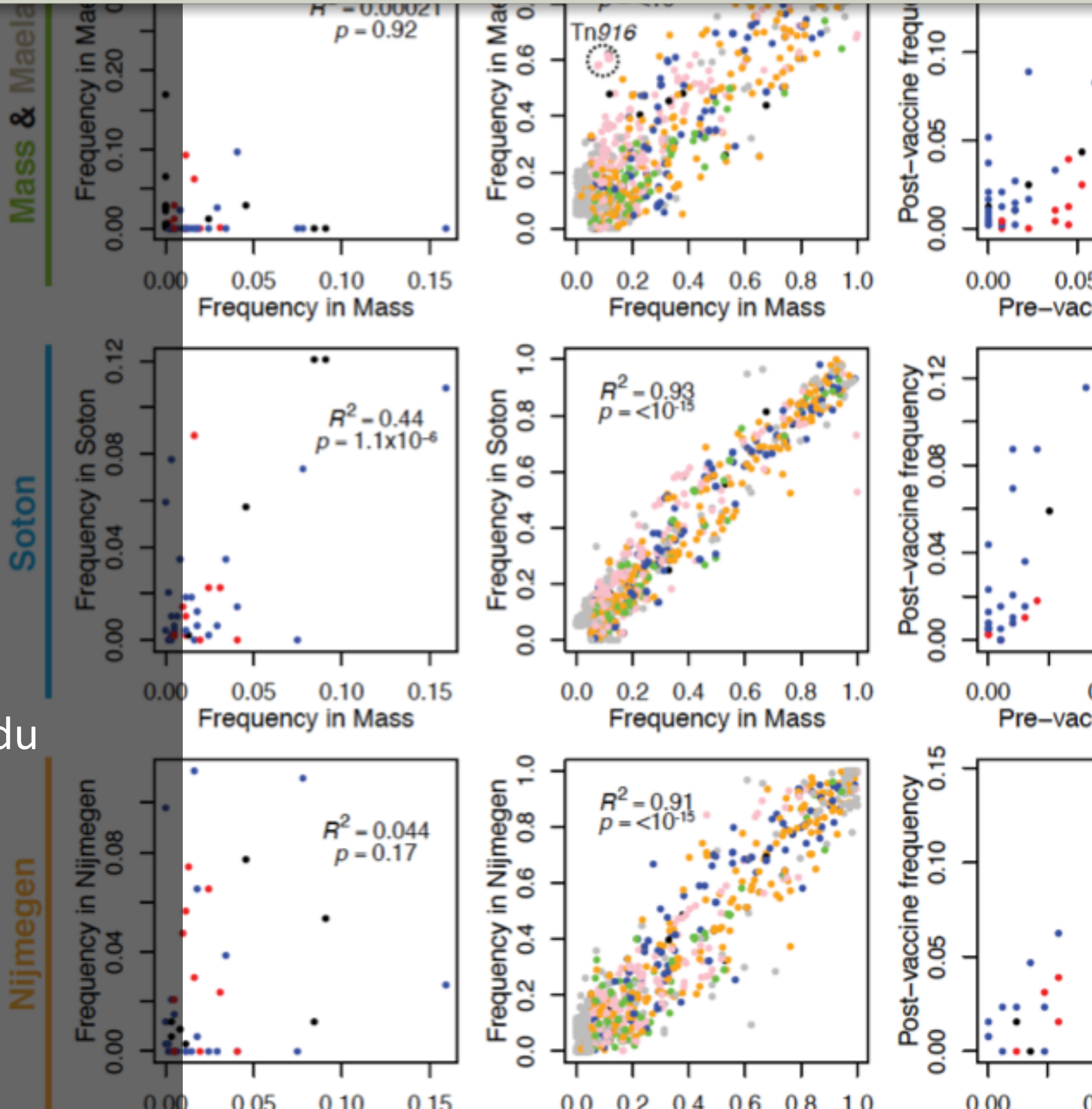
HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH



Pneumococcal genomics, vaccines and AMR

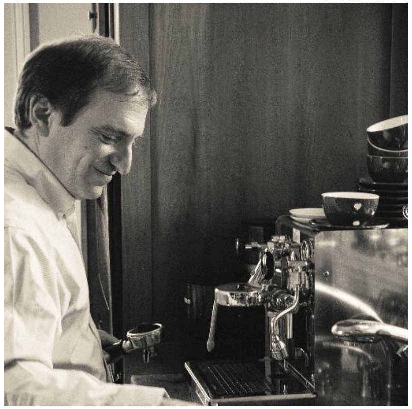
Bill Hanage
November 2021
whanage@hsph.harvard.edu

CENTER *for*
COMMUNICABLE
DISEASE DYNAMICS

Acknowledgements

Taj Azarian

Pamela Martinez



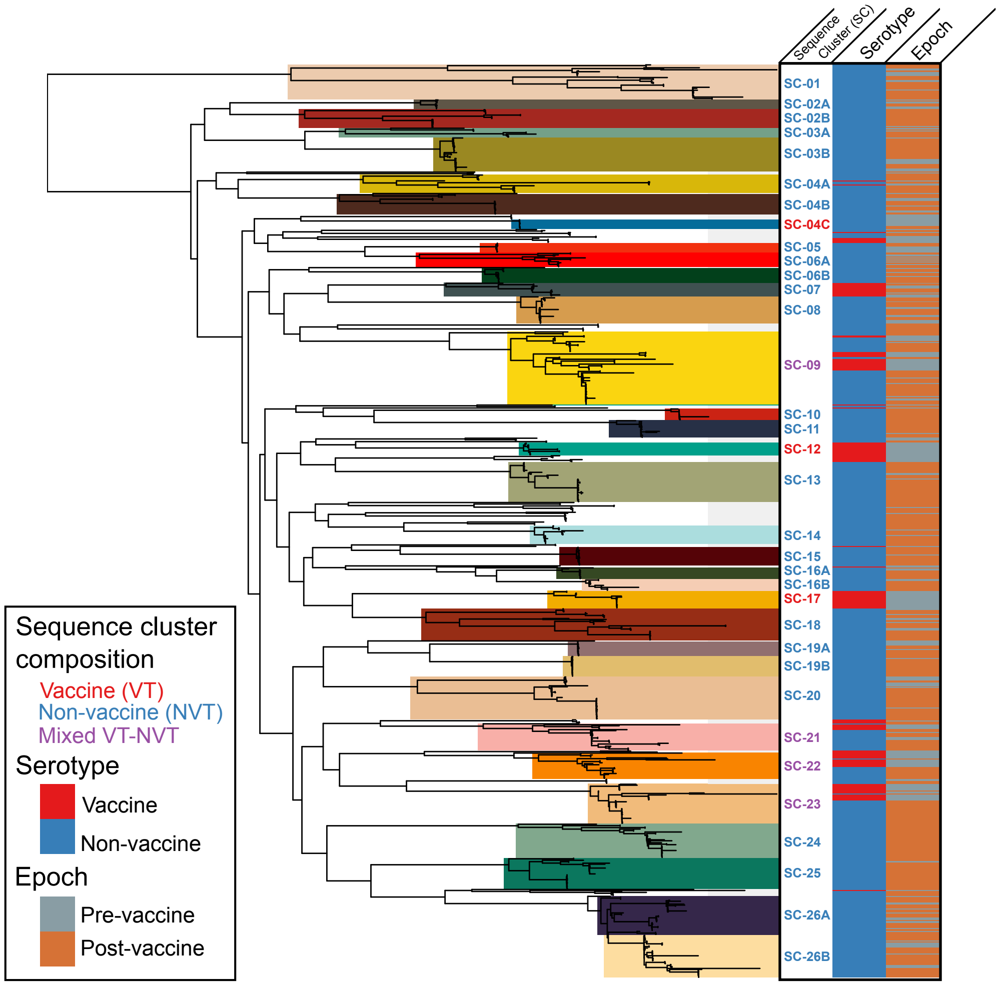
Marc Lipsitch

Forecasting the consequence of vaccination

- Following the removal of vaccine serotypes, non vaccine serotypes increase in carriage (and disease, depending on their virulence)
- Can we predict which will increase?

Vaccination in the South West US

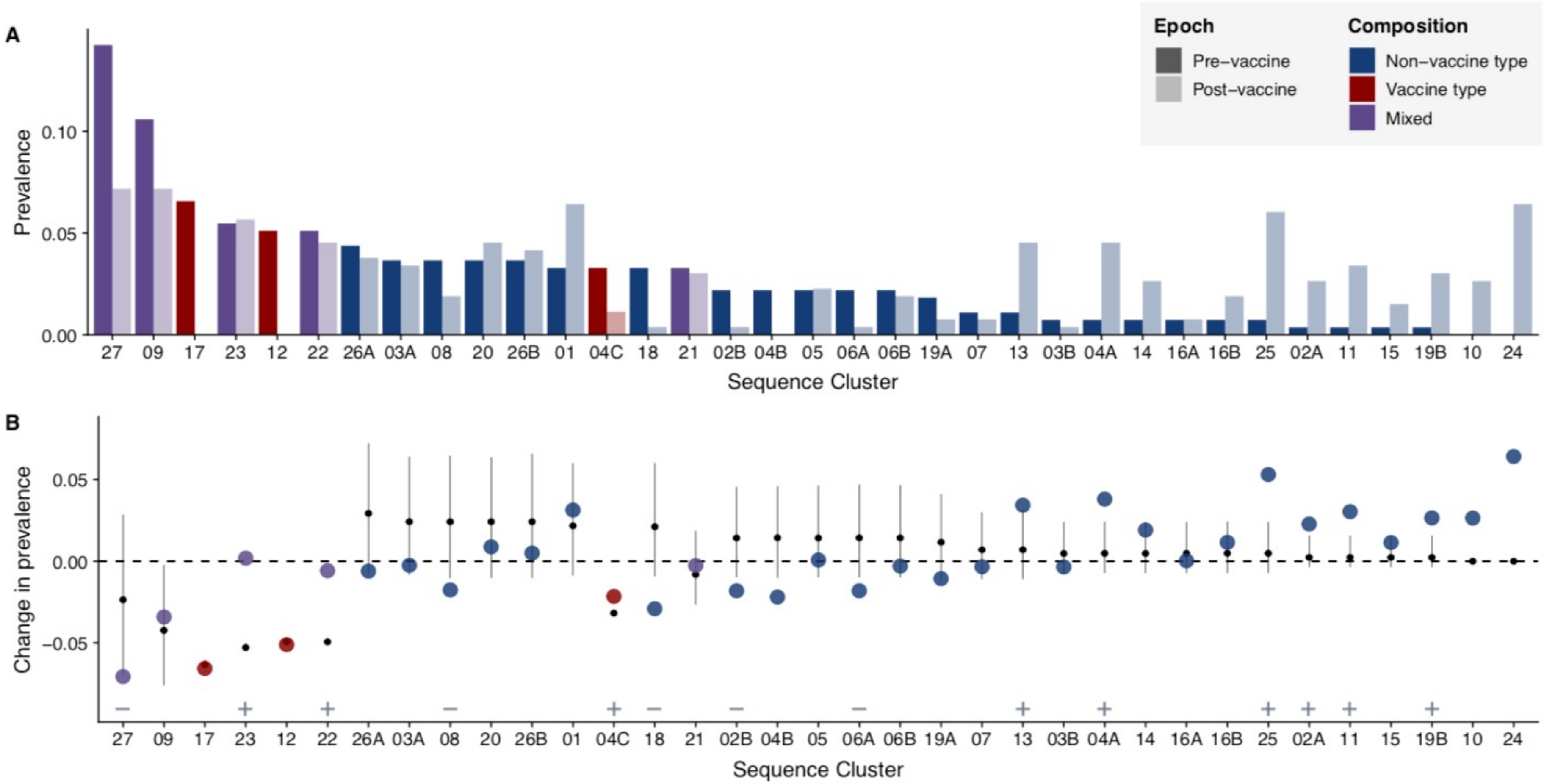
- Carriage samples straddling vaccination, from Native American communities (N=937)
- 35 “Sequence Clusters”
- Some are VT, some are NVT and some are mixed
- Vaccination did not change carriage prevalence



CENTER FOR AMERICAN INDIAN HEALTH

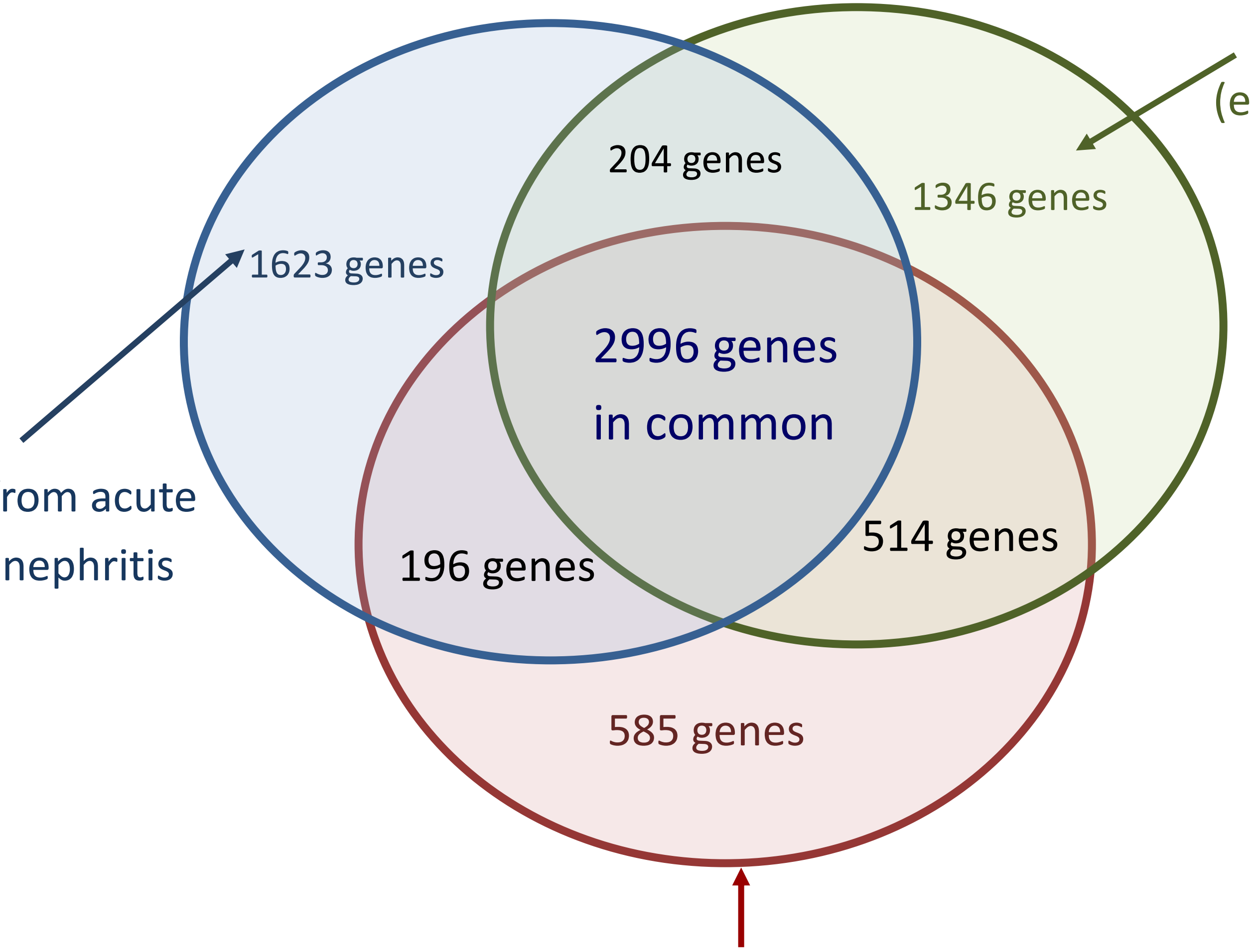
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Changes in the prevalence of sequence clusters following vaccine



Why do some sequence clusters increase more (or less) than expected?

Introducing the accessory genome



E. coli O157 H7
(enterohaemorrhagic)

1346 genes

204 genes

1623 genes

2996 genes
in common

E. coli from acute
pyelonephritis

196 genes

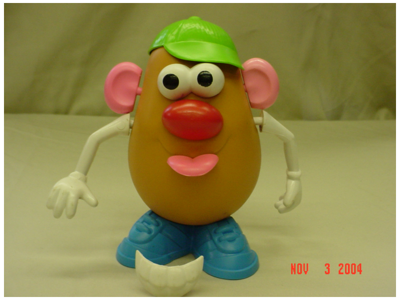
514 genes

585 genes

E. coli K12
(commensal)

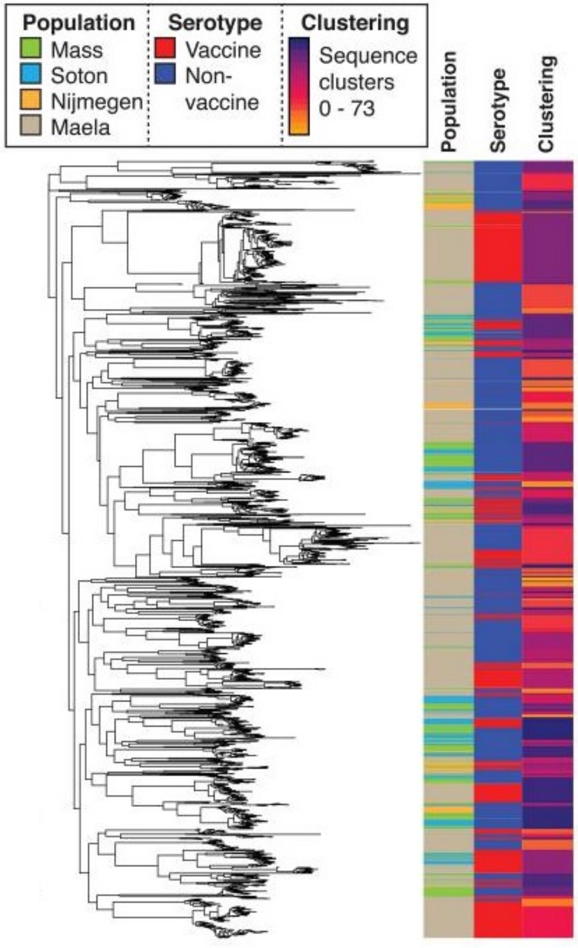
Only 39% of genes present in all three strains

Bacteria are Mr Potato Heads



Welch *et al*. 2002 PNAS **99:** 17020

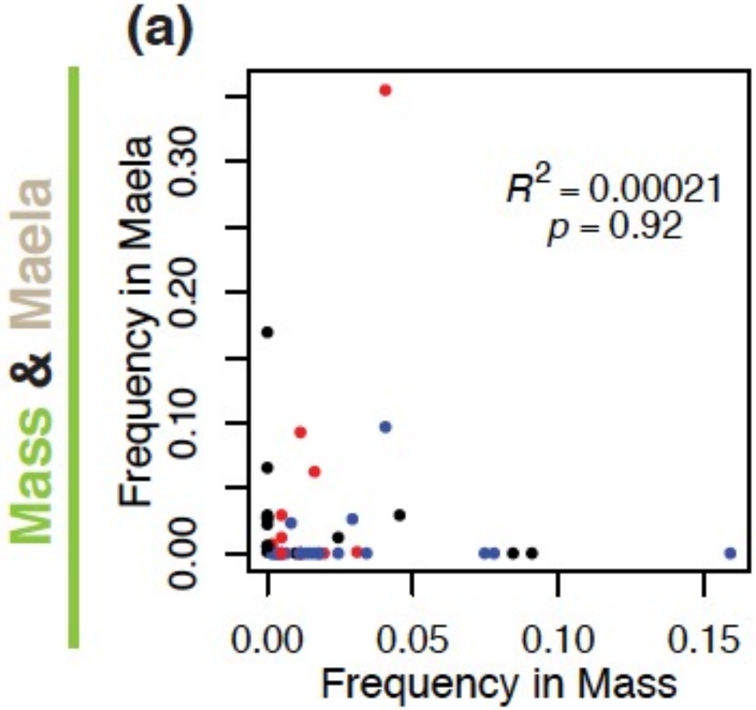
How does the accessory genome vary between four different sample sites?



- Samples from MA (N=614)
- Southampton, UK (N=516)
- Nijmegen, the Netherlands (N=337)
- Maela camp, Thailand (N=3,085)
- Total of 4,127 isolates, falling into 73 Sequence clusters, and 1,731 accessory COGs*
- *defined as present in 5%-95% of isolates, to avoid dodgy sequence from assembly errors and the like

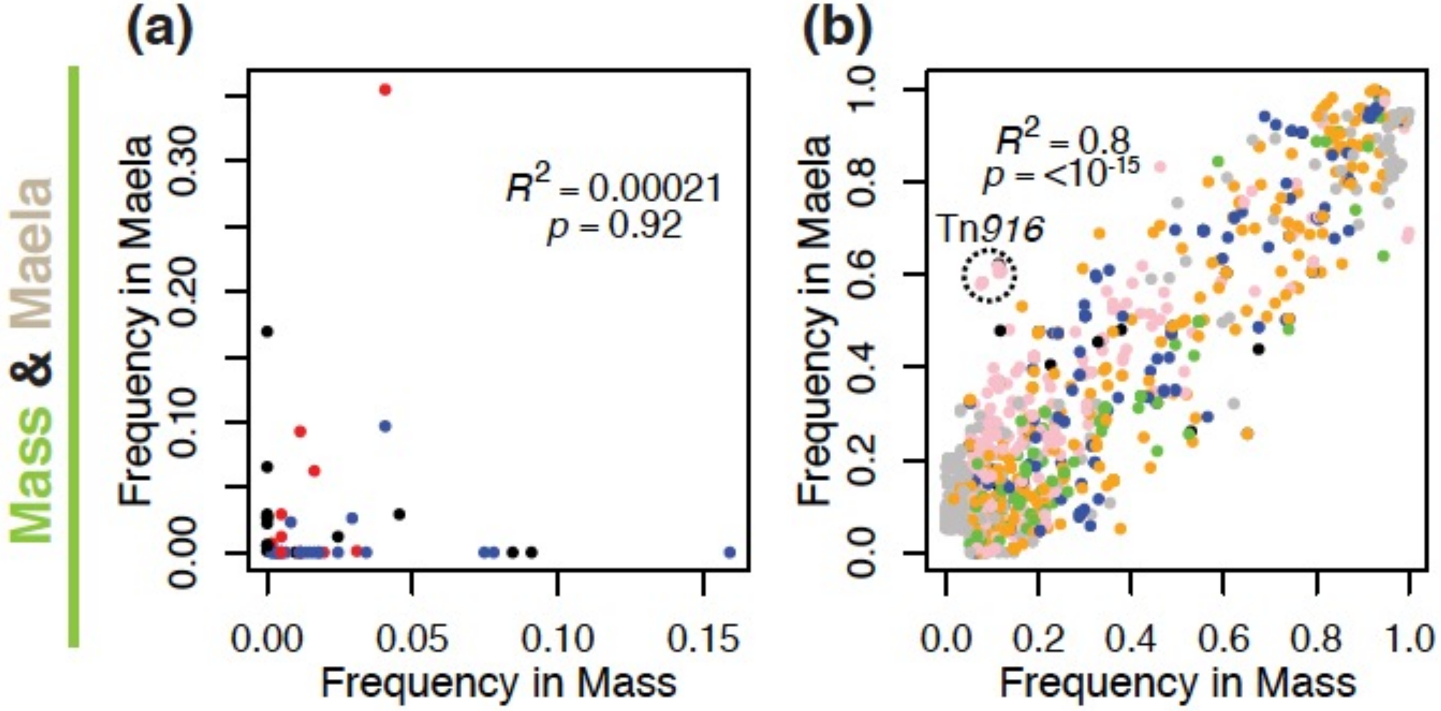
Corander et al Nat Eco Evol 2017

Comparing Massachusetts and Thailand



Corander et al Nat Eco Evol 2017

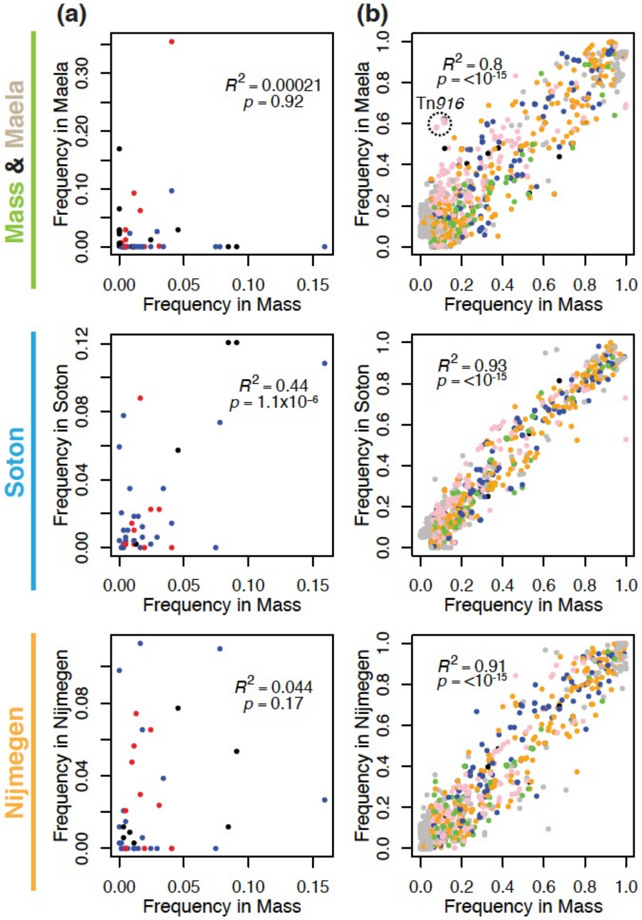
Comparing Massachusetts and Thailand



Corander et al Nat Eco Evol 2017

Population frequencies of accessory loci in different places are highly correlated

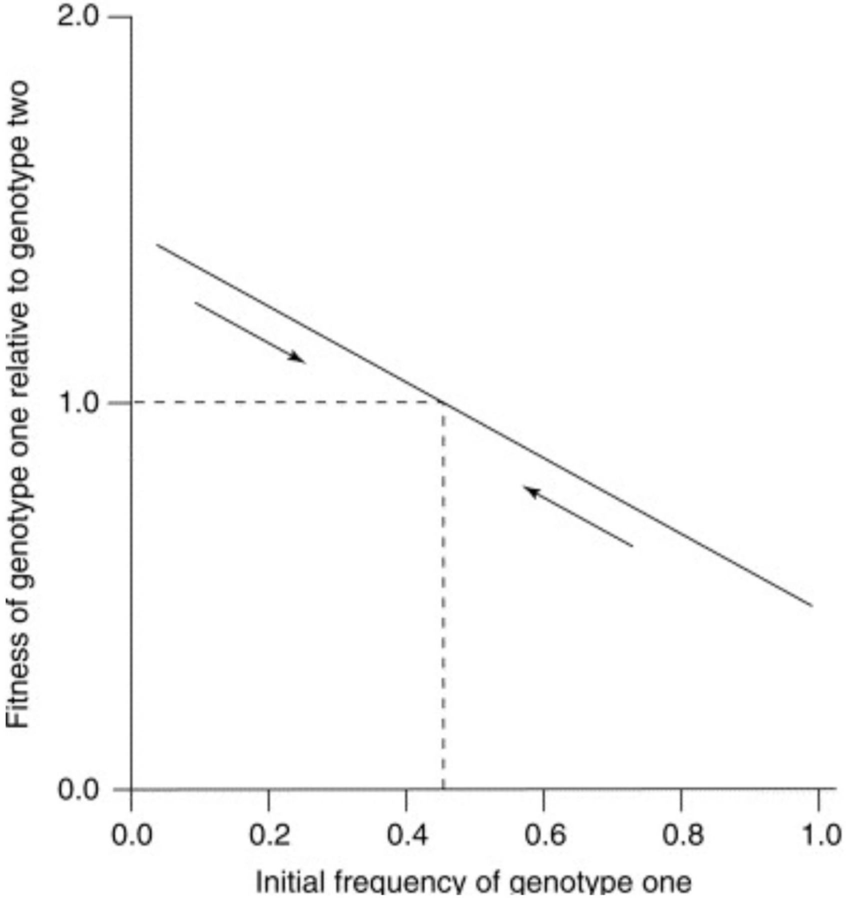
“Pickle plots”



Corander et al Nat Eco Evol 2017

Negative Frequency Dependent Selection

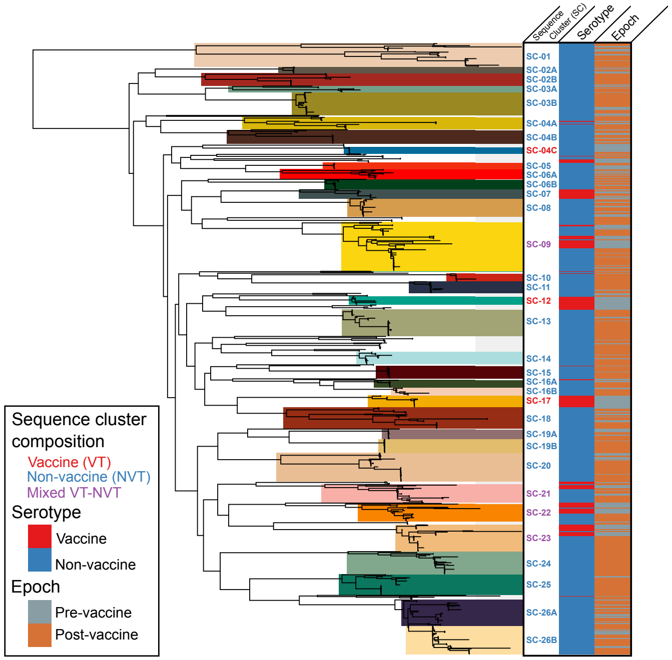
- Form of *balancing selection*
- Think about surface antigens: too common = too much immunity
- Or bacteriophage receptors, or many other genes

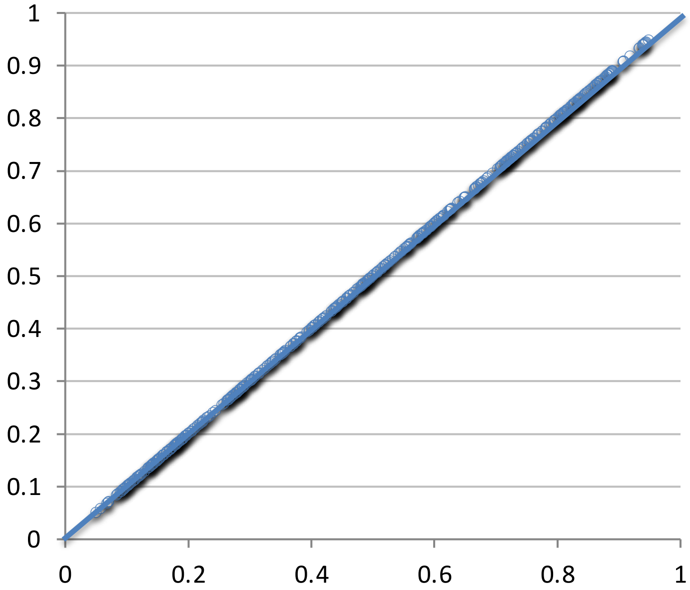


Hastings *Parasitology* 2006

If NFDS structures the population, can we use it to predict the impact of removing vaccine types?

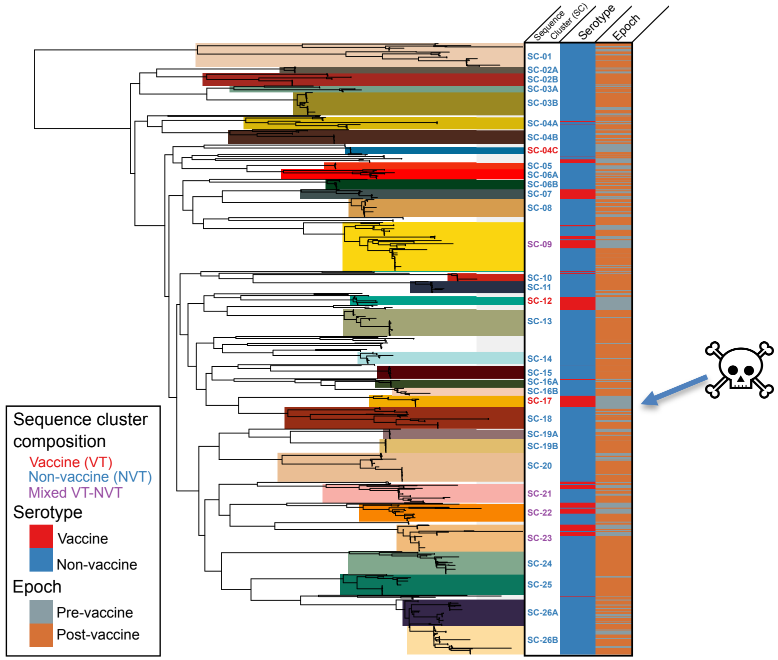
The predicted fitness

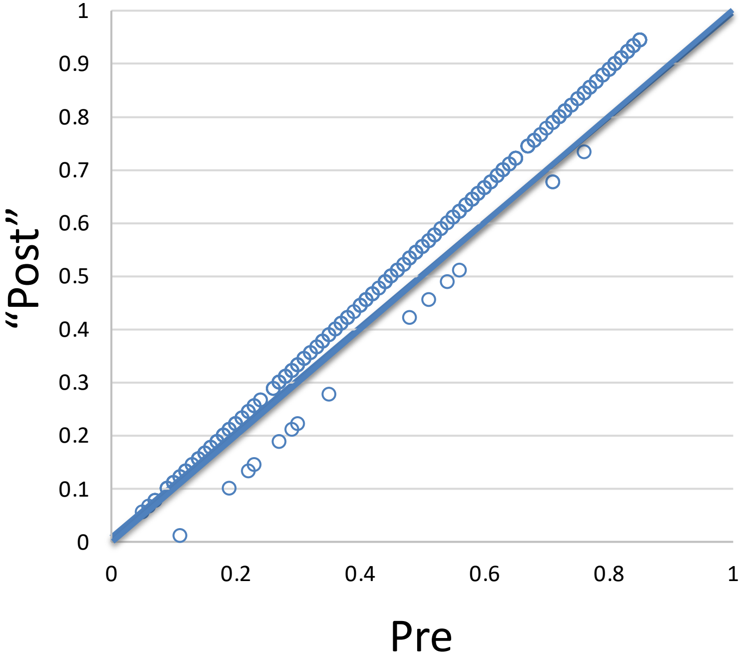




Isolates in the same sequence cluster tend to have similar accessory genome content (Croucher et al Nature Comms 2014)

Compute the consequences of removing a Vaccine type SC

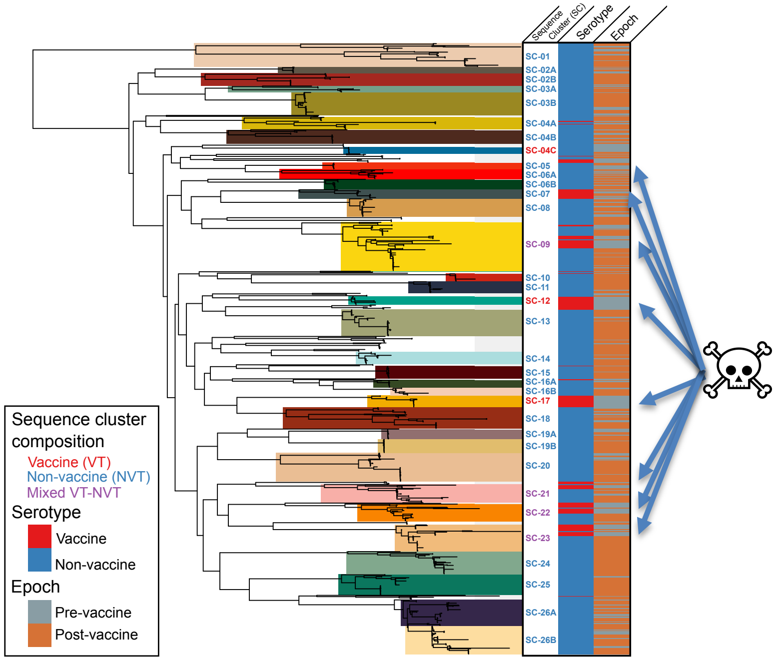


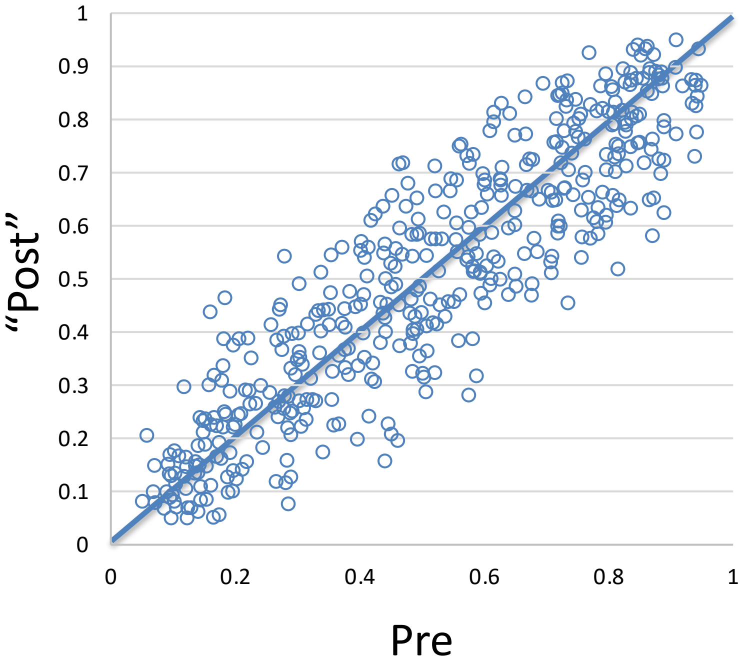


And its accessory loci

Case shown is removal of an SC making up 10% of the pre vaccine population

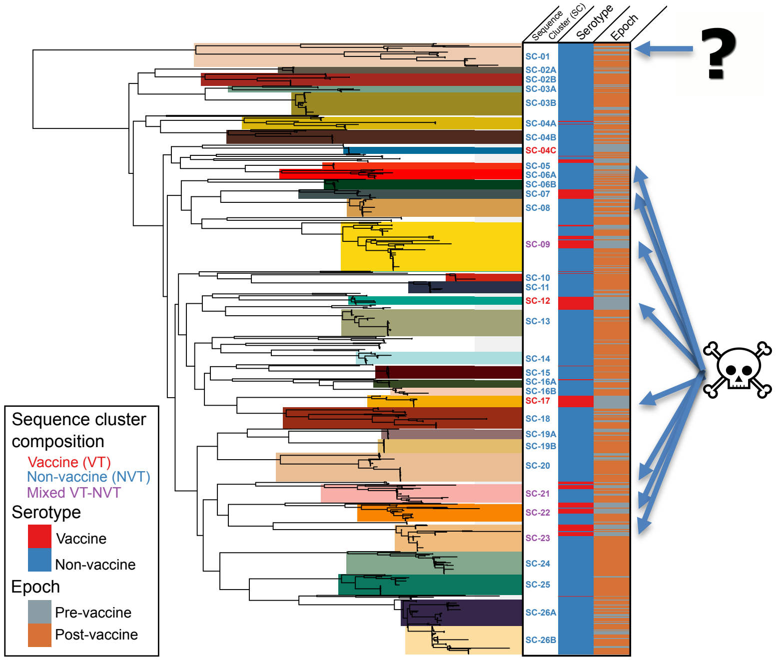
Consequence of removing all vaccine types

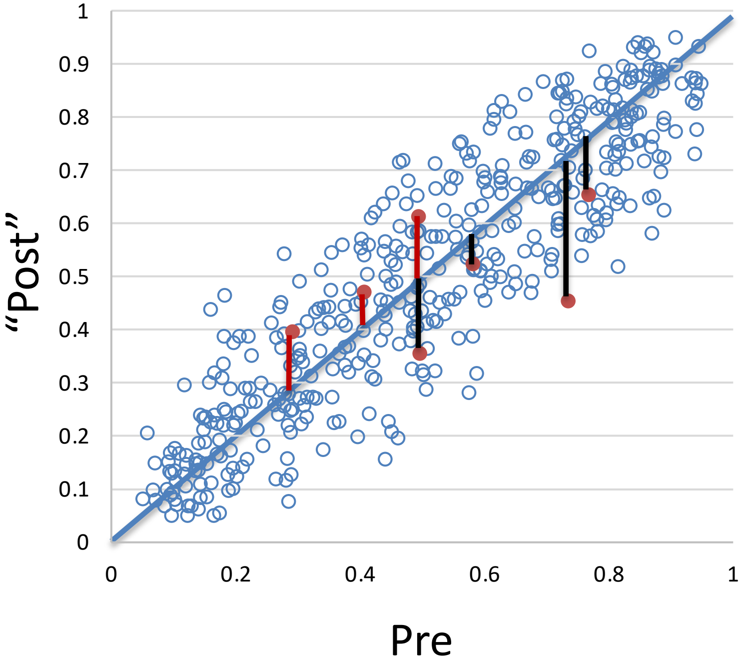




More scatter

Predicting the fitness of any SC

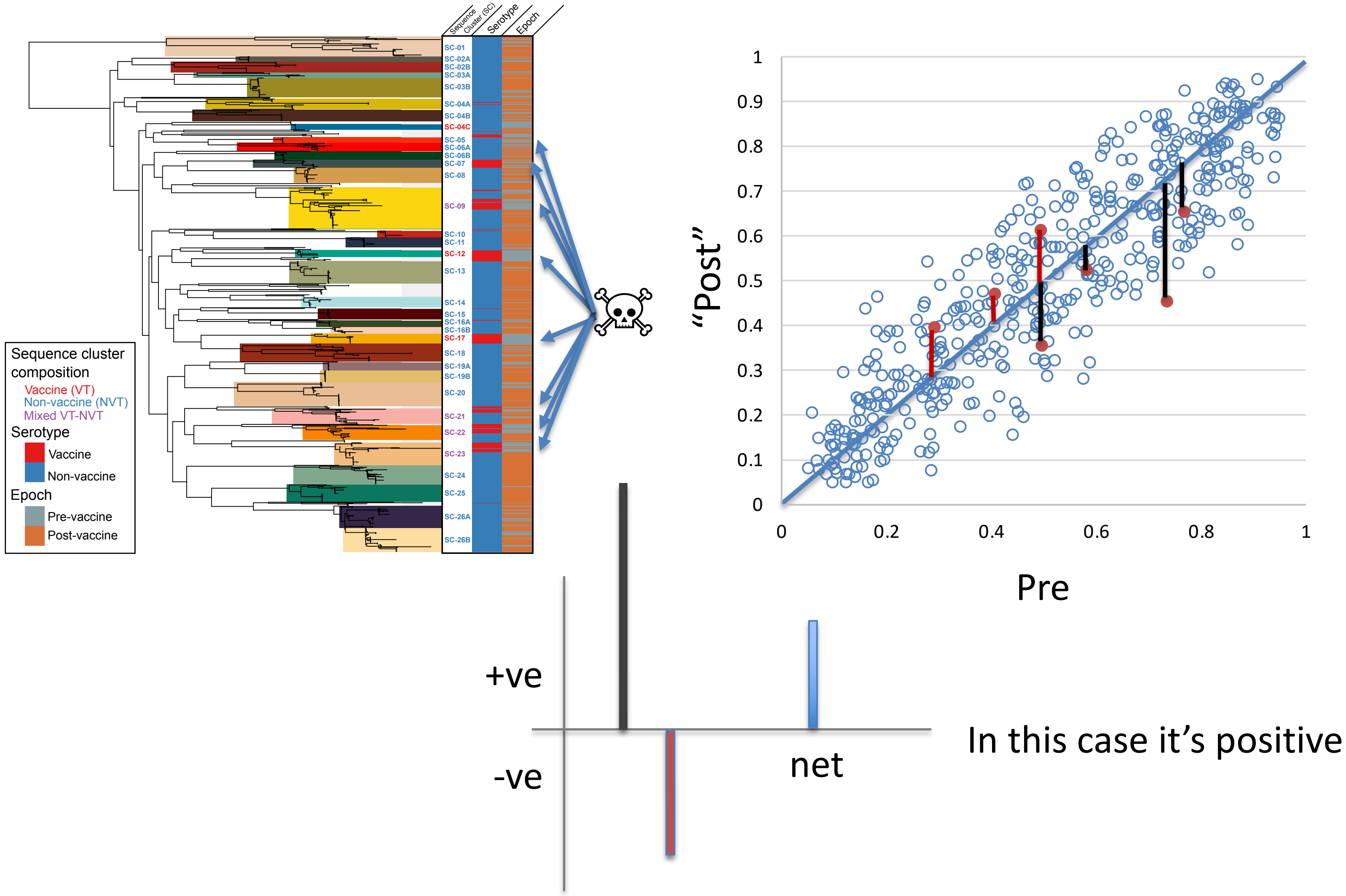




Identify the accessory genes that are in it

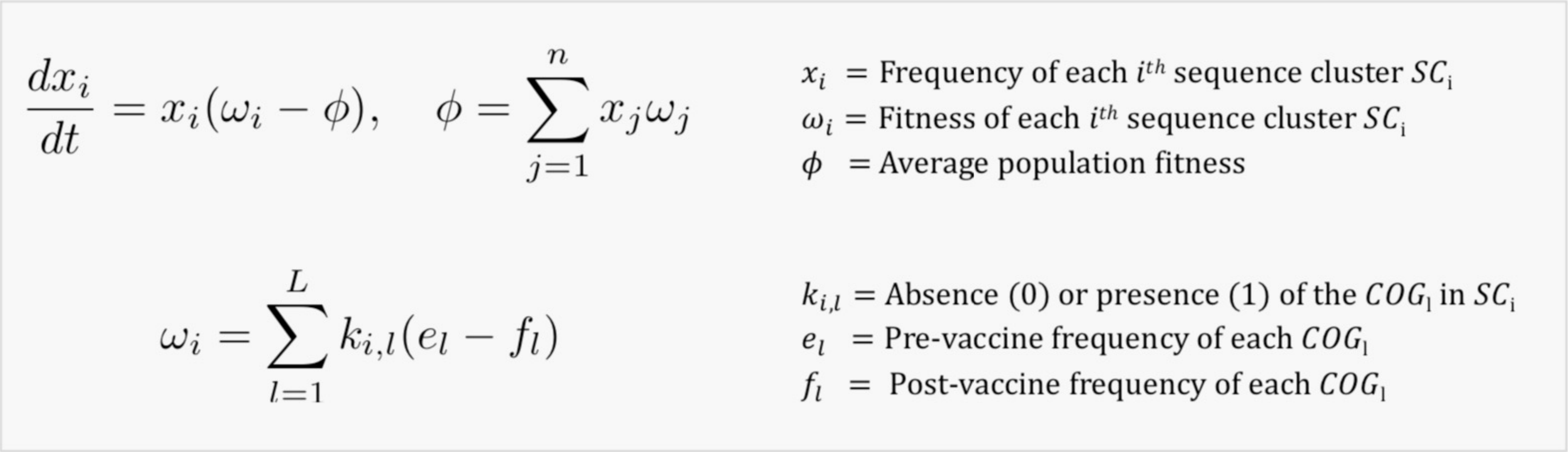
How far away are they from their pre-vaccine frequency?

Combine to find the net fitness



In this case it's positive

The replicator equation



$$\frac{dx_i}{dt} = x_i(\omega_i - \phi), \quad \phi = \sum_{j=1}^{n} x_j \omega_j$$

x_i = Frequency of each i^{th} sequence cluster SC_{i}

ω_i = Fitness of each i^{th} sequence cluster SC_{i}

ϕ = Average population fitness

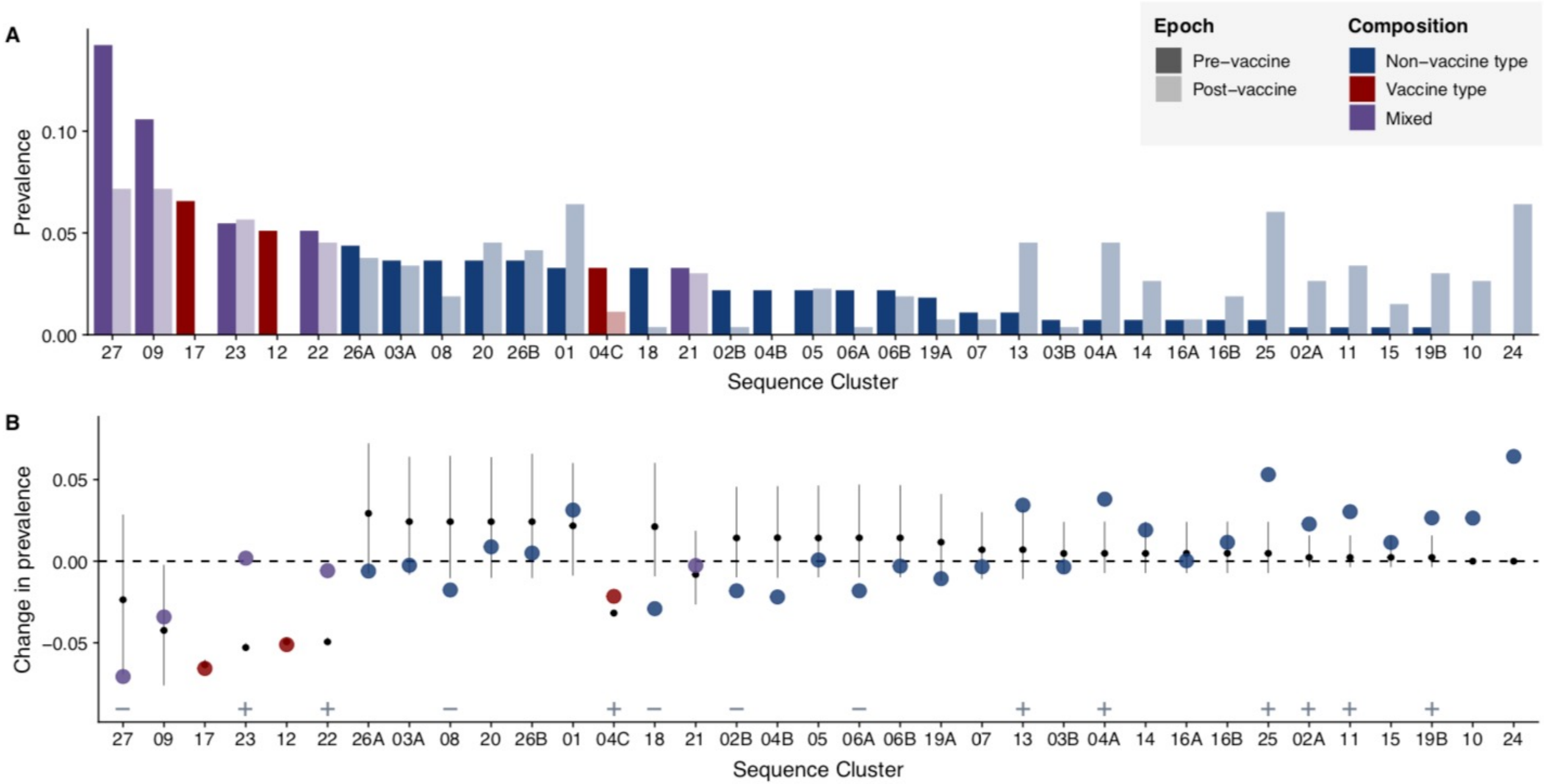
$$\omega_i = \sum_{l=1}^{L} k_{i,l}(e_l - f_l)$$

$k_{i,l}$ = Absence (0) or presence (1) of the COG_{l} in SC_{i}

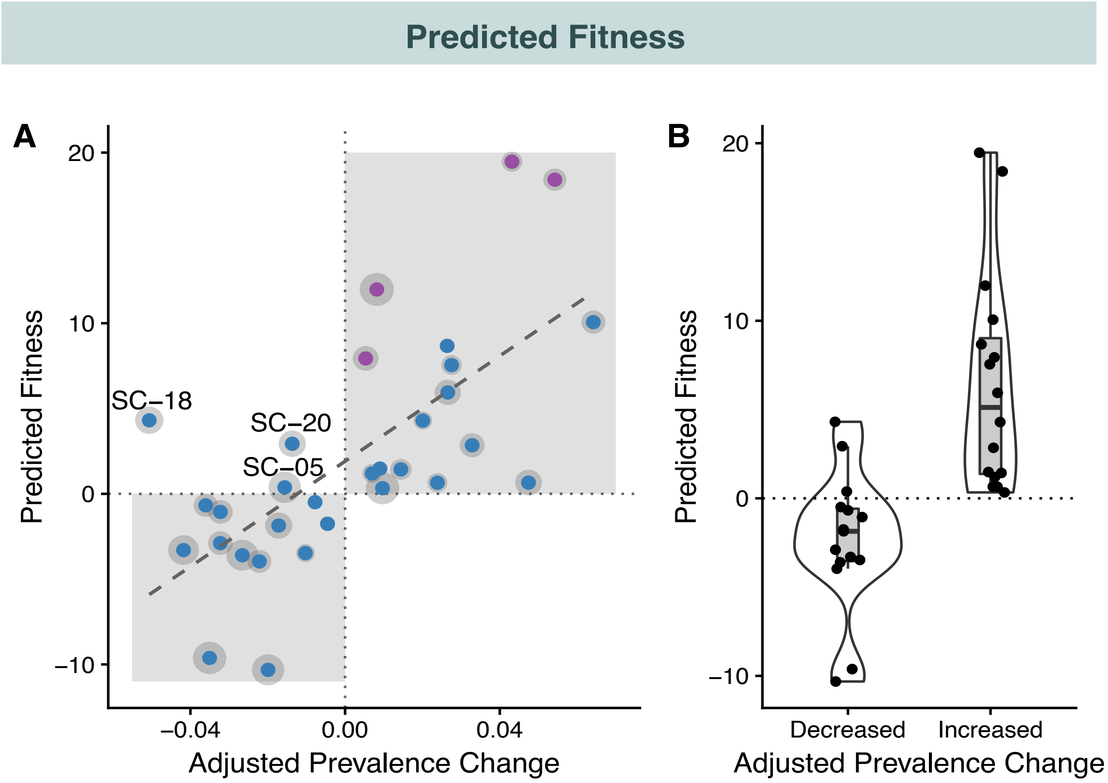
e_l = Pre-vaccine frequency of each COG_{l}

f_l = Post-vaccine frequency of each COG_{l}

Changes in the prevalence of sequence clusters following vaccine



Predicted fitness compared with prevalence change post vaccine



Two strains were not present in the population at the first time point

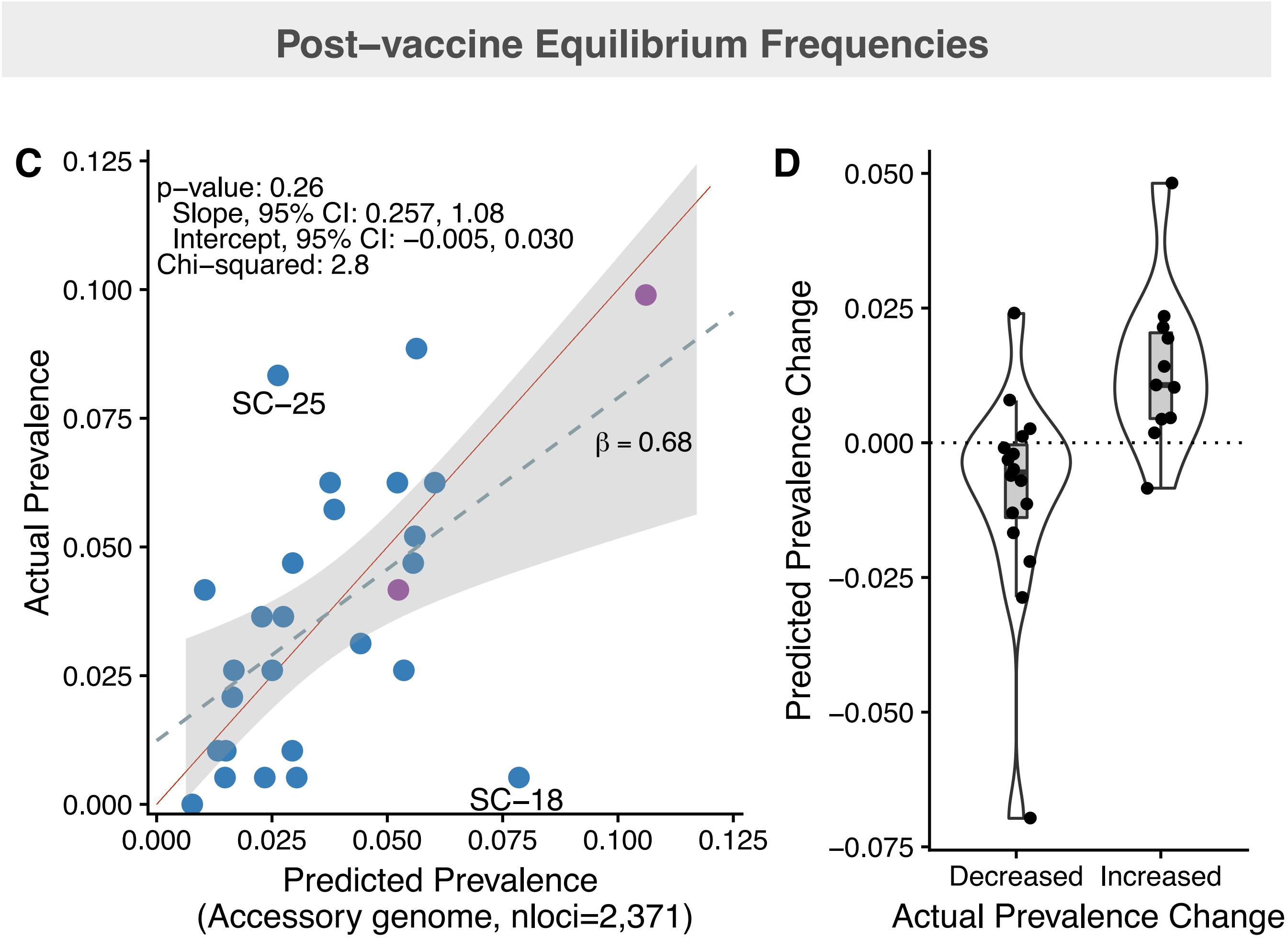
Their standardized predicted fitness can nevertheless be calculated

SC10 - 8.67

SC24 – 10.07

Quadratic programming

- What is the expected equilibrium?
- Not the same as the predicted fitness
- Quadratic programming can be used to estimate the optimum frequencies of each SC
- Again, assuming the pre vaccine *status quo* is a proxy for the selected frequency of each gene


Post–vaccine Equilibrium Frequencies
C
p–value: 0.26
Slope, 95% CI: 0.257, 1.08
Intercept, 95% CI: –0.005, 0.030
Chi–squared: 2.8
SC–25
SC–18
β = 0.68
Actual Prevalence
0.125
0.100
0.075
0.050
0.025
0.000
0.000
0.025
0.050
0.075
0.100
0.125
Predicted Prevalence
(Accessory genome, nloci=2,371)
D
Predicted Prevalence Change
0.050
0.025
0.000
–0.025
–0.050
–0.075
Decreased
Increased
Actual Prevalence Change

PLOS BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Frequency-dependent selection can forecast evolution in *Streptococcus pneumoniae*

Taj Azarian ✱ ✉, Pamela P. Martinez ✱, Brian J. Arnold, Xueting Qiu, Lindsay R. Grant, Jukka Corander, Christophe Fraser, Nicholas J. Croucher, Laura L. Hammitt, Raymond Reid, Mathuram Santosham, Robert C. Weatherholtz, Stephen D. Bentley, [...], William P. Hanage ✱ [view all]

Acknowledgments

CENTER *for* COMMUNICABLE DISEASE DYNAMICS

Taj Azarian
Pamela Martinez
Brian Arnold
Marc Lipsitch
Alanna Callendrello

Jonathan Finkelstein
Jukka Corander
Nick Croucher

wellcome trust sanger institute

Ste Bentley

CENTER FOR AMERICAN INDIAN HEALTH | JOHNS HOPKINS BLOOMBERG SCHOOL *of* PUBLIC HEALTH

Lindsay Grant
Laura Hammitt
Raymond Reed
Mathuram Santosham
Robert Weatherholz
Kate O'Brien

2 postdocs available! whanage@hsph.harvard.edu
@BillHanage

MIDAS

NATIONAL INSTITUTES OF HEALTH