# Predicting the evolution of human influenza

Marta Łuksza<sup>1</sup> and Michael Lässig<sup>2</sup>

<sup>1</sup> Biological Sciences, Columbia University <sup>2</sup> Institute for Theoretical Physics, University of Cologne Thanks

#### Influenza:

Natalja Strelkowa (Imperial College) Simone Pompei (Cologne) Ganna Rozhnova (Cologne) Mara Villa (Cologne) Theory of adaptive processes: Ville Mustonen (Sanger) Armita Nourmohammad (Princeton) Stephan Schiffels (Sanger) Gergeli Szollosi (Lyon) Torsten Held (Cologne) Daniel Klemmer (Cologne)





# 1. Predictability in (evolutionary) biology

## A physics experiment







## **Repeatability requires**

- control over initial conditions
- the existence of deterministic dynamical laws.

## **Evolutionary processes**





Repeatability is limited because all biological systems are complex.

#### **Evolutionary processes**





## Repeatability is limited because all biological systems are complex:

• All deterministic equations of motion are governed by interactions.

#### **Evolutionary processes**





influenza A (H3N2) strain tree

# Repeatability

#### **Evolutionary processes**





#### Repeatability is limited because all biological systems are complex.

- All deterministic equations of motion are governed by interactions.
- Adaptive mutations alter the system's dynamics.
- There are large-scale stochastic effects.

# Repeatability



#### Adaptive evolution of E. coli in the mouse gut

shows mutational randomness but repeatable phenotypes [Barroso-Batista PLoS Genetics 2014, see Isabel Gordo's talk].

This can be explained as a generic consequence of quantitative trait evolution

[A. Nourmohammad, T. Held, M.L., Current Opinion in Genetics and Development, 2014].

#### Adaptive evolution of influenza



## repeatably targets HA epitope sites



and leads to antigenic drift [Smith et al. Science 2004].

# Predictability

## **Physics theory**



$$\Delta h = \frac{g}{2}t^2$$





## **Predictability requires**

- control over initial conditions
- learnable deterministic dynamical laws.

# Predictability

#### **Physics theory**



COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. XIII, 001-14 (1960)

#### The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

> EUGENE P. WIGNER Princeton University





- 1. Predictability in (evolutionary) biology
- 2. Influenza's mode of evolution

# Modes and predictability of adaptive evolution

#### **Punctuated selection**:

selective sweeps are rare and independent.



➔ predictions are limited by monitoring.

# Modes and predictability of adaptive evolution

#### Influenza shows clonal interference:

competition of co-existing beneficial variants [N. Strelkowa, M.L., Genetics 2012].



→ predictions require modeling.

#### **Evolution of human influenza**

Strain tree based on hemagglutinin sequences of influenza strains since 1969

synonymous mutations,

allows inference of

nonsynonymous non-epitope mutations,

nonsynonymous epitope mutations.



mutational distance

## Analysis of frequency time-series



Epitope amino acid changes have an increased relative fixation probability:

g / g₀ = 2.3 → positive selection on most epitope changes, evidence for clonal interference

Non-epitope amino acid changes have a decreased relative fixation probability:

 $g / g_0 = 0.3$   $\rightarrow$  negative selection

[N. Strelkowa, M.L., Genetics 2012]

- 1. Predictability in (evolutionary) biology
- 2. Influenza's mode of evolution
- 3. Year-to-year prediction of influenza A (H3N2)

#### Does the strain content up to a given time predict future prevalent strains?



What can be predicted?

- how the tree will grow in the following season.
- distribution of frequencies in the future population.

**Strain frequencies**  $x_i$  correspond to fractions of the infected host population corresponding to a given strain.

#### Clades

- are sets of strains with a common ancestor
- are destined for fixation or loss
- have frequency time-series

$$X_{\alpha}(t) = \sum_{i: \ \alpha, t} x_i$$

- are suitable units of prediction.



depends on non-epitope and epitope sequence

$$f_i = f_0 - \mathcal{L}(a_i) - \sum_{j: t_j < t_i} x_j C(a_i, a_j)$$

non-epitope mutations: increase mutational load epitope mutations: decrease cross-immunity load



depends on non-epitope and epitope sequence

$$f_i = f_0 - \mathcal{L}(a_i) - \sum_{j:t_j < t_i} x_j C(a_i, a_j)$$

non-epitope mutations: increase mutational load epitope mutations: decrease cross-immunity load



Sequence distance-based approximation:

$$\mathcal{L}(a_i) = \sigma_{\rm ne} d_{\rm ne}(a_i, a_*)$$

depends on non-epitope and epitope sequence

$$f_i = f_0 - \mathcal{L}(a_i) - \sum_{j: t_j < t_i} x_j C(a_i, a_j)$$

non-epitope mutations: increase mutational load epitope mutations: decrease cross-immunity load



Sequence distance-based approximation:

$$C(a_i, a_j) = -\sigma_{\rm ep} \exp[-d_{\rm ep}(a_i, a_j)/D_0]$$

depends on non-epitope and epitope sequence

$$f_i = f_0 - \mathcal{L}(a_i) - \sum_{j:t_j < t_i} x_j C(a_i, a_j)$$

non-epitope mutations: increase mutational load epitope mutations: decrease cross-immunity load



**Epistatic and inhomogeneous effects for clades** are inferred from accumulation of synonymous mutations.

# **Prediction scheme**



- 1. Model training from past 4 8 years
- 2. Clade frequency prediction from one year to the next

$$X_{\alpha}(t) = \sum_{i:\alpha,t} x_i \longrightarrow \hat{X}_{\alpha}(t+1) = \sum_{i:\alpha,t} x_i \exp[f_i],$$

**3.** Validation from posterior clade frequencies in the following year.

#### Clade frequency evolution between consecutive years

Clade growth/decline is measured by the frequency ratio W = X(t+1) / X(t).



**Epitope** and **non-epitope** mutations carry predictive information.



Fitness predictions confirm the particular role of East and Southeast Asia in seeding antigenic variants [Russell et al. Science 2008].

Cross-immunity overlap of a vaccine strain with circulating strains



# Conclusion

- A joint fitness model of antigenic and deleterious effects successfully predicts the frequency evolution of HA sequence clades from one year to the next.
- Our results call for a more comprehensive epidemiology of influenza.
- The prediction scheme can integrate diverse types of data.



- The model provides a principled method for vaccine strain selection.
- The predictive power at the phenotypic level remains to be tested.





COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. XIII, 001-14 (1960)

#### The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

> EUGENE P. WIGNER Princeton University

A Reasonable Effectiveness of Mathematics in Evolutionary Biology?