Publication

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

Equations

Predect next season's frequencies X for clade v based on prior season

$$X_v(t) = \sum_{i:v,t} x_i$$
Observed frequency X of clade v at a given season t for strain i

$$f_i$$
Malthusian fitness or growth rate, measured in units of 1yr

$$\hat{X}_v(t+1) = \sum_{i:v,t} x_i exp(f_i)$$
Predicted frequency of that clade in season 1 year or later

Strain fitness where mutations in epitope regions confer postive and non-epitope regions are negative

$$\mathcal{C}(a_i, a_j)$$
Cross-immunity amplitude: antigenic similarity of strains i and j

$$\mathcal{L}(a_i)$$
Fitness cost: cumulative effect of recent non-epitope amino acid changes

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

$$\hat{W}_v = \frac{\hat{X}_v(t+1)}{X_v(t)}$$
Wrightian fitness: i.e. if next season is larger than prior season then $W_v > 1$

Vaccine selection

$$\mathcal{C}_v(t) = \sum_{i:t} x_i \mathcal{C}(a_i, a_v)$$
Average cross-immunity between vaccine strain and the circulating strain in a given season

$$\hat{\mathcal{C}}_v(t+1) = \sum_{i:t} x_i \mathcal{C}(a_i, a_v) exp(f_i)$$
Optimal vaccine maximizes this reduction

Methods summary

$$t = \text{Season starting in October year } y \text{ -1 to April year } y$$

$$a_v = \text{Given HA sequence that is the common clade ancester}$$

$$\mathcal{D}_{ep} = \text{Linear amino acid distance between pairs of sequences in epitope codons}$$

$$c(\mathcal{D}_{ep}) = exp\left(\frac{-\mathcal{D}_{ep}}{\mathcal{D}_0}\right) = \text{nonlinear cross-immunity amplitude}$$

$$f_i^{ep} = \sigma_{ep} \sum_{j:t_j < t_i} x_j c(\mathcal{D}_{ep}(a_i, a_j))$$

$$f_i^{ne} = -\sigma_{ep} \mathcal{D}_{ne}(a_i, a_i^*)$$

$$a_i^* = \text{last ancestor of strain } i \text{ in a previous season}$$

Assumptions

- Clades are geographically well-mixed

- Clade frequencies are sufficiently large to measure $X_v(t) > 0.15$

- Each strain has a Malthusian fitness or growth rate f_i

- Epitope changes are predominantly under positive selection

- Non-epitope mutations are predominantly under negative selection

- SIR

- Ignores higher-order antigenic interactions of more than two strains

- Ignores birth-death turnover of the host population

Publication

Huddleston, J., Barnes, J.R., Rowe, T., Xu, X., Kondor, R., Wentworth, D.E., Whittaker, L., Ermetal, B., Daniels, R.S., McCauley, J.W. and Fujisaki, S., 2020. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution. Elife, 9, p.e60067.

$$\frac{1}{Z(t)} = \text{Normalized to 1}$$

$$\hat{x}_i(t + \delta t) = \frac{1}{Z(t)} x_i(t) exp(f_i(t)\delta t) \qquad \text{Estimated future frequency of each strain } i$$

$$f_i(t) = \beta_{ne} f_{i,ml}(t) + \beta_{lbi} f_{i,lbi}(t) \qquad \text{Additive model of various fitness terms}$$

$$u = t + 1 \qquad \text{time unit comparison}$$

$$h(s_i, s_j) = \text{hamming distance between strain } i \text{ and strain } j$$

$$d_i(u) = \sum_{j \in s(u)} x_j(u) h(s_i, s_j) \qquad \text{strain-specific distance to the future}$$

$$d_i(\hat{u}) = \sum_{j \in s(\hat{u})} x_j(\hat{u}) h(s_i, s_j) \qquad \text{Estimated ... strain-specific distance to the future}$$

$$f_{i,ep}(t) = \sum_{j:t_j < t_i} -max(x_j) exp\left(\frac{-\mathcal{D}_{ep}(a_i, a_j)}{\mathcal{D}_0}\right) \qquad \text{Antigenic drift}$$

$$f_{i,DMS}(t) = \sum_{r \in r, a_i != r, a_j} log_2 \frac{\pi_{r,a_i}}{\pi_{r,a_j}} \qquad \text{Functional constraint}$$

Publication

Neher, Richard A., Colin A. Russell, and Boris I. Shraiman. "Predicting evolution from the shape of genealogical trees." Elife 3 (2014): e03568.

$$m_{\uparrow i} = \tau \left(1 - e^{\frac{-b_i}{\tau}}\right) + e^{\frac{-b_i}{\tau}} \sum_j m_{\uparrow i_j} \qquad \text{up-messages of the parent of node } i$$

$$m_{\downarrow i} = \tau \left(1 - e^{\frac{-b_i}{\tau}}\right) + e^{\frac{-b_i}{\tau}} \left[m_{\downarrow i} + \sum_{k \neq j} m_{\uparrow i_k}\right] \qquad \text{down-messages of the children of node } i$$

$$\lambda_i(\tau) = m_{\downarrow i} + \sum_j m_{\downarrow i_j} \qquad \text{Exponentially discounted tree length}$$

$$n = \text{number of offspring}$$
$$t = \text{time}$$
$$x = \text{ancestor fitness}$$
$$P(n|x,t) = \text{Probability of the number of offspring } n \text{ after time t and given x}$$

$$P(n|x,tv,t+\delta t) = [1 - \delta t(2 + x + u)]P(n|x,t) + \delta t < uP(n|x+s,t) > +\delta t(1+x) \sum_{n'=0}^{n} P(n-n'|x,t)P(n'|x,t) \quad \text{backward master equation}$$

$$1 + x = \text{birth rate}$$
$$1 = \text{death rate}$$