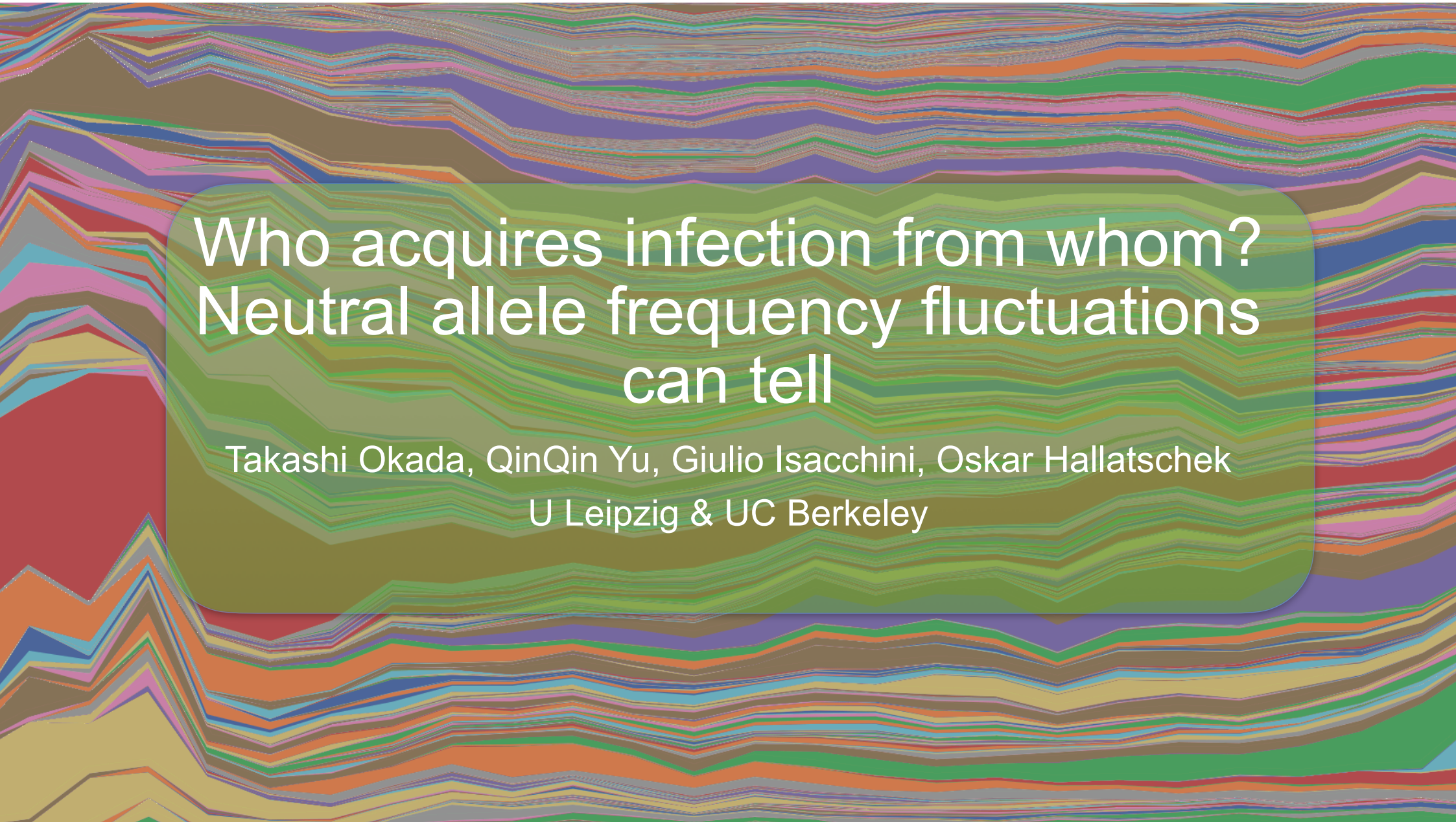


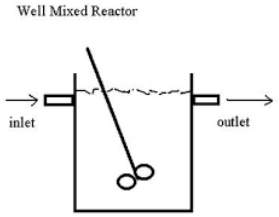
Who acquires infection from whom? Neutral allele frequency fluctuations can tell

Takashi Okada, QinQin Yu, Giulio Isacchini, Oskar Hallatschek

U Leipzig & UC Berkeley

Modeling spreading processes

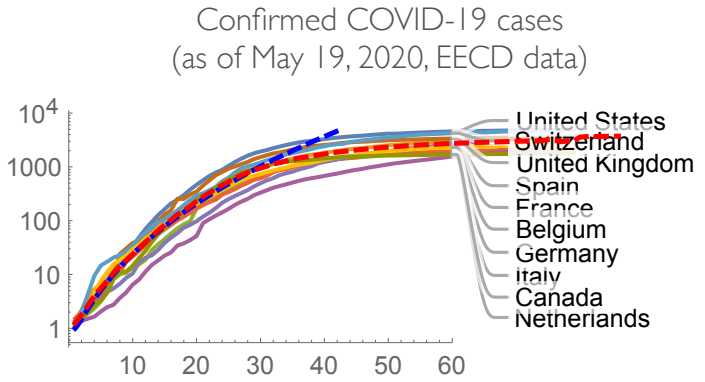
In the simplest models, the abundance $X(t)$ of a new variant obeys



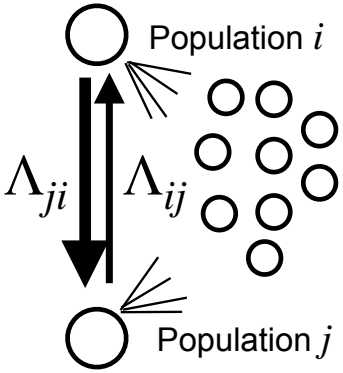
$$\partial_t X = r(t)X \quad + \sqrt{X/N_e}\,\eta\ ,$$

Growth

Demographic noise



… in structured models:

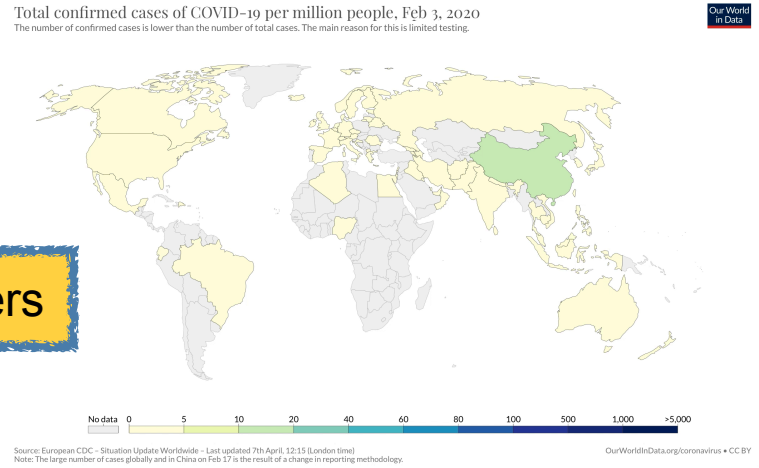


$$\partial_t X_i = \sum_j \Lambda_{ij}(t)X_j + r(t)X_i + \sqrt{X_i/N_e}\,\eta\ ,$$

Migration

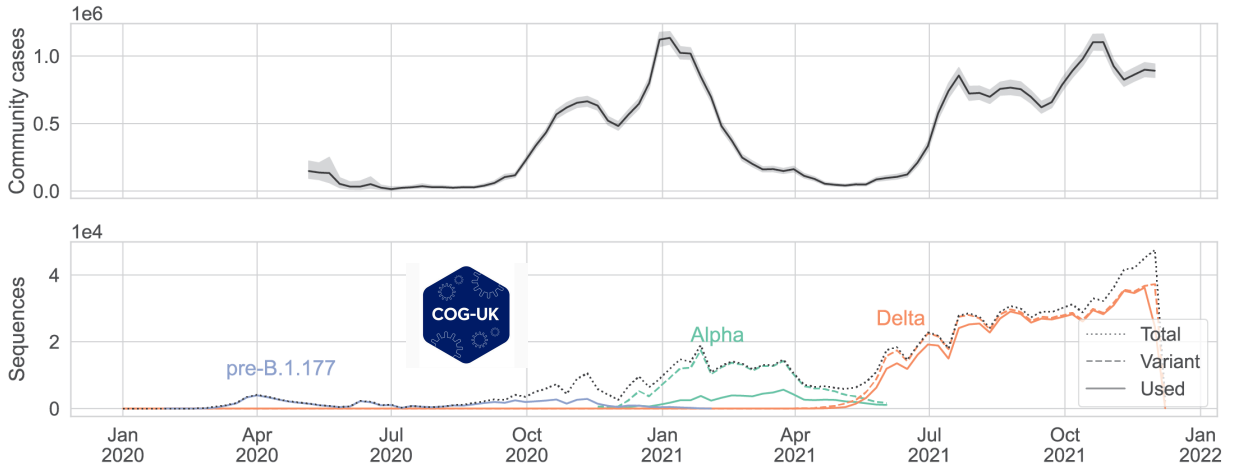
Forecasting/backcasting relies on good parameters

How can we directly measure epidemiological parameters?



Our focus: Sars-CoV-2 in England

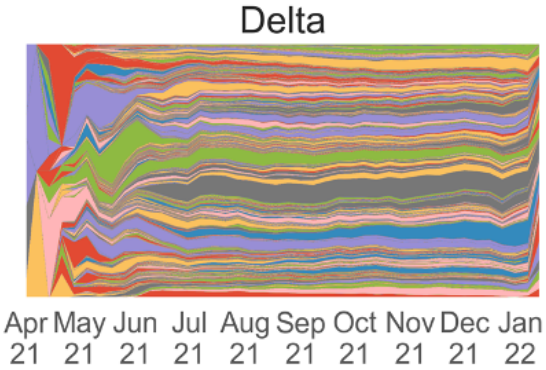
- England heavily invested in COVID sequencing (big thanks to COG-UK!!!)
- Strangely but conveniently, data shows long-lasting plateaus of high incidents rates (especially for Delta)
- There are lots of ≈neutral variants, and they fluctuate a lot.
- Similar time series arise in barcoding experiments, metagenomics, …



Hypothesis: Neutral fluctuations can tell us about demographic noise and migration ….

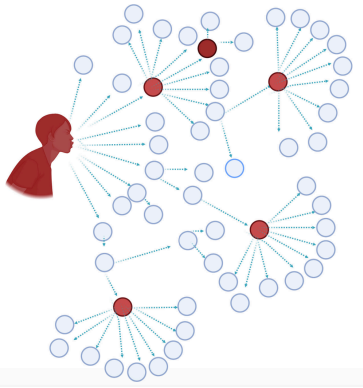
Learning from neutral allele frequency fluctuations

1. Quantify fluctuation strength. Consistent with SEIR or super spreaders?



$$Z \sim NB(R_t, k)$$

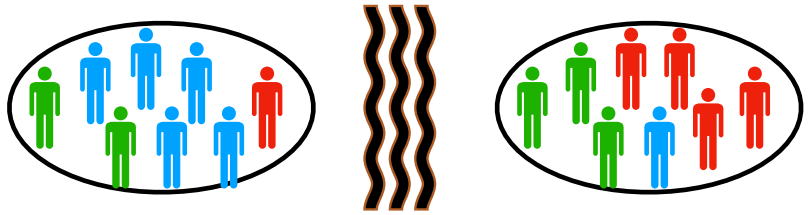
PDF of infected often modeled as a Negative Binomial
Lloyd-Smith et al, Nature, 2005



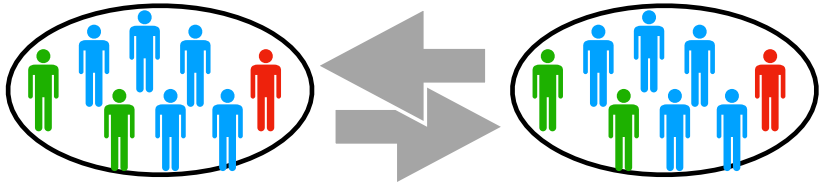
Lakdawala and Menachery, Trends in Microbiology, 2021

Excess fluctuations might hint at emergence of jackpot effects
(Qinqin Yu, et al, bioRxiv 2022.11.21.517390)

2. Compare fluctuations in different groups of people. Correlations reflect epidemiological coupling (infection *rates*).

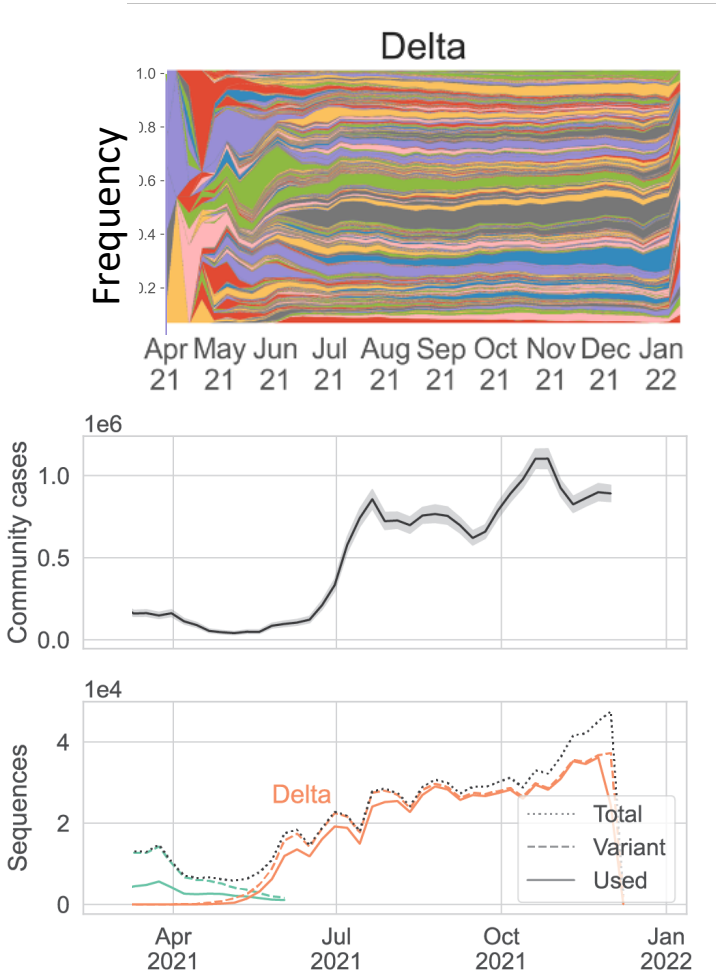


Independent fluctuations



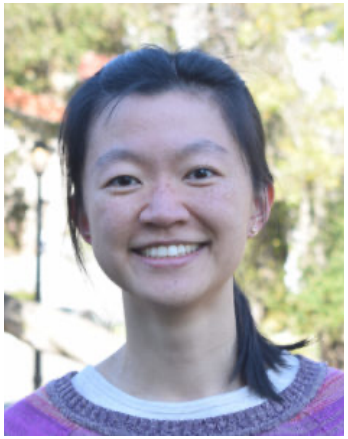
Frequency fluctuations are correlated

Inferring the strength of genetic drift



- The variation due to genetic drift **adds over time**
- The variation due to sampling biases **does not add over time**

→ Use this signal to infer genetic drift and sampling biases



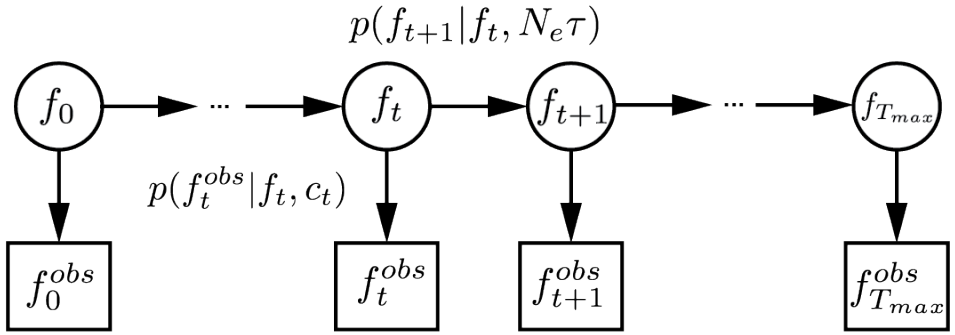
QinQin Yu

Genetic drift $\mathrm{var}[f_{t+\Delta t} - f_t] = \frac{f_t(1-f_t)}{N_e}\Delta t$

Sampling bias $\mathrm{var}[f_t^{obs} - f_t] = \frac{c_t}{M_t} f_t$

f_t = lineage frequency
f_t^{obs} = observed lineage frequency
$N_e(t)$ = time-dependent effective population size
c_t = strength of sampling bias (c_t = 1, random sampling)
M_t = number of sequences

Hidden Markov Model of frequency time series



Inferring sampling noise:

$$\{c_t, c_{t+1}, ..., c_{t+T}\}^{inf} = \underset{\{c_t, c_{t+1}, ..., c_{t+T}, N_e\tau\}}{\arg\min} \Big[\ln \sum_{t_1, t_2 = t}^{t+T} \frac{(\kappa_{t_1,t_2}^{obs} - \kappa_{t_1,t_2}(c_{t_1}, c_{t_2}, N_e\tau))^2}{\Delta\kappa_{t_1,t_2}^{obs}} \Big]$$

$$\kappa_{t_1,t_2}(c_{t_1}, c_{t_2}, N_e\tau) = \mathrm{var}[p(f_{t_1}^{obs}|f_{t_1}, c_{t_1})] + \mathrm{var}[p(f_{t_2}^{obs}|f_{t_2}, c_{t_2})] + \mathrm{var}[p(f_{t_2}|f_{t_1}, N_e\tau)]$$

- f_t = superlineage (*) frequency
- f_t^{obs} = observed superlineage frequency
- N_e = effective population size
- τ = generation time
- M_t = number of sequences
- c_t = strength of sampling bias (c_t = 1, random sampling)

(*) We create ”superlineages” by combining lineages together until they reach a threshold number of counts

Maximize likelihood function to determine most likely $N_e\tau$

$$\{N_e\tau\}_t^{inf} = \underset{N_e\tau}{\arg\max} \Big[\int_0^1 d\vec{f} p(f_{t-\frac{T}{2}}^{obs}|f_{t-\frac{T}{2}}, c_{t-\frac{T}{2}}) \prod_{t=t-\frac{T}{2}+1}^{t+\frac{T}{2}} p(f_t^{obs}|f_t, c_t^{inf}) p(f_t|f_{t-1}, N_e\tau) \Big]$$

Transition probability from genetic drift

$$\phi_t \equiv \sqrt{f_t}$$

$$p(\phi_{t+1}|\phi_t) = \mathcal{N}(\phi_t, \frac{1}{4N_e\tau})$$

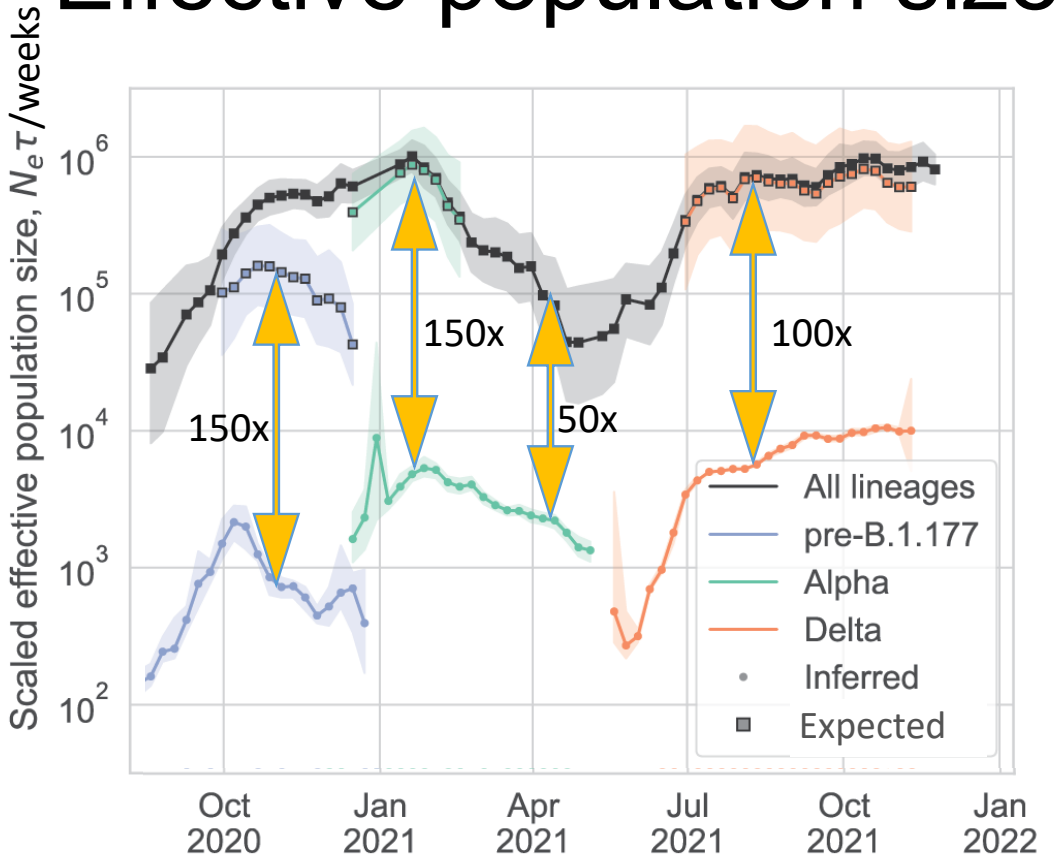
Emission probability from sampling bias

$$\phi_t^{obs} \equiv \sqrt{f_t^{obs}}$$

$$p(\phi_t^{obs}|\phi_t) = \mathcal{N}(\phi_t, \frac{c_t}{4M_t})$$

Related: Jonathan P Bollback, Thomas L York, and Rasmus Nielsen. Genetics 179.1 (2008), pp. 497–502.
Anna Ferrer-Admetlla et al. Genetics 203.2 (2016), pp. 831–846.

Effective population size in England across time



Expectation from SEIR model:

$$N_e\tau = \frac{(E+I)^2}{2R_t\gamma_I I} \approx 2\frac{I}{\gamma_I} \approx 3I\,\text{weeks}$$

$$I \approx \text{const.}$$

I = number of infected individuals
E = number of exposed individuals
R_t = effective reproduction number
$1/\gamma_I$ = average time from infection to recovery

Volz et al, Genetics, 2009
Frost et al, Phil. Trans. R. Soc. B, 2010

The inferred $N_e\tau$ is **much lower** than expected!

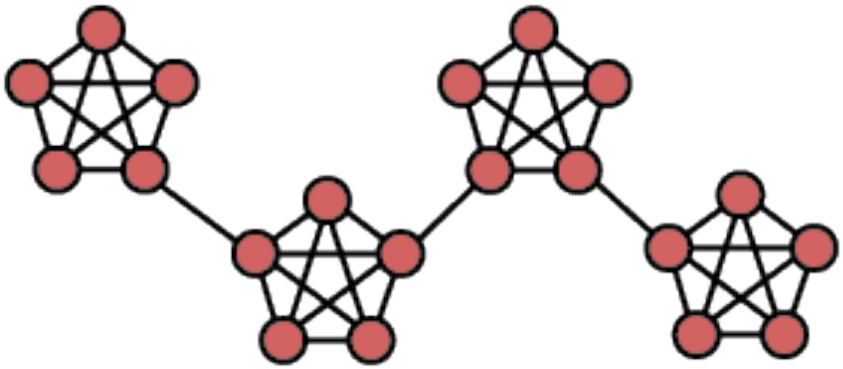
(N.B. Sampling noise changed over time, and was sometimes different between variants)

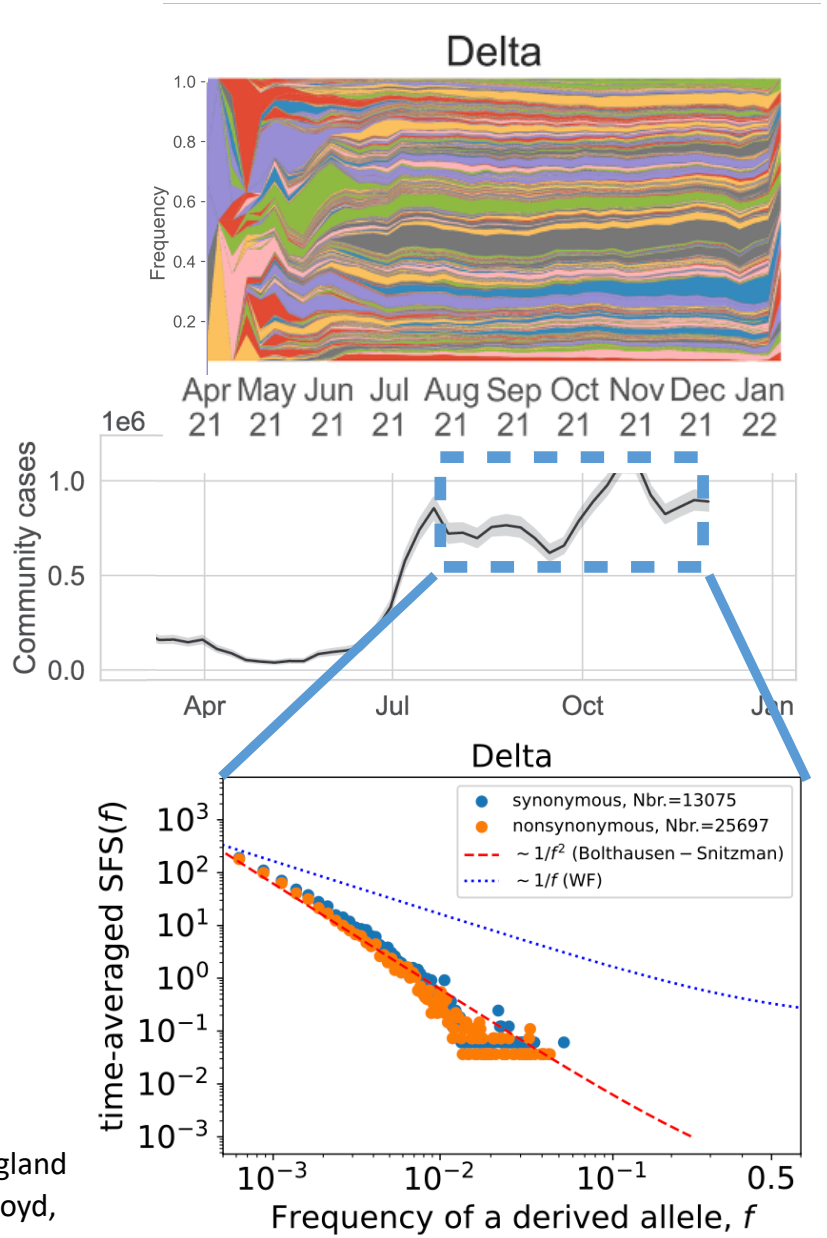
Learning from neutral allele frequency fluctuations

1. Quantifying demographic noise

Effective population size much smaller than expected [by O(10^2)], which cannot be explained by the impact of super spreaders alone.

… might indicate hidden structure

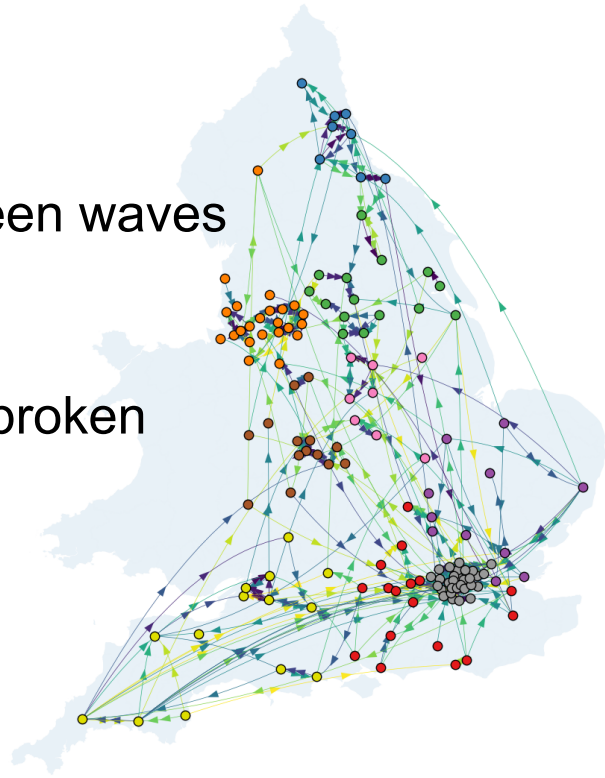




Lineage frequency time series reveal elevated levels of genetic drift in SARS-CoV-2 transmission in England
QinQin Yu, Joao Ascensao, Takashi Okada, The COVID-19 Genomics UK (COG-UK) consortium, Olivia Boyd, Erik Volz, Oskar Hallatschek, (2022) bioRxiv 2022.11.21.517390

Conclusions

- Correlated fluctuations encode interactions
- Interactions reflect geography but differ substantially between waves
- Relaxation time of 8 weeks for Delta in England
  (=twice as fast as for alpha!)
- Long-range connections matter; detailed balance partially broken



Outlook

- How do these couplings compare to estimates from mobility data or from mobility proxies (cell phone) ?
- Applications where we don't have mobility proxies:
  - **Infection matrix between age groups, ethnicities, …**
  - Metagenomics of microbiomes (natural & experimental)
  - "Historical" DNA - Recombination creates lot's of uncorrelated time series

Acknowledgments

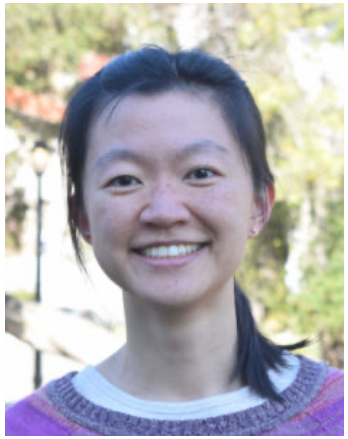
Takashi Okada
Visiting Scholar (snow associate professor at Kyoto University)

Lead on inferring interactions from time series

Giulio Isacchini
Postdoc

Extension to recombining genomes / ancient DNA

see Poster #



QinQin Yu
Now Postdoc @ Harvard Longwood

Lead on analyzing the strength of genetic drift

Joao Ascensao
PhD Candidate in Bioengineering

Methods development

Olivia Boyd
PhD Student
Imperial College London

Phylogenetics of SARS-CoV-2

Erik Volz
Advisor

Special thanks to:
Mike Boots
Katia Koelle
Priya Moorjani
Daniel Reeves
Erik Volz
Daniel Weissman

Hallatschek Lab
Jonas Denk
Alison Feder
Giulio Isacchini
Yuya Karita
Stephen Martis
Aditya Prasad
Valentin Slepukhin
Christian Westendorf