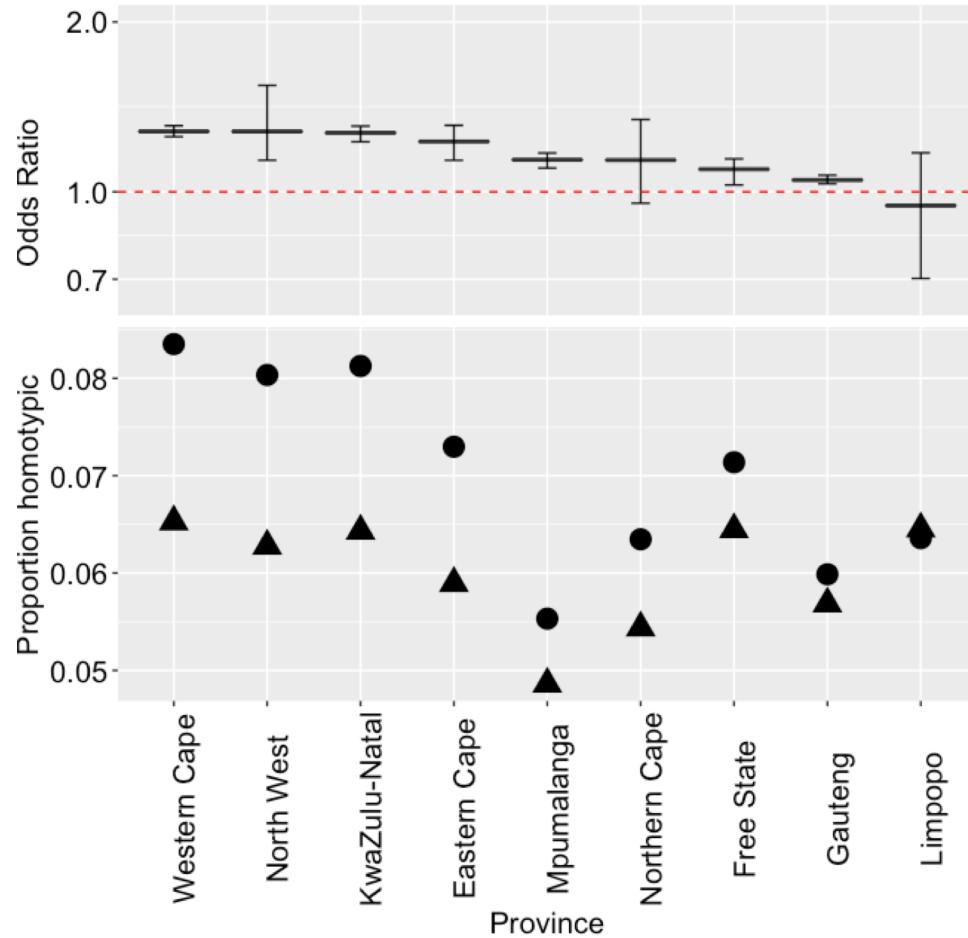


Serotype spread.



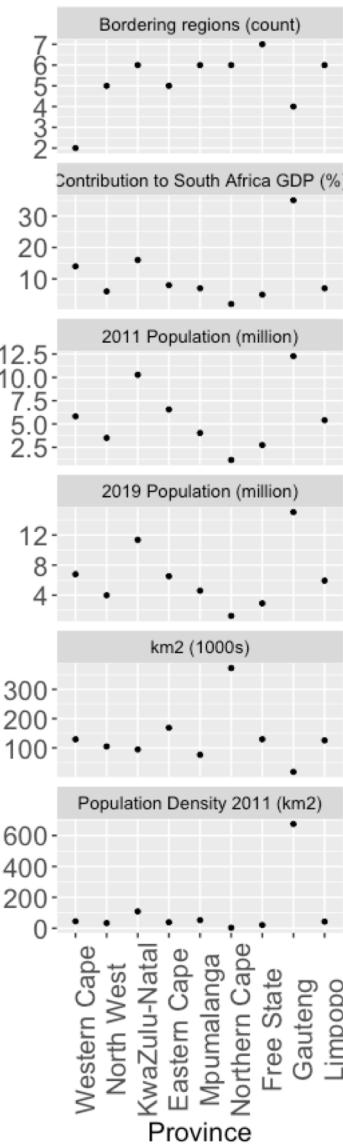
More homotypic serotypes within provinces than between provinces

A



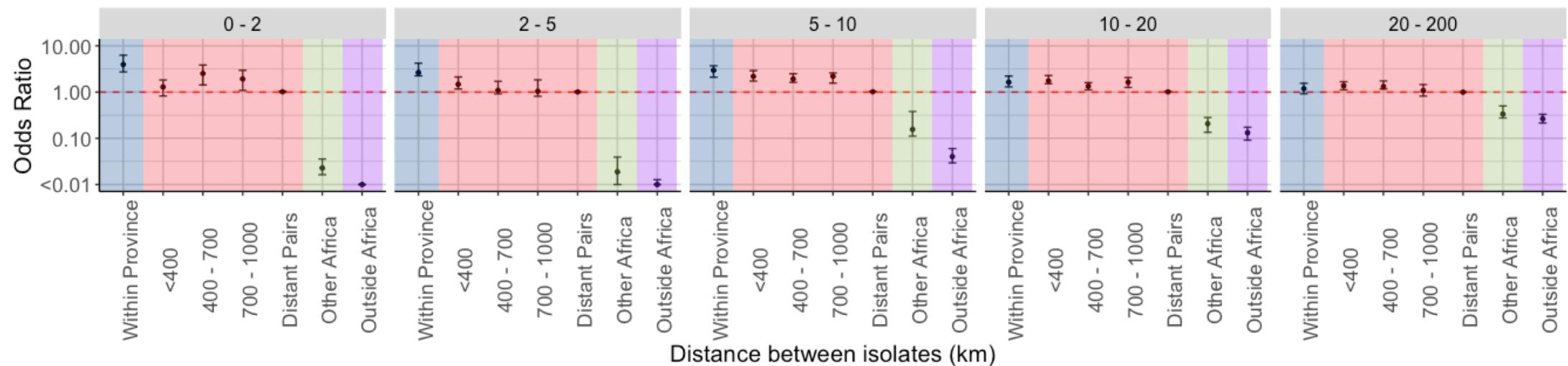
B

C

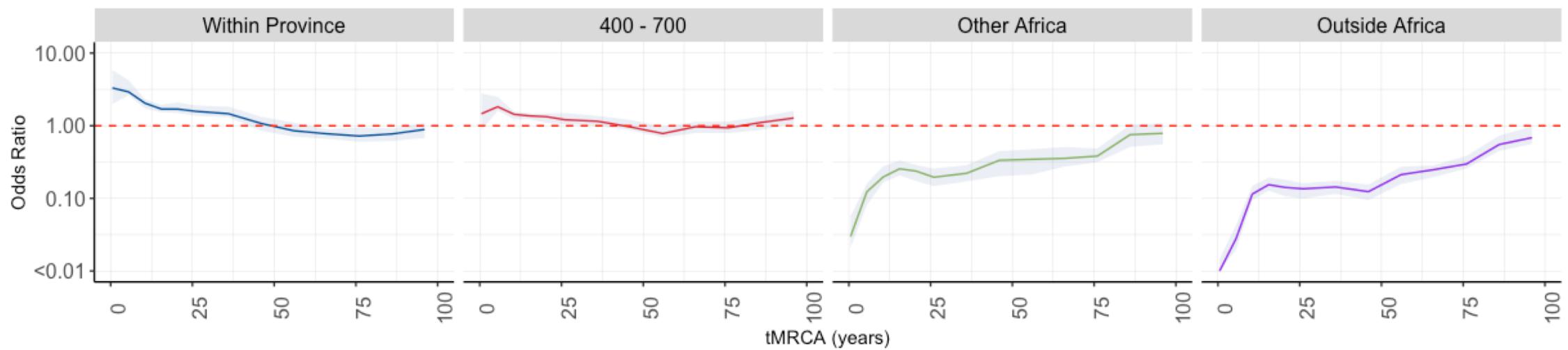


50 years for strains to become homogenous across South Africa (WEIGHTED)

A



B



Impact of different types of sampling

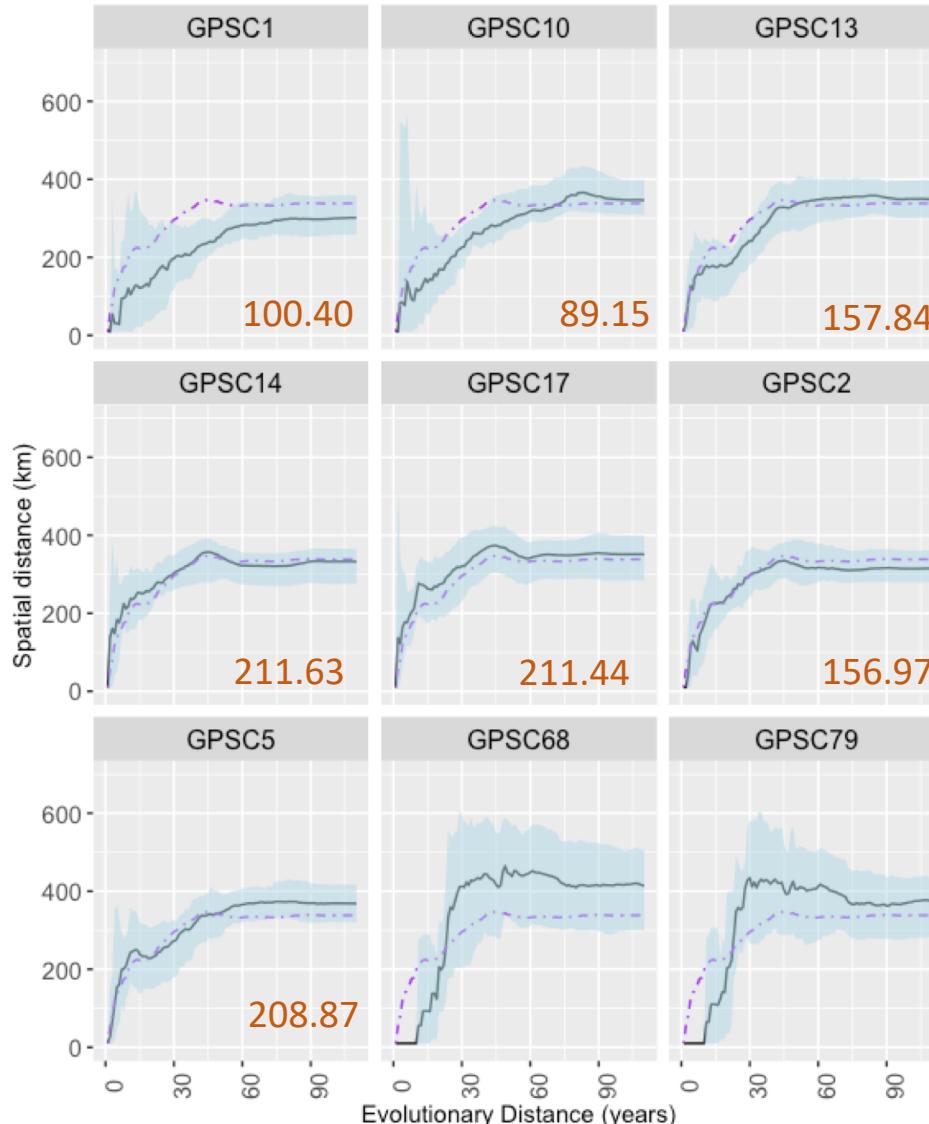
Rates for individual GPSCs are similar to overall

Purple dashed line is overall.

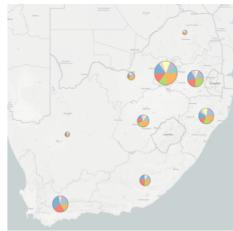
Light blue band is 95% confidence intervals.

Labeled mean (μ_k) rate of transmission at 10 years.

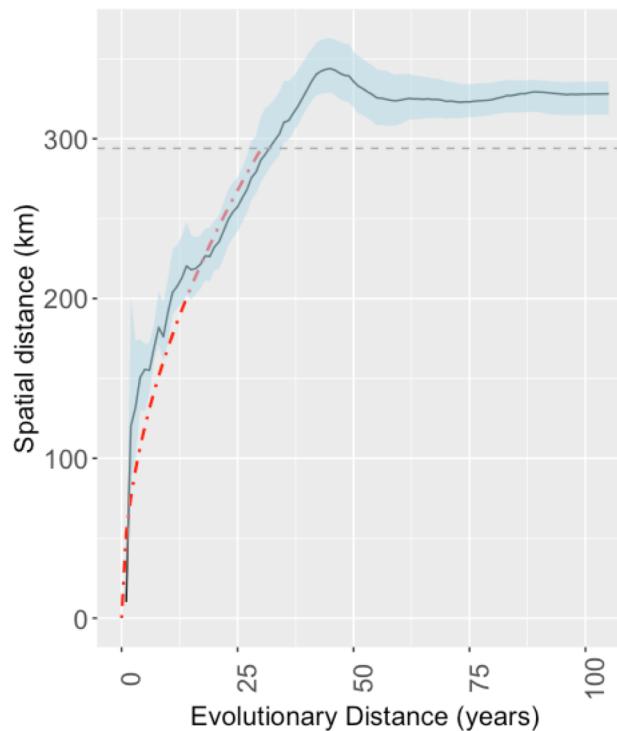
Sub-sampling to 50 per GPSC in each region without replacement. Cumulative over time.



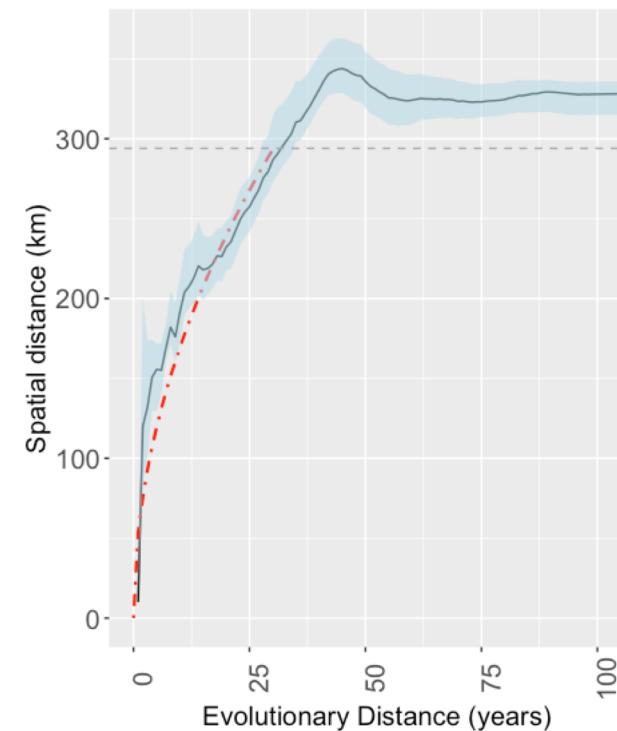
Accounting for sampling bias



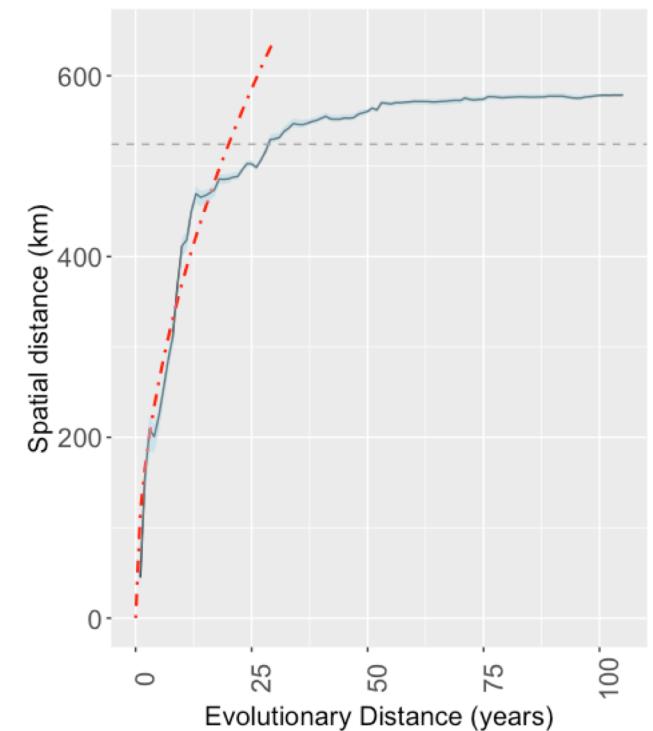
Subsample to 300



Subsample to 50

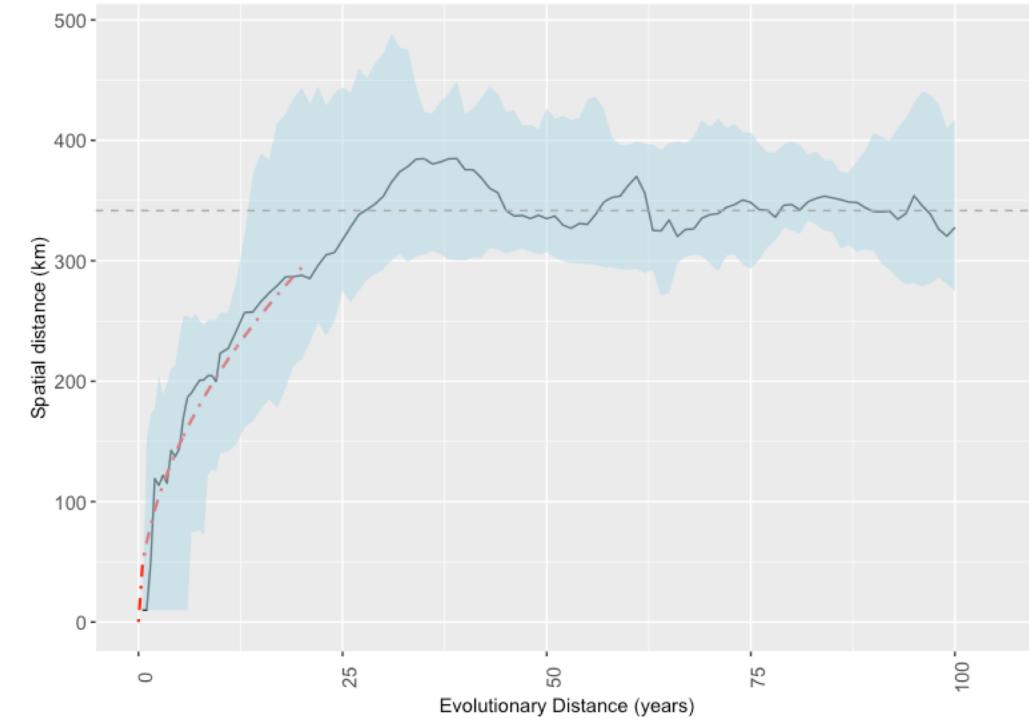


Weighted by GPSC and Region*

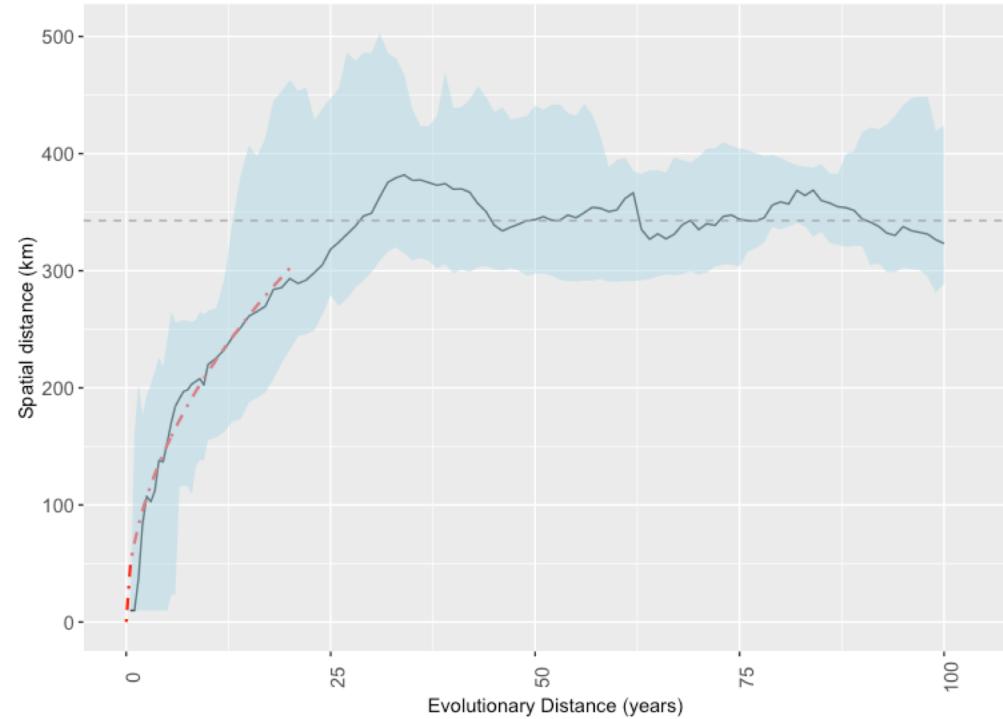


*over representation of Gauteng - Mpumalanga

Testing replacements

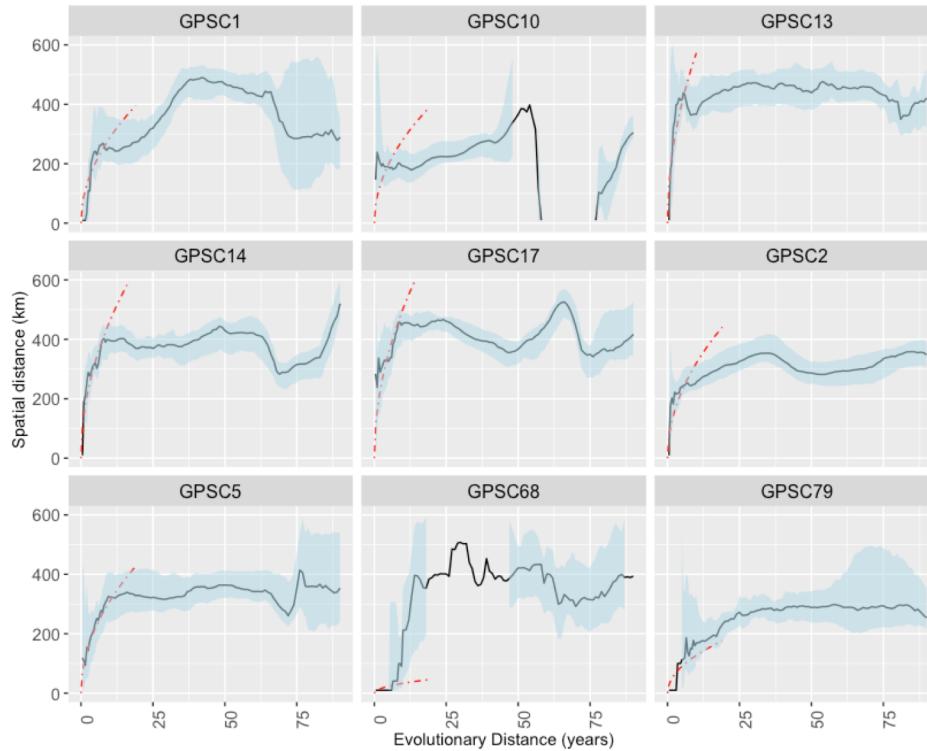


```
selectedSeqs=NULL
for (i in 1:9) {
  a <- which(vector_region[gpst]==region_vec[i])
  b<- if (length(a) == 1) (a) else sample(a,min(50,length(a)),replace=F)
  selectedSeqs <- c(selectedSeqs,b)
}
totvec.tmp<-sample(selectedSeqs,replace = T)
```



```
selectedSeqs=NULL
for (i in 1:9) {
  a <- which(vector_region[gpst]==region_vec[i])
  b<- if (length(a) == 1) (a) else sample(a,min(50,length(a)),replace=T)
  selectedSeqs <- c(selectedSeqs,b)
}
totvec.tmp<-selectedSeqs
```

No subsampling of GPSC



Subsampling to 50

