Predicting the (very near) future: forecasting pathogen evolution

Molecular Epidemiology of Infectious Diseases

Lecture 11

April 13th, 2020

Most of the approaches we've considered are retrospective... can we say anything about the future?

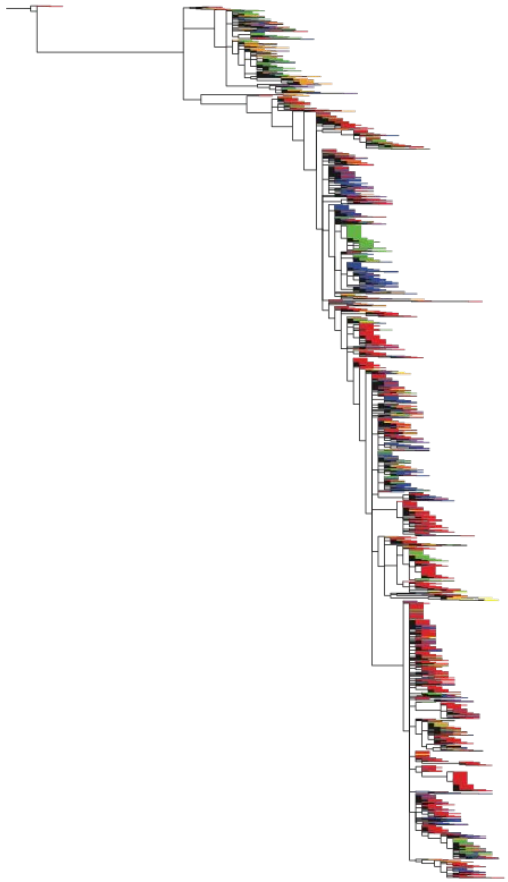
“No scientific theory is worth anything unless it enables us to predict something which is actually going on. Until that is done, theories are a mere game of words, and not such a good game as poetry”

J.B.S Haldane (Adventures of a Biologist, 1937)

Influenza A (H3N2)

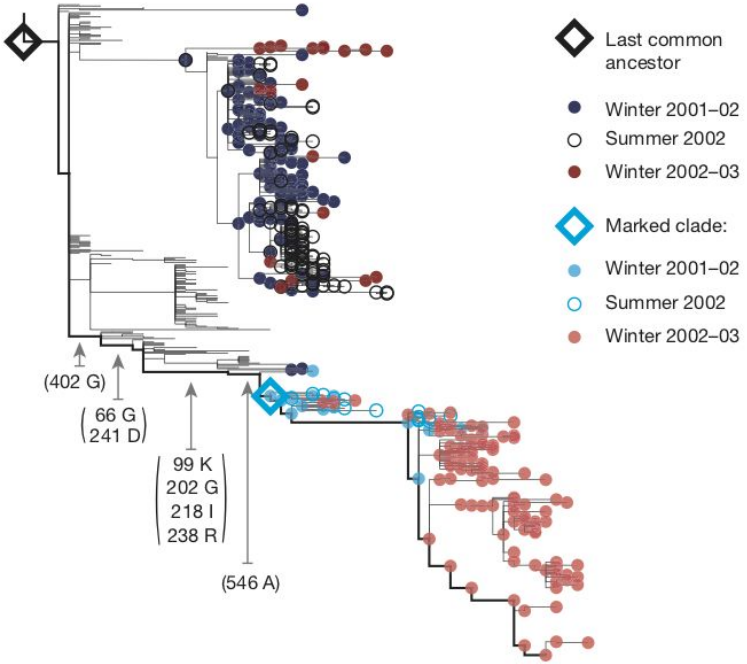
New antigenic variants periodically replace older strains:

- New antigenic variants emerge and escape antibody-based immunity against earlier strains.
- **Antigenic drift** leads to a ladder-like structure with a trunk lineage
- Flu vaccines need to be updated yearly to avoid antigenic mismatch.



Russell *et al.* (Science, 2008)

Forecasting short-term flu evolution



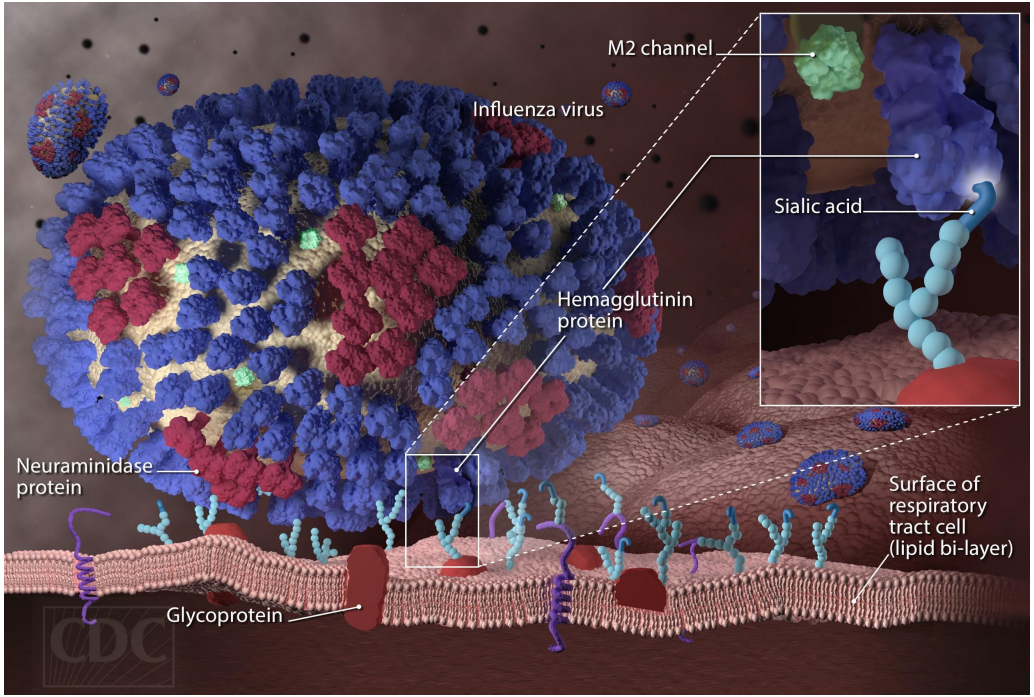
Consider the evolution dynamics of different influenza *clades*

The frequency X_v of a particular clade can be predicted based on the fitness f_i of individual strains i in a clade:

$$\hat{X}_v(t+1) = \sum_{i:v,t} x_i \exp(f_i)$$

Luskza & Lassig (Nature, 2014)

Influenza hemagglutinin and cell entry



Forecasting short-term flu evolution

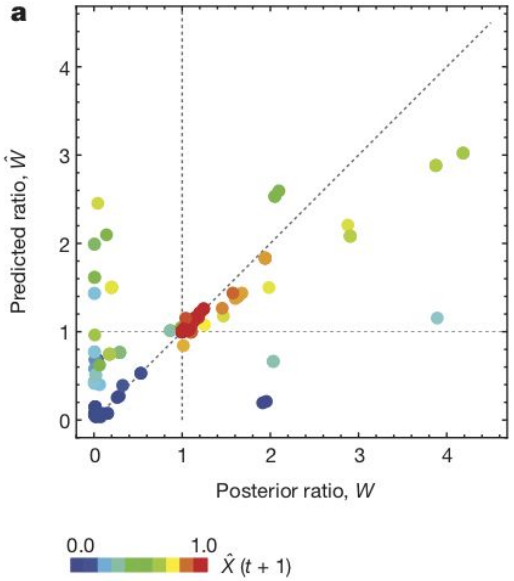
Luskza & Lassig (2014) consider two main factors that influence the fitness $\boldsymbol{f_i}$ of a strain:

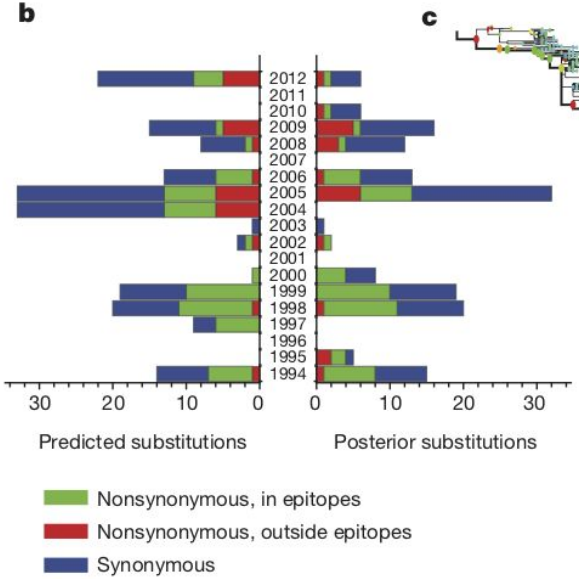
1) The amplitude of cross-immunity $\mathbf{C(a_i,a_j)}$ between strain *i* and all other strains *j* that have previously circulated in the host population
2) The fitness cost $\mathbf{L(a_i)}$ of deleterious mutations at non-antigenic sites

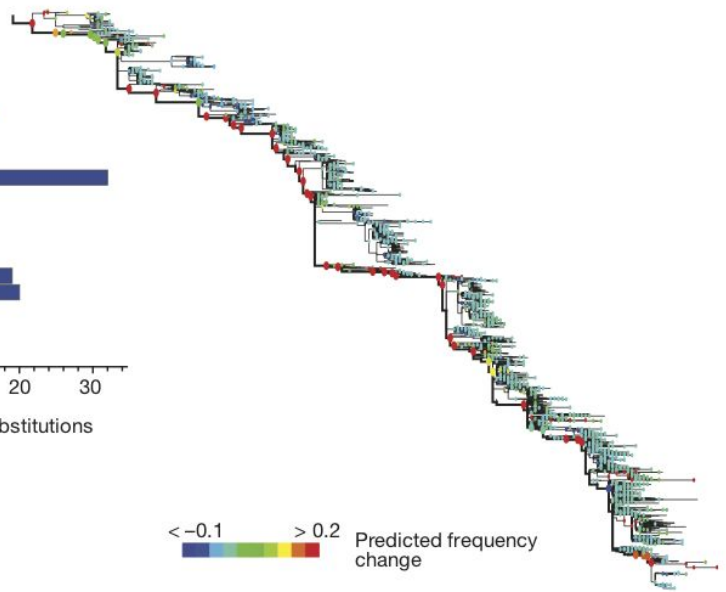
Their overall fitness mapping function is:

$$f_i = f_0 - \mathcal{L}(\mathbf{a}_i) - \sum_{j:\, t_j < t_i} x_j \, \mathcal{C}(\mathbf{a}_i, \mathbf{a}_j)$$

Forecasting short-term flu evolution





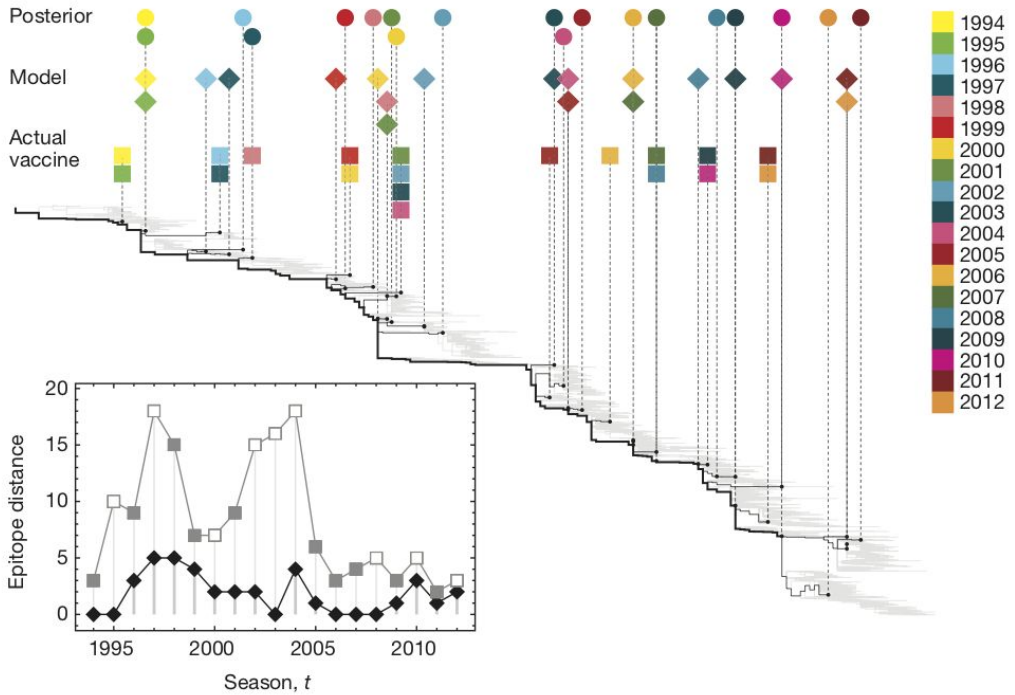


$$W_v = \frac{X_v(t+1)}{X_v(t)}$$

Luskza & Lassig (Nature, 2014)

Forecasting short-term flu evolution

Evolutionary predictions can aid design of vaccines with optimal immunity to dominant strains in the next flu season.



Luskza & Lassig (Nature, 2014)

Can we predict pathogen evolution more generally?

What do we need to know?

What mutations/genotypes are available?

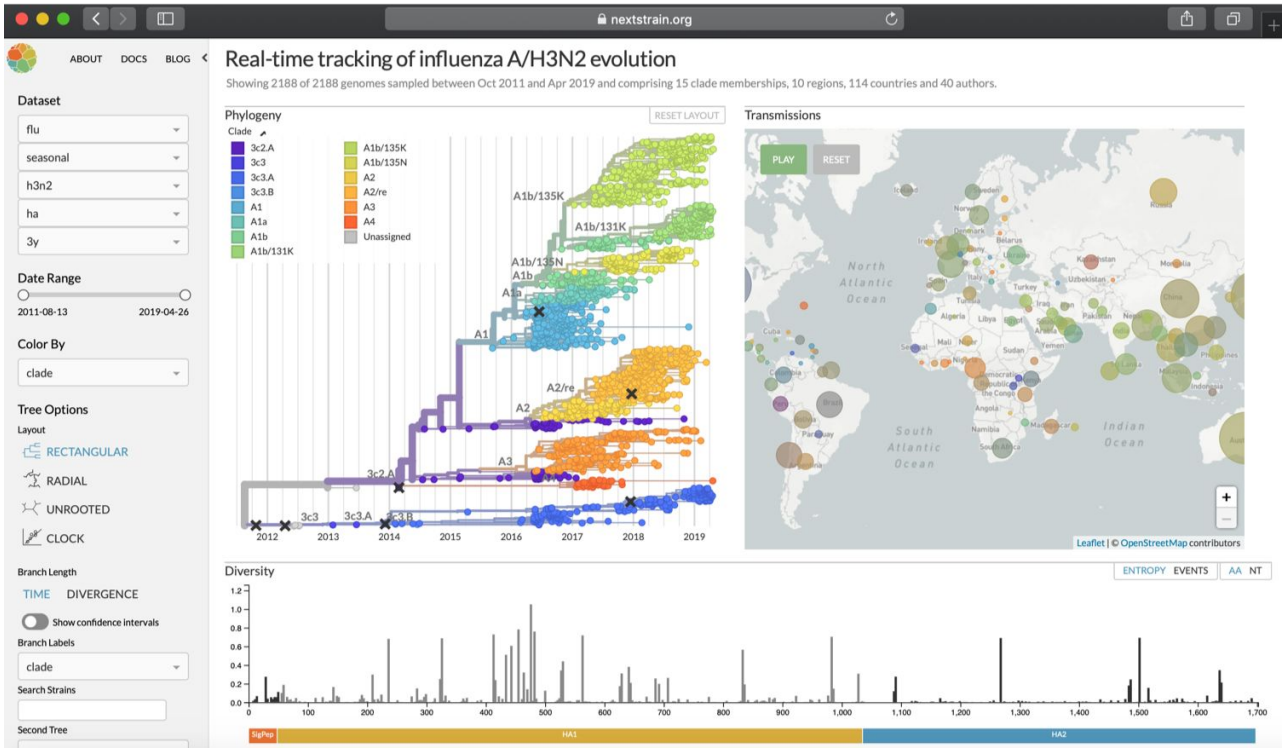
How do genotypes map to fitness-related phenotypes?

How does fitness translate to epidemic potential at the population level?

Mutational limits on prediction

At the very least, we need to know what mutations/genotypes are in a population to be able to predict anything about evolution

Genomic surveillance



Mutational limits on prediction

At the very least, we need to know what mutations/genotypes are in a population to be able to predict anything about evolution

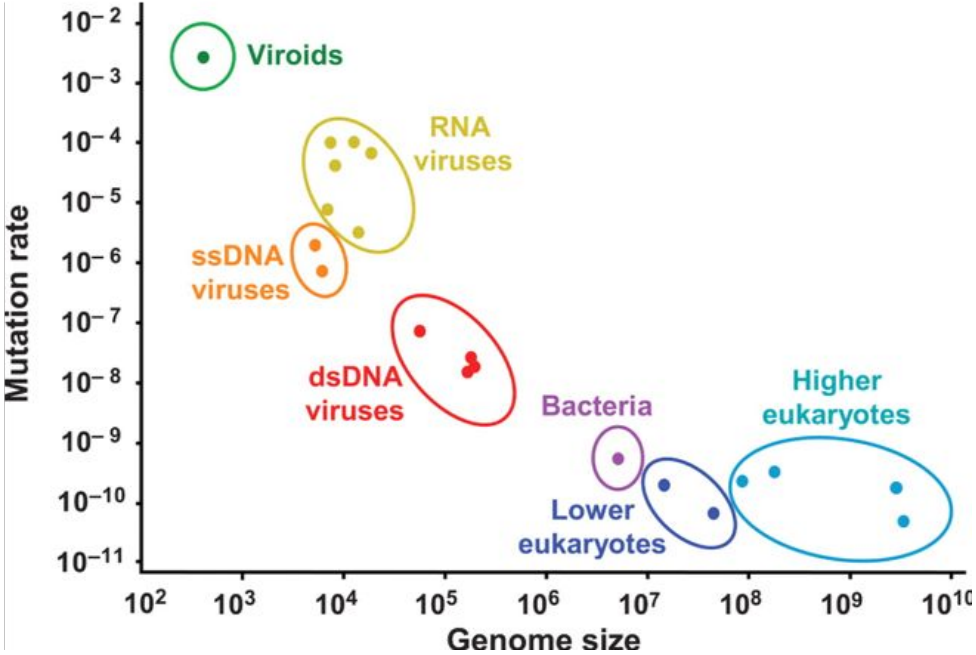
Meaningful predictions are probably limited to short-term predictions about standing genetic variation (or immediately accessible mutations).

Long-term predictions are limited by the stochastic nature of the mutation process and what mutations will enter a population

Rapidly mutating microbes

Microbial evolution is often not mutation limited - high mutation rates and large population sizes often ensure that all possible mutations occur on relatively short timescales.

Evolutionary predictions may be extended to all locally accessible genotypes

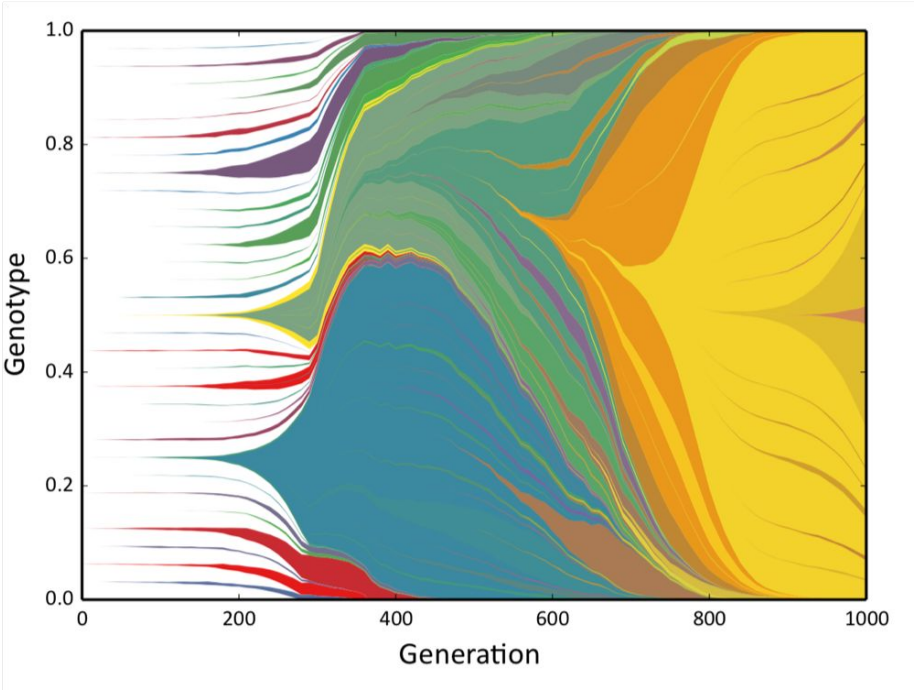


Gago et al. (Science, 2009)

Clonal interference

Clonal interference arises in large asexual populations with high mutations rates.

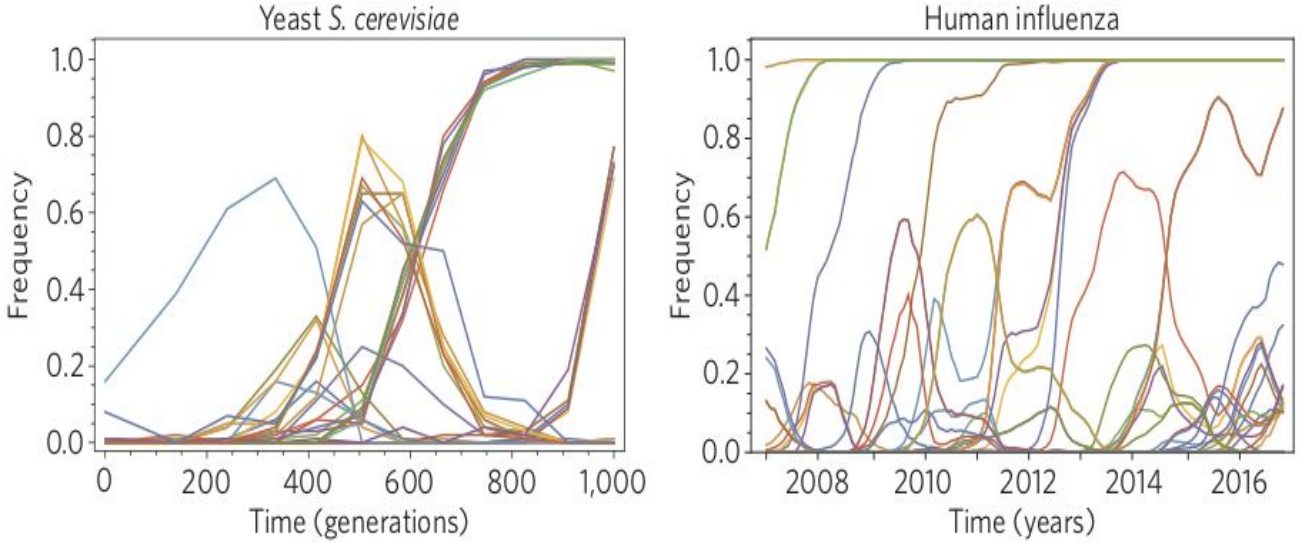
Multiple lineages with beneficial mutations compete with one another.



Cvijovic et al. (Trends in Genetics, 2018)

Clonal interference

Clonal interference is a common feature of many microbial populations:



Lässig et al. (Nat. Ecol Evol, 2017)

Clonal interference

Clonal interference enhances overall predictability:

Increases odds of evolution finding most fit genotype even if this requires multiple mutations.

Role of genetic drift becomes negligible.

Increases chances that “best” genotype with the largest fitness advantages goes to fixation.

What do we need to know?

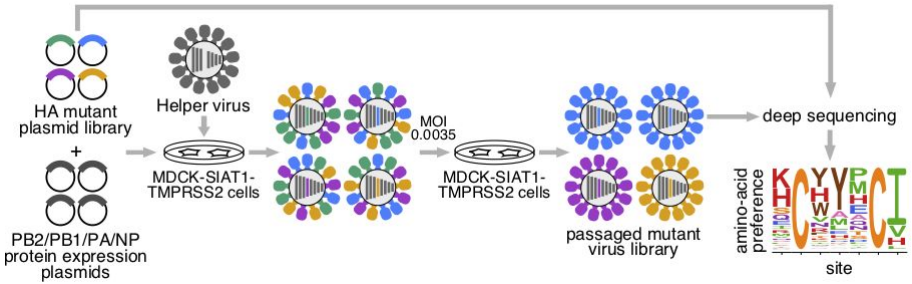
What mutations/genotypes are available?

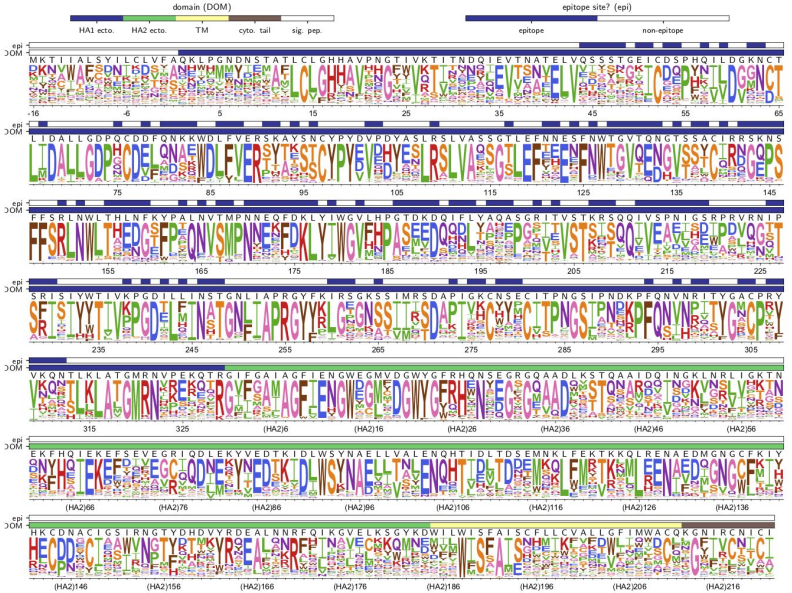
How do genotypes map to fitness-related phenotypes?

How does fitness translate to epidemic potential at the population level?

Deep mutational scanning

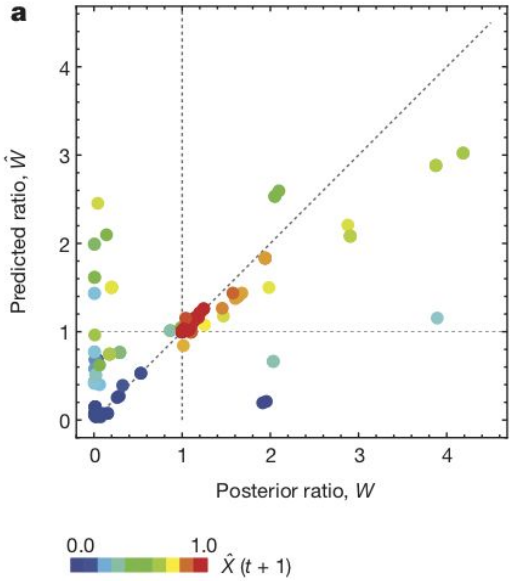
Reverse genetics approaches can be used to systematically explore the genotype to phenotype map using large libraries of mutants.

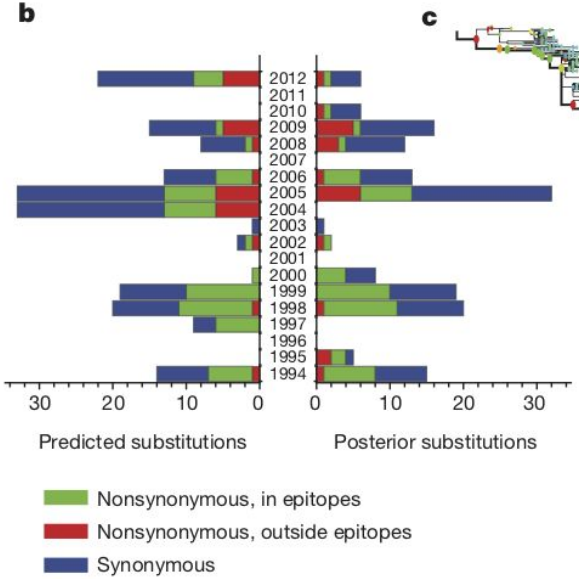


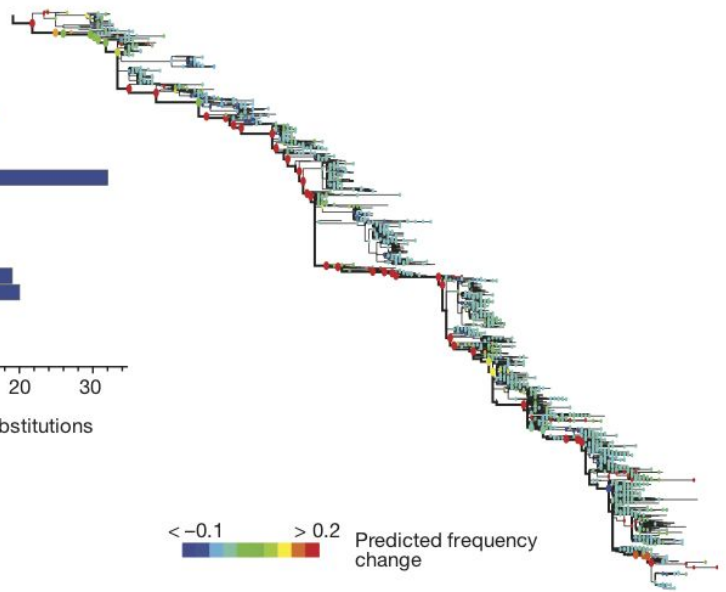


Lee et al. (PNAS, 2018)

Forecasting short-term flu evolution







$$W_v = \frac{X_v(t+1)}{X_v(t)}$$

Luskza & Lassig (Nature, 2014)

Context dependence

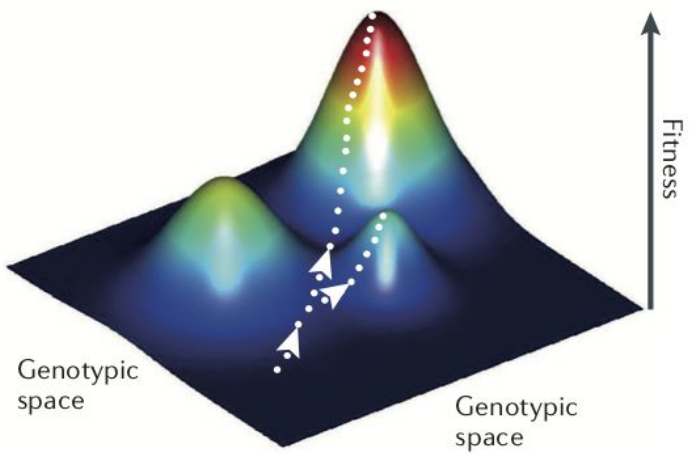
How predictable phenotypes are based on genotypes largely depends on whether phenotypes are context dependent:

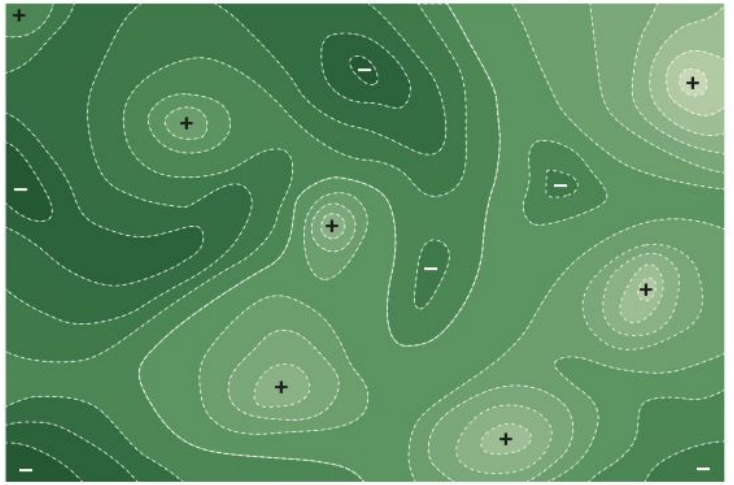
Epistasis: dependence on genetic background including interactions among mutations

Pleiotropy: the effects of mutations on multiple traits or the same trait across different environments.

Epistasis in fitness landscapes

Epistasis largely controls the smoothness/ruggedness of the fitness landscape.

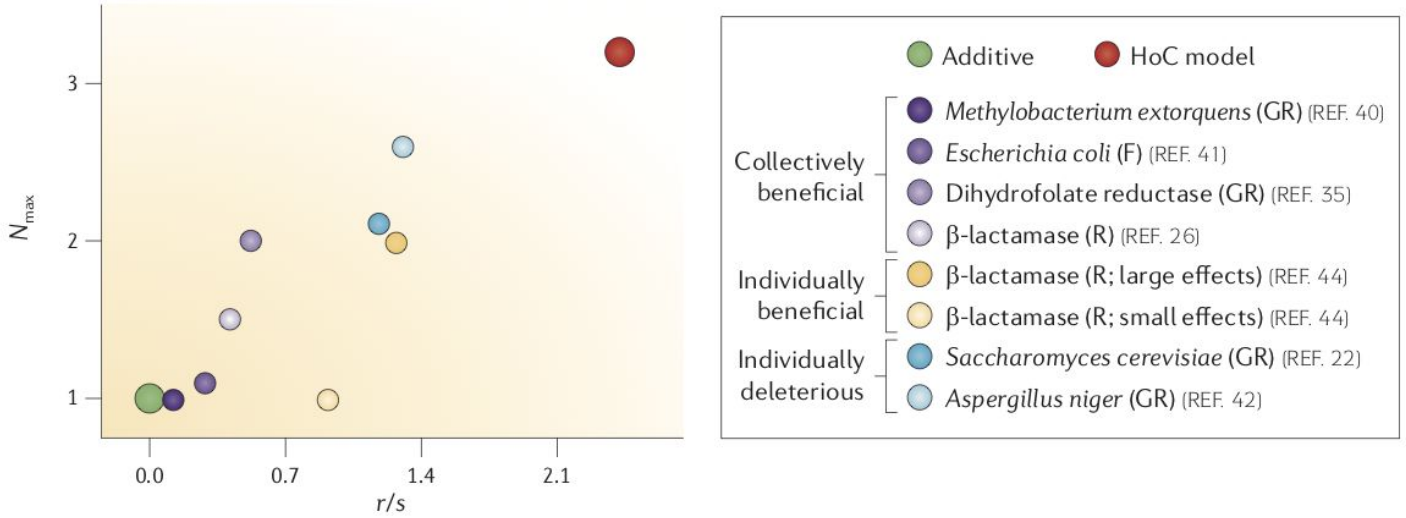




de Visser *et al*. (Nature Rev. Genetics, 2014)

Epistasis in fitness landscapes

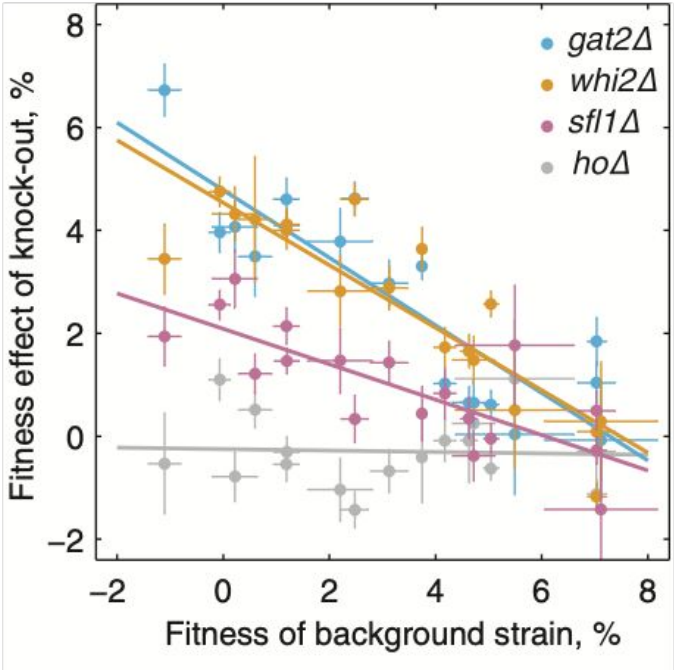
How rugged are empirical fitness landscapes?



de Visser *et al*. (Nature Rev. Genetics, 2014)

Global epistasis

Mutations often exhibit ***global epistasis*** where their fitness effects depend on starting fitness but are “independent of the specific identify of mutations present in the background”.

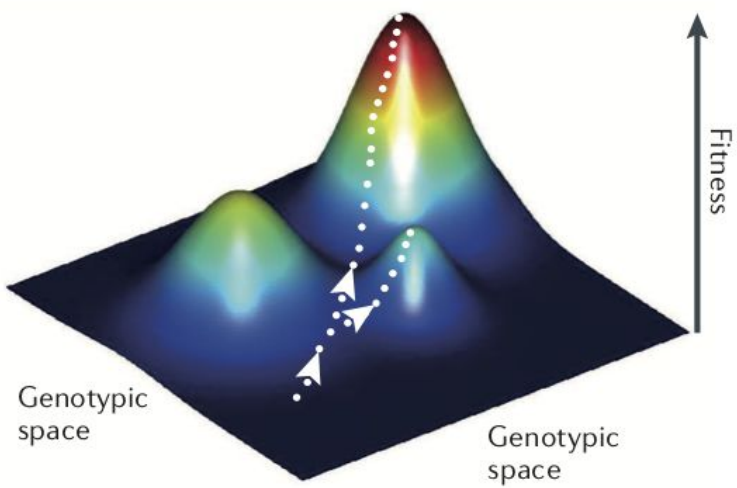


Kryazhimskiy *et al*. (Science, 2014)

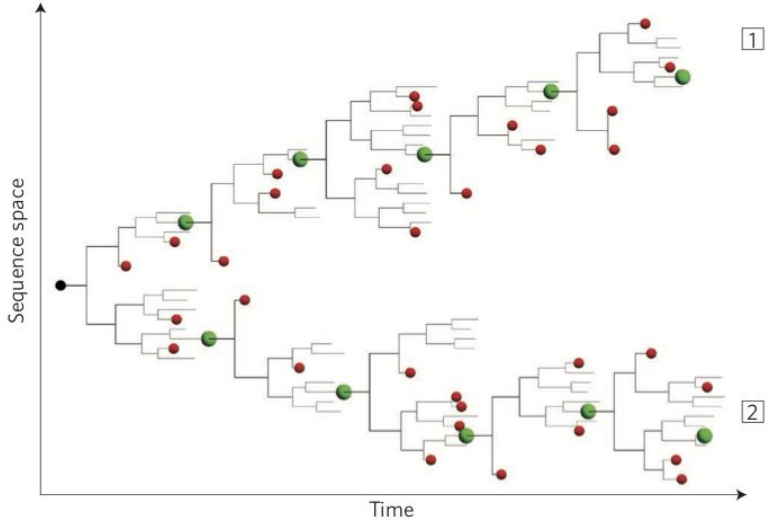
Global epistasis

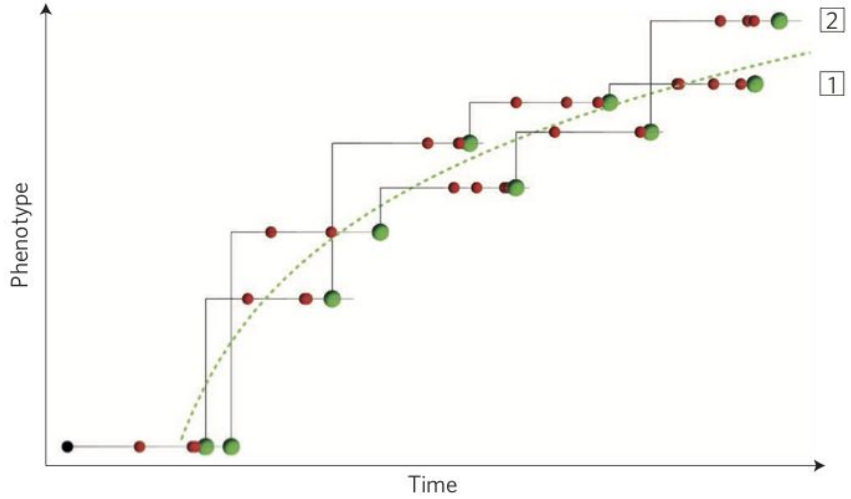
Mutations often exhibit ***global epistasis*** where their fitness effects depend on starting fitness but are “independent of the specific identify of mutations present in the background”.

This is often seen as “diminishing returns” on the effects of beneficial mutations in already fit genotypes.



Can we predict phenotypic evolution?





Lässig et al. (Nat. Ecol Evol, 2017)

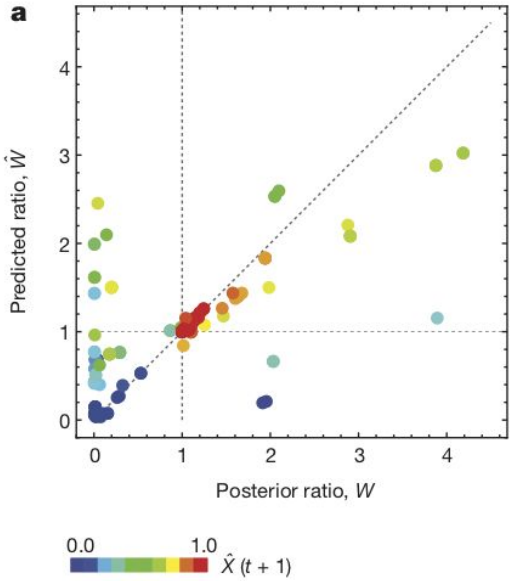
What do we need to know?

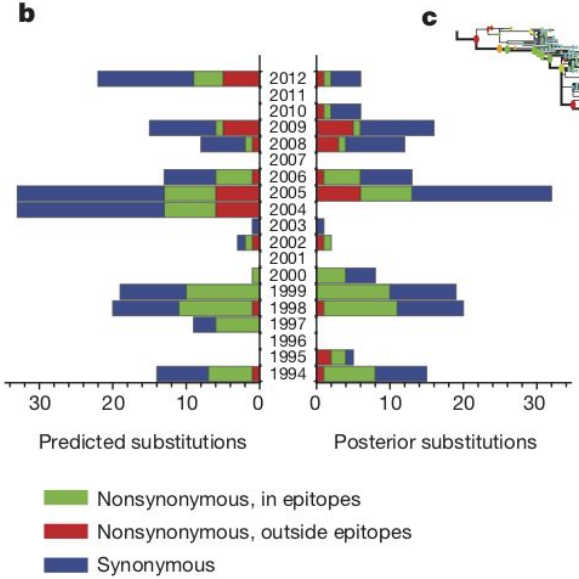
What mutations/genotypes are available?

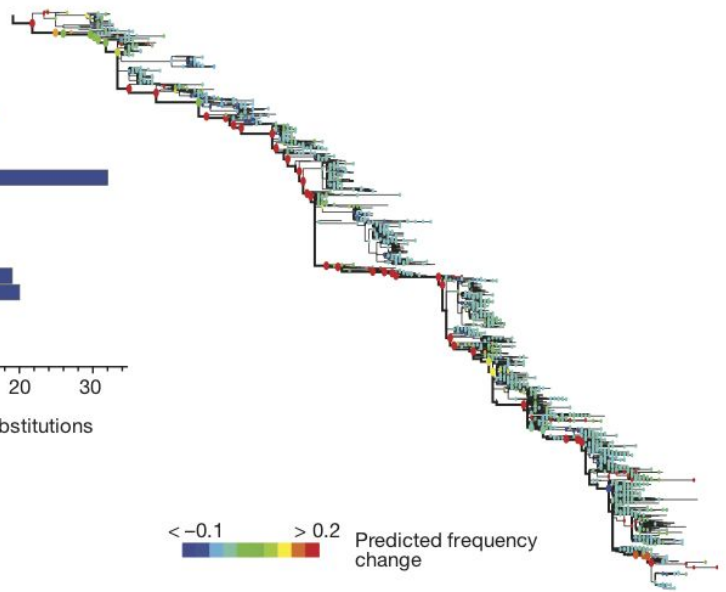
How do genotypes map to fitness-related phenotypes?

How does fitness translate to epidemic potential at the population level?

Forecasting short-term flu evolution







$$W_v = \frac{X_v(t+1)}{X_v(t)}$$

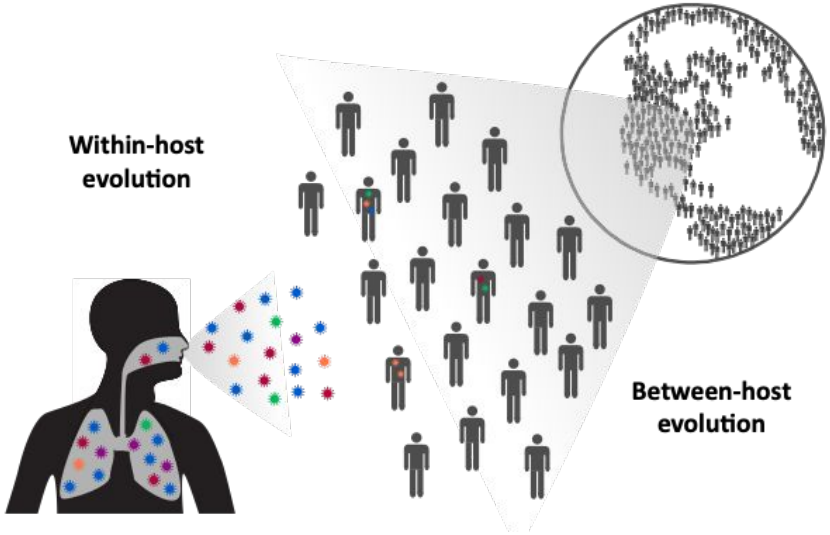
Luskza & Lassig (Nature, 2014)

“Any prediction of evolution is essentially an estimate of fitness differences between strains”

Luksza & Lassig (2014)

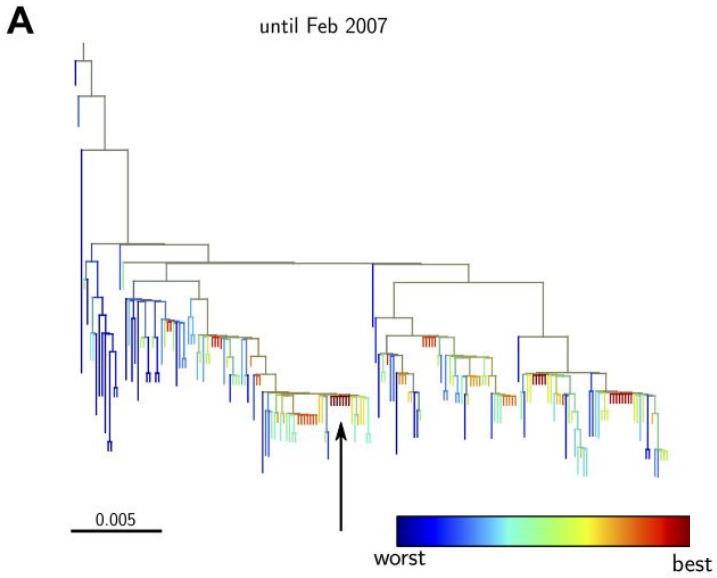
Translating between scales

To make accurate predictions we need to know how pathogen phenotypes related to within-host fitness translate to population-level fitness between hosts.



Predicting evolution from tree shape

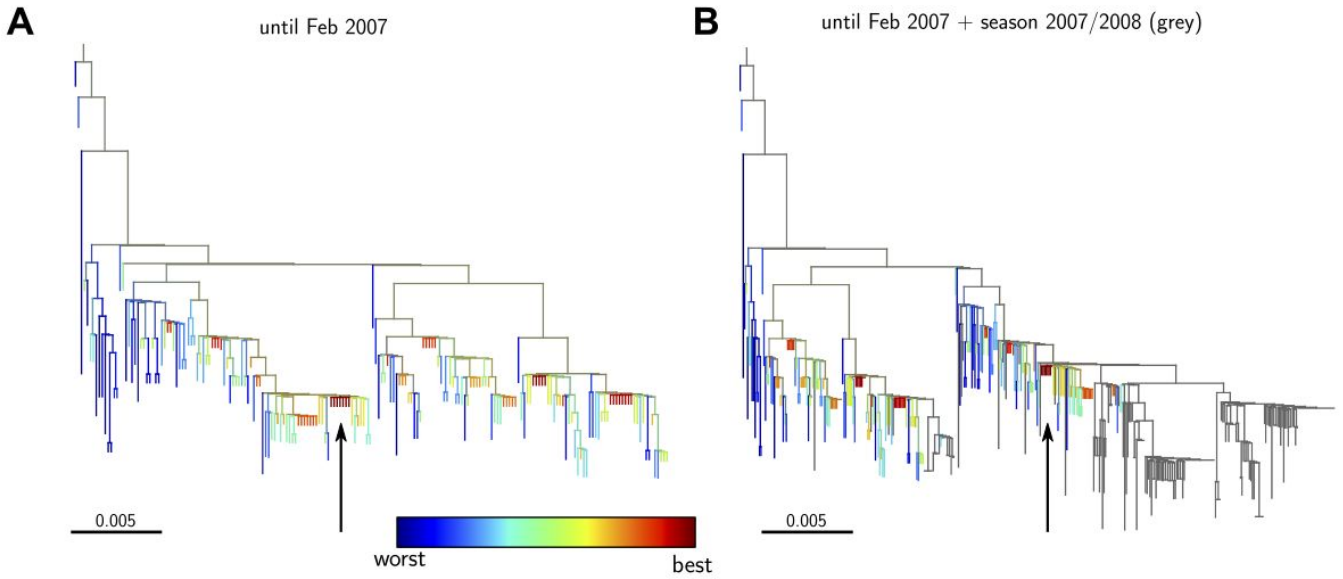
Branching rates in pathogen phylogenies correlate strongly with fitness



Neher et al. (*eLife*, 2014)

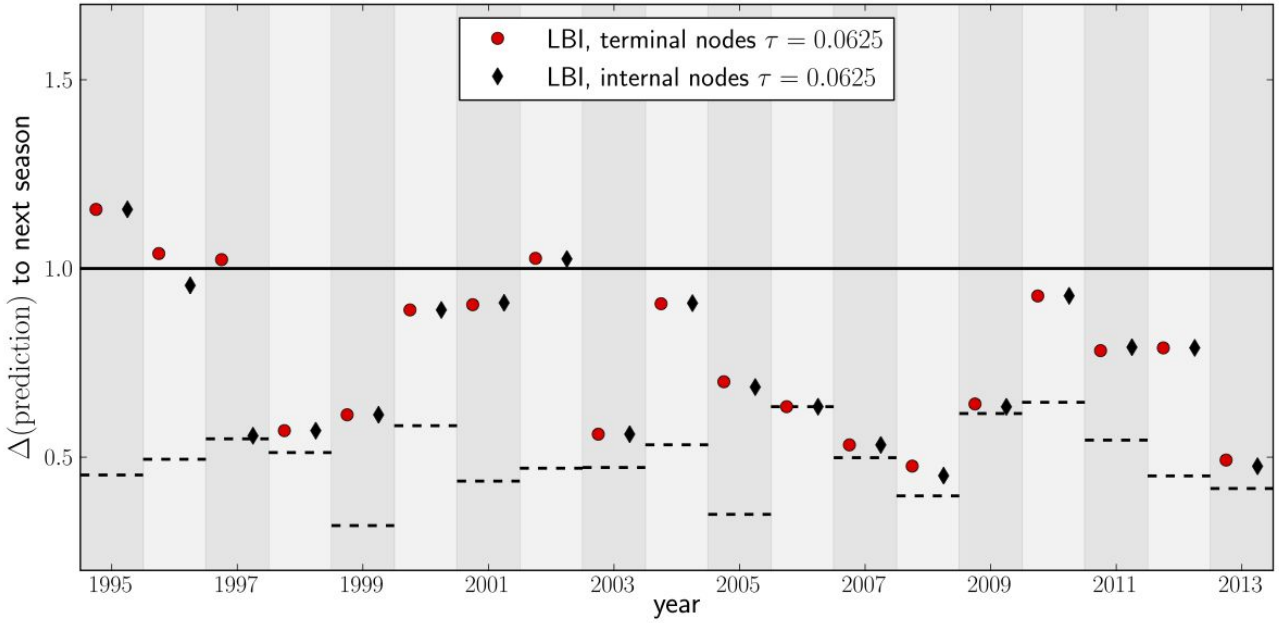
Predicting evolution from tree shape

Branching rates in pathogen phylogenies correlate strongly with fitness



Neher et al. (*eLife*, 2014)

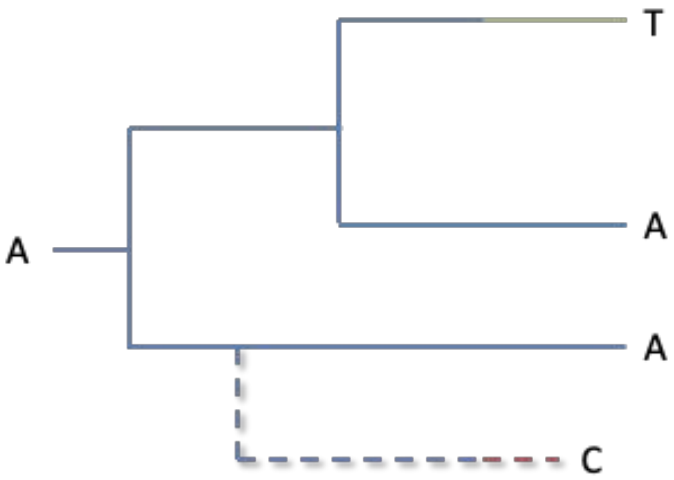
Predicting evolution from tree shape



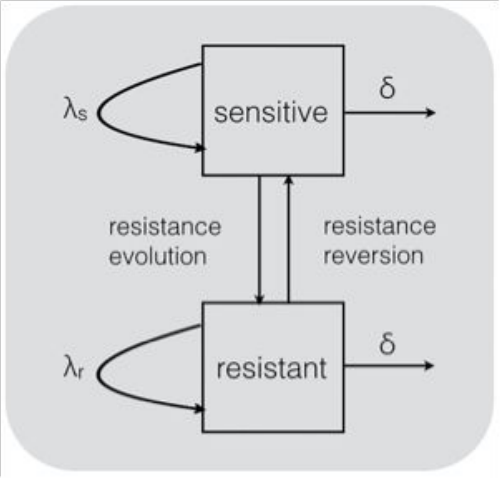
Neher et al. (*eLife*, 2014)

Multi-type birth-death models

Allows for different types of individuals (e.g. genotypes) that can vary in their birth or death rates and therefore their fitness values.



Fitness of HIV drug resistance mutations



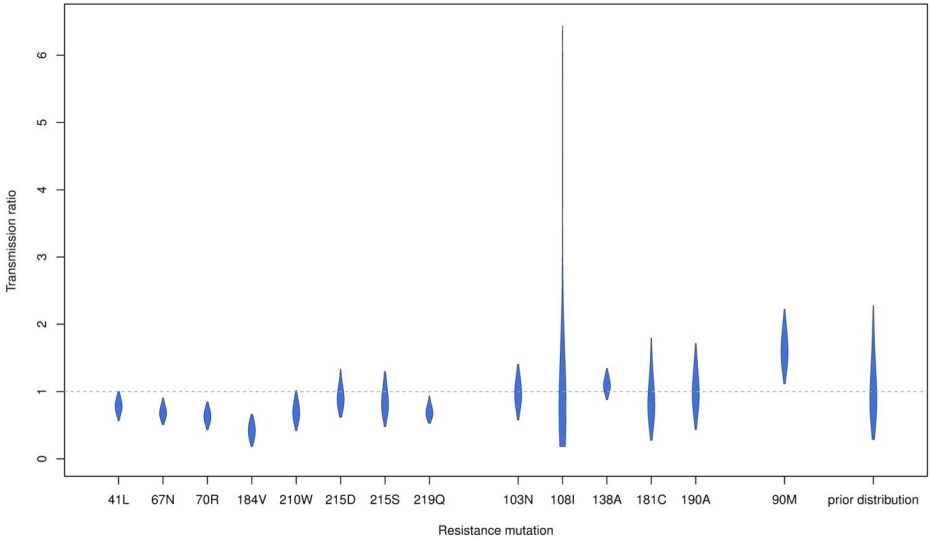
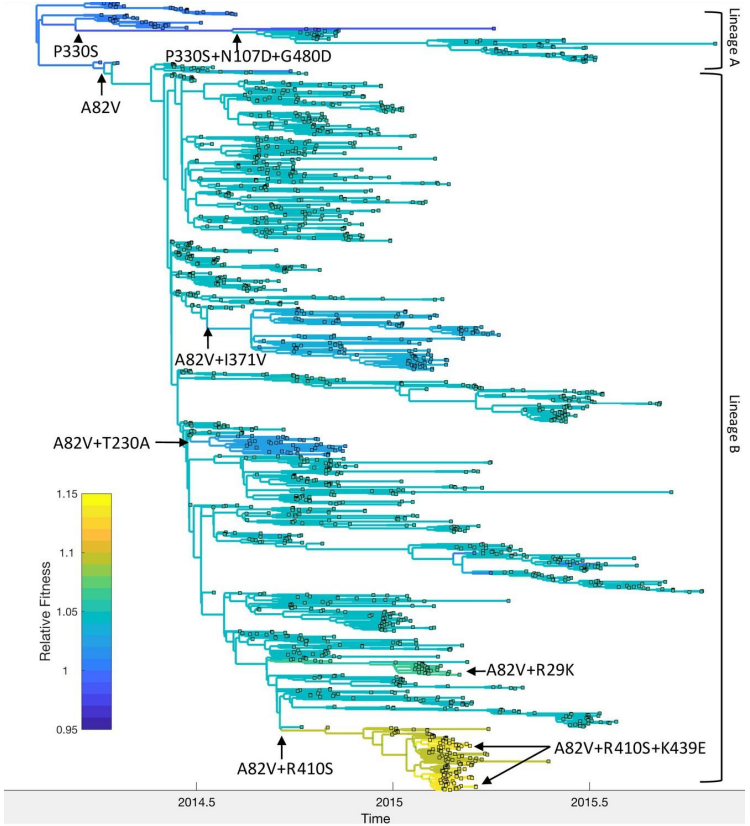


Table 1. Resistance mutations with numbers of corresponding clusters and samples, related drugs and drug usage dates within Switzerland.

	nRTI									NNRTI					PI
Resistance mutation	41L	67N	70R	184V	210W	215D	215S	215Y	219Q	103N	108I	138A	181C	190A	90M
Number (#) of clusters of size ≥ 2	56	23	19	35	18	18	16	25	20	25	10	46	8	8	14
# Sequences in clusters	927	667	712	1011	481	569	494	807	605	725	334	1014	329	311	389
# Resistant samples in clusters	93	39	26	44	26	41	31	28	28	38	11	109	10	12	38
Drug (SHCS drug codes)	AZT D4T	AZT D4T	AZT D4T	3TC ABC FTC	AZT D4T	AZT D4T	AZT D4T	AZT D4T	AZT D4T	NVP EFV	NVP EFV	RPV	NVP EFV ETV RPV	NVP EFV	NFV SQV
Drug usage ≥ 1%	1987	1987	1987	1995.5	1987	1987	1987	1987	1987	1997	1997	2013	1997	1997	1996
Drug usage < 1%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2008

Kühnert *et al.* (PLoS Pathogens, 2018)

Adaptation of Ebola virus to humans



Rasmussen and Stadler (eLife, 2019)

What do we need to know?

What mutations/genotypes are available?

How do genotypes map to fitness-related phenotypes?

How does fitness translate to epidemic potential at the population level?

Analogy: Forecasting the weather

Despite the fact that the physical models required to predict the weather were developed in the 19th century, it still took another hundred years for reliable forecasts to emerge because of the need for massive amounts of atmospheric data and computing power.

But once short-term forecasts could be made, methods could be iteratively tested and improved, and forecasting advanced remarkably quickly.

A brief history of weather forecasting: https://www.newyorker.com/magazine/2019/07/01/why-weather-forecasting-keeps-getting-better

The future of evolutionary predictions

We have the theory, methods and data to predict short-term evolution

- Predictive genotype-to-fitness models
- High-throughput phenotypic data
- Genomic surveillance data and molecular epidemiological methods

We will likely get it wrong many times before we get it right but the fact that we can repeatedly test predictions on short timescales means that we can iteratively and rapidly improve our evolutionary forecasts.

In class discussion on Wednesday

Please read these two papers for class on Wednesday:

Łuksza, M., & Lässig, M. (2014). A predictive fitness model for influenza. Nature, 507(7490), 57-61.

Morris, D. H., Gostic, K. M., Pompei, S., Bedford, T., Łuksza, M., Neher, R. A., ... & McCauley, J. W. (2018). Predictive modeling of influenza shows the promise of applied evolutionary biology. Trends in Microbiology, 26(2), 102-118.

In class discussion on Wednesday

After you read these papers, please think about and be prepared to discuss:

1. How predictable is evolution in your favorite host-pathogen system?
2. What information is needed to make accurate predictions?
3. What is the time horizon of predictability?
4. What factors promote or limit predictability?
5. What is the biggest source of uncertainty surrounding predictions?