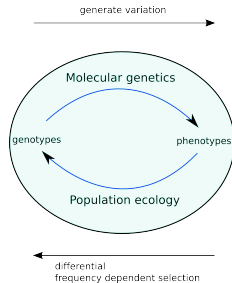
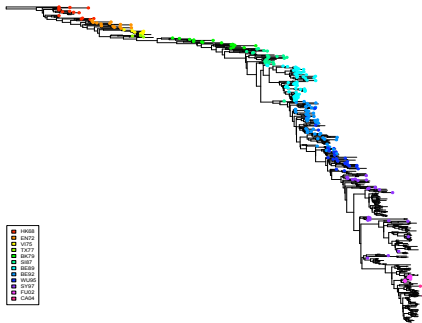
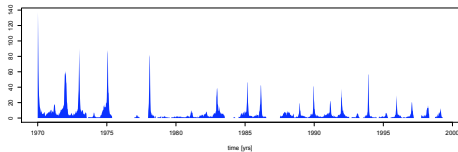


STATISTICAL INFERENCE FOR VIRAL DISEASES
USING
EPIDEMIOLOGICAL AND GENETIC
SUMMARY STATISTICS

Oliver Ratmann (Duke Biology)

Christophe Fraser (DIDE Imperial), Ge Donker (NIVEL), Katia Koelle (Duke Biology)

Epidemiological & evolutionary dynamics of influenza A (H3N2): overlap \rightarrow interact \rightarrow reflect



Statistical inference using epidemiological and genetic data

Bayesian inference

- x_0 observed incidence time series AND viral phylogeny
- phylodynamic model that defines likelihood $f(x_0|\theta)$ implicitly

⇒ Bayes' posterior density

$$f(\theta|x_0) = f(x_0|\theta)\pi(\theta) / f(x_0)$$

Approximate Bayesian Computation

circumvent evaluation of $f(x_0|\theta)$ in two steps:

- simulate from likelihood, $x \sim f(\cdot|\theta)$
- weight simulation under θ by degree ϵ with which x and x_0 match

Statistical inference using epidemiological and genetic data

Bayesian inference

- x_0 observed incidence time series AND viral phylogeny
- phylodynamic model that defines likelihood $f(x_0|\theta)$ implicitly

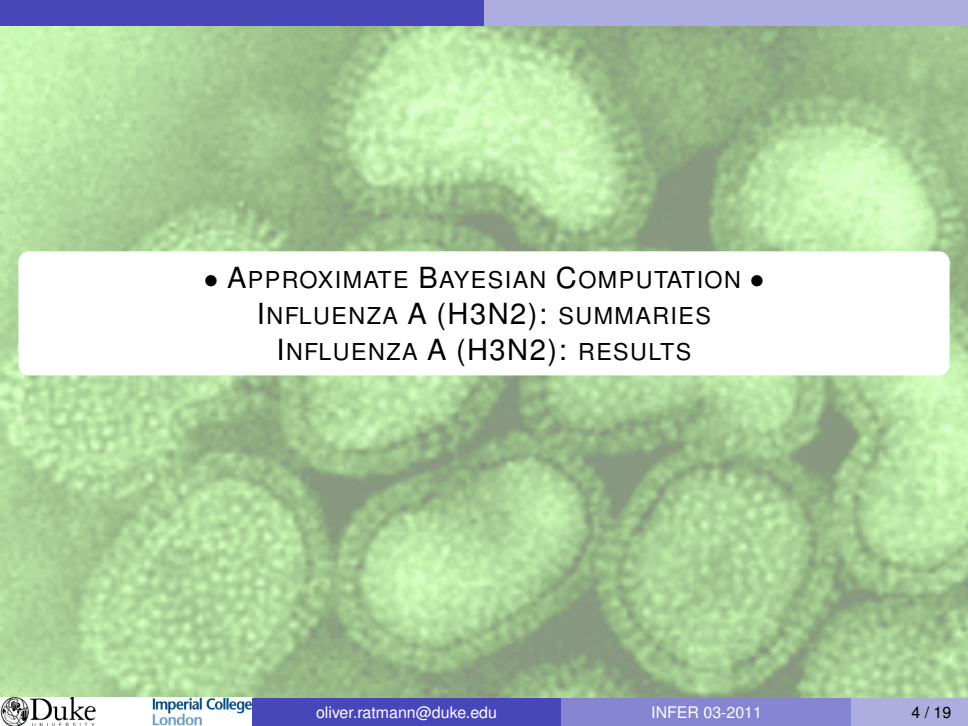
⇒ Bayes' posterior density

$$f(\theta|x_0) = f(x_0|\theta)\pi(\theta) / f(x_0)$$

Approximate Bayesian Computation

circumvent evaluation of $f(x_0|\theta)$ in two steps:

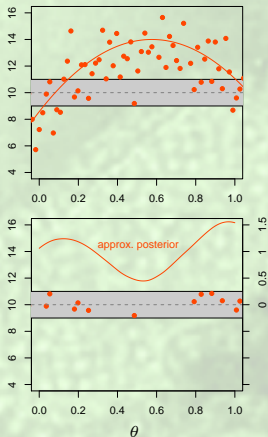
- simulate from likelihood, $x \sim f(\cdot|\theta)$
- weight simulation under θ by degree ε with which x and x_0 match

A microscopic image of influenza A (H3N2) virus particles, showing their characteristic spherical shape and surface structure. The background is a light green color.

- APPROXIMATE BAYESIAN COMPUTATION ●
- INFLUENZA A (H3N2): SUMMARIES
- INFLUENZA A (H3N2): RESULTS

Approximate Bayesian Computation

- eg S_1 : # antigenic clusters



- set ABC kernel $\kappa_\tau(\varepsilon)$ eg to $1/\tau \mathbb{1}\{|\varepsilon| \leq \tau/2\}$

Rejection-sampler

- 1 Sample $\theta \sim \pi(\theta|M)$
- 2 Simulate $x \sim f(x|\theta)$, compute summaries $\mathbb{S}(x) = \{S_1(x), \dots, S_K(x)\}$
- 3 Compute auxiliary errors $\varepsilon_k = \rho_k(S_k(x), S_k(x_0))$
- 4 Accept (θ, ε) with prob proportional to

$$\prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k)$$

Approximate Bayesian Computation

ABC: a particular auxiliary variable Monte Carlo method

- ABC projection $\xi_{x_0} : x \rightarrow (\varepsilon_1, \dots, \varepsilon_K)$, $\varepsilon_k = \rho_k(S_k(x), S_k(x_0))$
- for given θ , errors are distributed according to

$$\xi_{x_0, \theta}(E_1 \times \dots \times E_K) \\ = f\left(\xi_{x_0}^{-1}(E_1 \times \dots \times E_K) \mid \theta, M\right) = \int \mathbb{1}\{x \in \xi_{x_0}^{-1}(E_1 \times \dots \times E_K)\} f(dx \mid \theta, M)$$

- augmented sampling density of ABC is

$$f_{\text{ABC}}(\theta, \varepsilon \mid x_0) \propto \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k) \times \xi_{x_0, \theta}(\varepsilon_1, \dots, \varepsilon_K) \pi(\theta)$$

ABC kernel \times prior predictive error density given θ

- .. **augmented likelihood** still cannot be computed pointwise for $z = (\theta, \varepsilon)$
- .. **and** interested in auxiliary variable for model criticism (Ratmann PNAS 2009)

Approximate Bayesian Computation

ABC: a particular auxiliary variable Monte Carlo method

- ABC projection $\xi_{x_0} : x \rightarrow (\varepsilon_1, \dots, \varepsilon_K)$, $\varepsilon_k = \rho_k(S_k(x), S_k(x_0))$
- for given θ , errors are distributed according to

$$\xi_{x_0, \theta}(E_1 \times \dots \times E_K)$$

$$= f\left(\xi_{x_0}^{-1}(E_1 \times \dots \times E_K) \mid \theta, M\right) = \int \mathbb{1}\{x \in \xi_{x_0}^{-1}(E_1 \times \dots \times E_K)\} f(dx \mid \theta, M)$$

- augmented sampling density of ABC is

$$f_{\text{ABC}}(\theta, \varepsilon \mid x_0) \propto \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k) \times \xi_{x_0, \theta}(\varepsilon_1, \dots, \varepsilon_K) \pi(\theta)$$

ABC kernel \times prior predictive error density given θ

- .. augmented likelihood still cannot be computed pointwise for $z = (\theta, \varepsilon)$
- .. and interested in auxiliary variable for model criticism (Ratmann PNAS 2009)

Approximate Bayesian Computation

ABC: a particular auxiliary variable Monte Carlo method

- ABC projection $\xi_{x_0} : x \rightarrow (\varepsilon_1, \dots, \varepsilon_K)$, $\varepsilon_k = \rho_k(S_k(x), S_k(x_0))$
- for given θ , errors are distributed according to

$$\xi_{x_0, \theta}(E_1 \times \dots \times E_K)$$

$$= f\left(\xi_{x_0}^{-1}(E_1 \times \dots \times E_K) \mid \theta, M\right) = \int \mathbb{1}\{x \in \xi_{x_0}^{-1}(E_1 \times \dots \times E_K)\} f(dx \mid \theta, M)$$

- augmented sampling density of ABC is

$$f_{\text{ABC}}(\theta, \varepsilon \mid x_0) \propto \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k) \times \xi_{x_0, \theta}(\varepsilon_1, \dots, \varepsilon_K) \pi(\theta)$$

ABC kernel \times prior predictive error density given θ

- .. **augmented likelihood** still cannot be computed pointwise for $z = (\theta, \varepsilon)$
- .. **and** interested in auxiliary variable for model criticism (Ratmann PNAS 2009)

Approximate Bayesian Computation

ABC: no need to calculate the augmented likelihood

- if we propose from the intractable component

MCMC-sampler

- 1 Propose $\theta' \sim q(\theta \rightarrow \cdot)$ and propose $\varepsilon' \sim \xi_{x_0, \theta'}$
- 2 Accept $z' = (\theta', \varepsilon')$ with probability

$$\min \{ 1, \text{mh}(z \rightarrow z') \}, \quad \text{mh}(z \rightarrow z') = \frac{q(\theta' \rightarrow \theta)}{q(\theta \rightarrow \theta')} \times \frac{\pi(\theta') \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon'_k)}{\pi(\theta) \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k)}$$

and otherwise stay at z .

- Because, for $q(z \rightarrow z') = q(\theta \rightarrow \theta')\xi_{x_0, \theta'}(\varepsilon')$,

$$\{q(z \rightarrow z')\text{mh}(z \rightarrow z')\} / \{q(z' \rightarrow z)\text{mh}(z' \rightarrow z)\} = f_{\text{ABC}}(z'|x_0) / f_{\text{ABC}}(z|x_0)$$

Approximate Bayesian Computation

ABC: no need to calculate the augmented likelihood

- if we propose from the intractable component

MCMC-sampler

- 1 Propose $\theta' \sim q(\theta \rightarrow \cdot)$ and propose $\varepsilon' \sim \xi_{x_0, \theta'}$
- 2 Accept $z' = (\theta', \varepsilon')$ with probability

$$\min \{ 1, \text{mh}(z \rightarrow z') \}, \quad \text{mh}(z \rightarrow z') = \frac{q(\theta' \rightarrow \theta)}{q(\theta \rightarrow \theta')} \times \frac{\pi(\theta') \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon'_k)}{\pi(\theta) \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k)}$$

and otherwise stay at z .

- Because, for $q(z \rightarrow z') = q(\theta \rightarrow \theta') \xi_{x_0, \theta'}(\varepsilon')$,

$$\{q(z \rightarrow z') \text{mh}(z \rightarrow z')\} / \{q(z' \rightarrow z) \text{mh}(z' \rightarrow z)\} = f_{\text{ABC}}(z'|x_0) / f_{\text{ABC}}(z|x_0)$$

Approximate Bayesian Computation

ABC: no need to calculate the augmented likelihood

- if we propose from the intractable component

MCMC-sampler

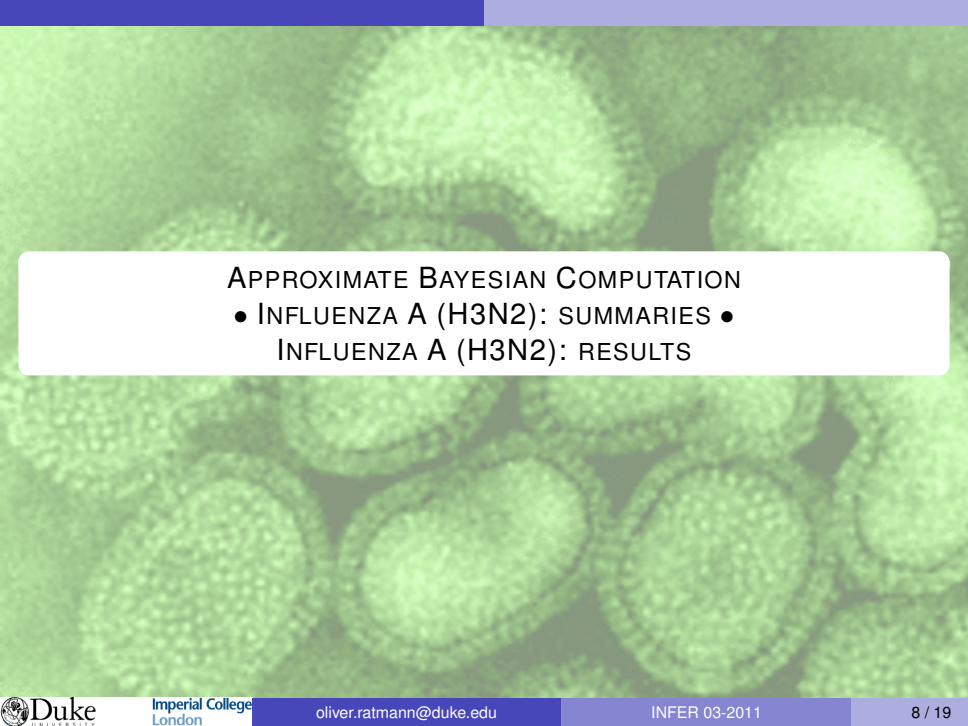
- 1 Propose $\theta' \sim q(\theta \rightarrow \cdot)$ and propose $\varepsilon' \sim \xi_{x_0, \theta'}$
- 2 Accept $z' = (\theta', \varepsilon')$ with probability

$$\min \{ 1, \text{mh}(z \rightarrow z') \}, \quad \text{mh}(z \rightarrow z') = \frac{q(\theta' \rightarrow \theta)}{q(\theta \rightarrow \theta')} \times \frac{\pi(\theta') \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon'_k)}{\pi(\theta) \prod_{k=1}^K \kappa_{\tau_k}(\varepsilon_k)}$$

and otherwise stay at z .

- Because, for $q(z \rightarrow z') = q(\theta \rightarrow \theta')\xi_{x_0, \theta'}(\varepsilon')$,

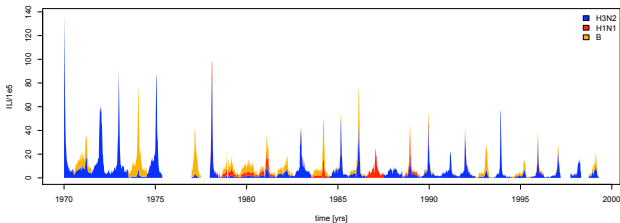
$$\{q(z \rightarrow z')\text{mh}(z \rightarrow z')\} / \{q(z' \rightarrow z)\text{mh}(z' \rightarrow z)\} = f_{\text{ABC}}(z'|x_0) / f_{\text{ABC}}(z|x_0)$$

A microscopic image of influenza A (H3N2) virus particles, showing their characteristic spherical shape and surface structure. The background is a light green color.

APPROXIMATE BAYESIAN COMPUTATION

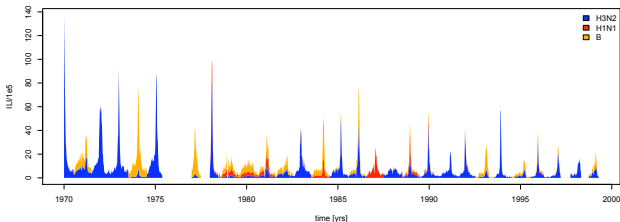
- INFLUENZA A (H3N2): SUMMARIES •
- INFLUENZA A (H3N2): RESULTS

Summaries characterizing seasonal influenza A (H3N2) incidence



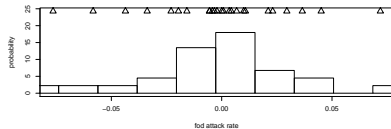
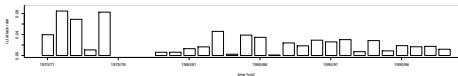
- interannual variability
- periodicity
- explosiveness
- overall magnitude

Summaries characterizing seasonal influenza A (H3N2) incidence

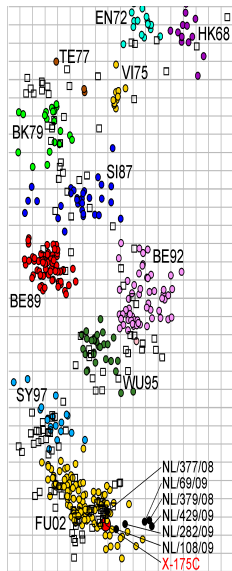


- interannual variability
- periodicity
- explosiveness
- overall magnitude

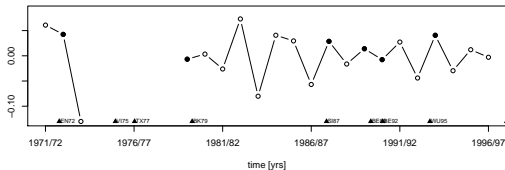
- interannual variability, eg: differences in annual attack rate (Δ_y)



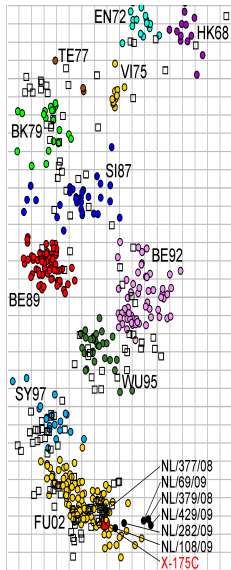
Summaries characterizing influenza A (H3N2) antigenic evolution



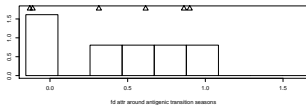
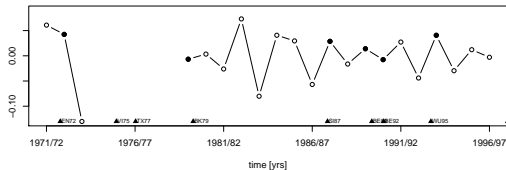
- Number of antigenic clusters (Smith et al 2004)
- No large changes in annual attack rate at transition yrs



Summaries characterizing influenza A (H3N2) antigenic evolution

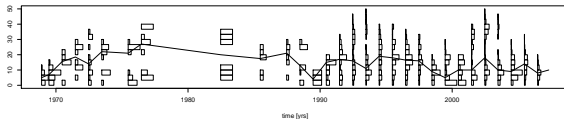
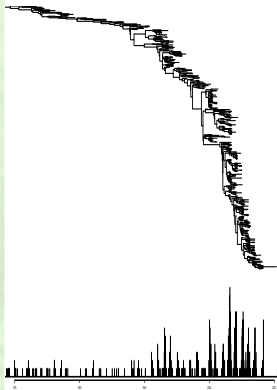


- Number of antigenic clusters (Smith et al 2004)
- No large changes in annual attack rate at transition yrs



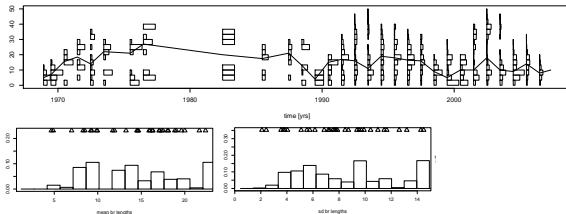
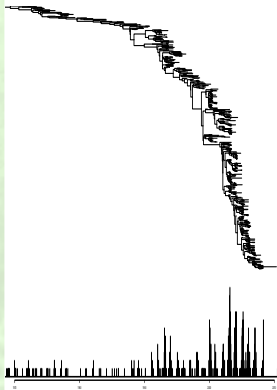
Summaries characterizing the influenza A (H3N2) HA phylogeny

- pw diversity between strains collected in season

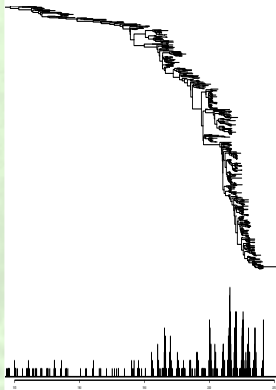


Summaries characterizing the influenza A (H3N2) HA phylogeny

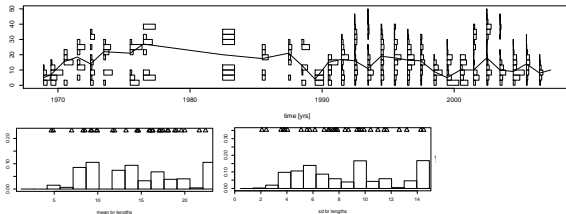
- pw diversity between strains collected in season



Summaries characterizing the influenza A (H3N2) HA phylogeny

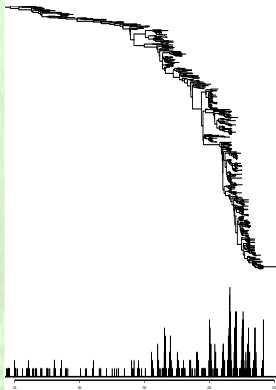


- pw diversity between strains collected in season

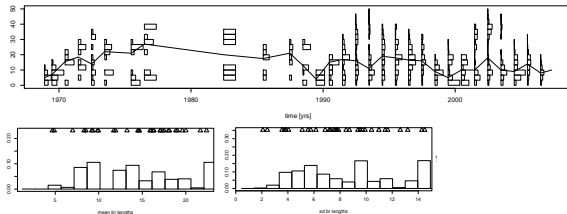


- divergence of serially sampled strains to root
- TMRCA of strains collected in season

Summaries characterizing the influenza A (H3N2) HA phylogeny

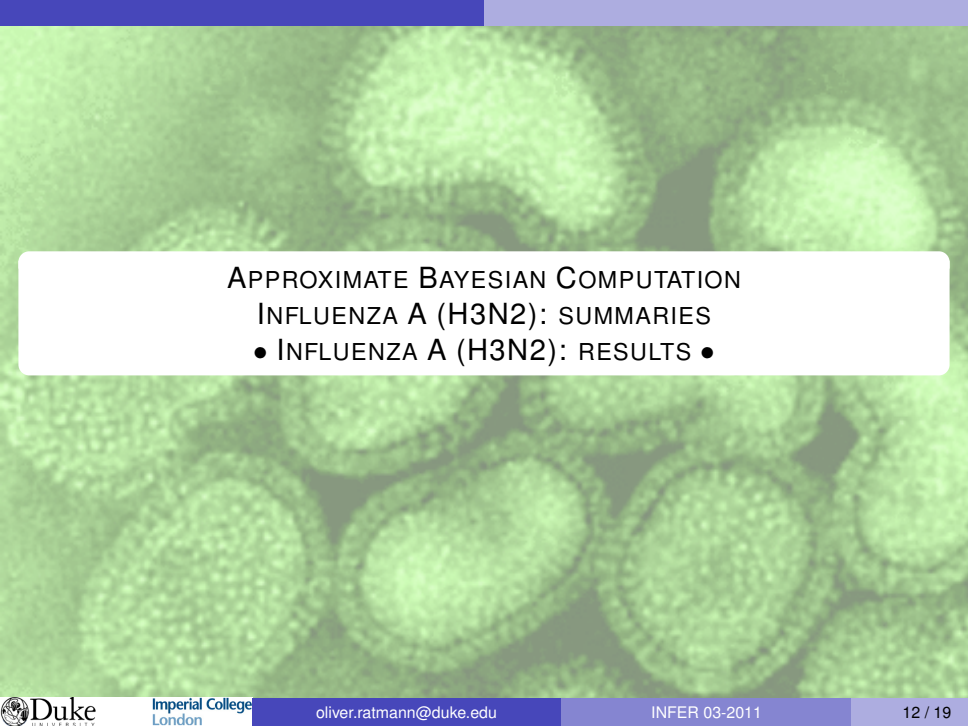


- pw diversity between strains collected in season



- divergence of serially sampled strains to root
- TMRCA of strains collected in season

- substantial # pilot runs to determine which summary to include based on ability to further constrain posterior Θ (Nunes Balding 2010)

A microscopic image of influenza A (H3N2) virus particles, showing their characteristic spherical shape and surface structure. The background is a light green color.

APPROXIMATE BAYESIAN COMPUTATION
INFLUENZA A (H3N2): SUMMARIES

- INFLUENZA A (H3N2): RESULTS •

SIRS with sinusoidal seasonal forcing

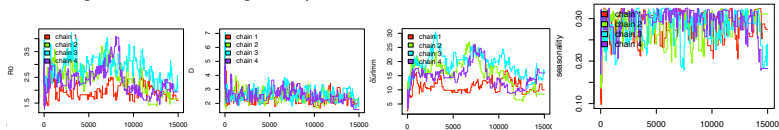
- MCMC using epidemiological summaries
 - fix: demography, birth/death rate, low migration

SIRS with sinusoidal seasonal forcing

- MCMC using epidemiological summaries

fix: demography, birth/death rate, low migration

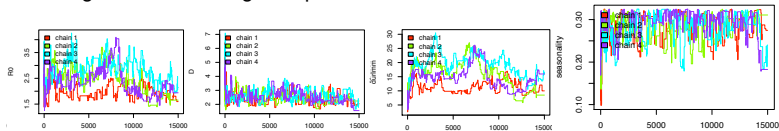
⇒ strong seasonal forcing to explain interannual seasonal variation



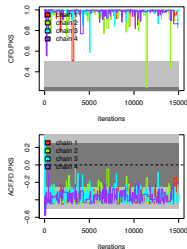
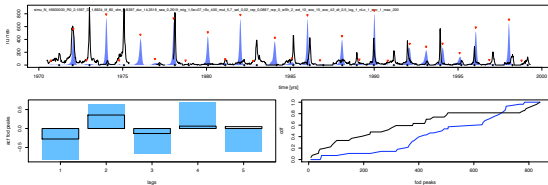
SIRS with sinusoidal seasonal forcing

- MCMC using epidemiological summaries
fix: demography, birth/death rate, low migration

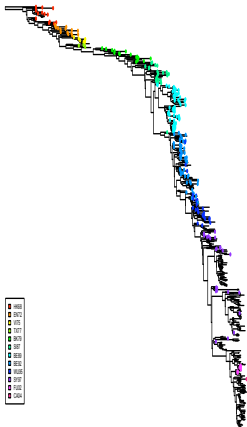
⇒ strong seasonal forcing to explain interannual seasonal variation



⇒ too regular and too strong sustained oscillations



Antigenic tempo model (Koelle et al JRSoc 2010)



- track status of infection with multiple **phenot distinct variants**

$$i = 1, \dots, n: \quad \frac{dS_i}{dt} = \mu(N - S_i) - \beta_t \frac{S_i}{N} \sum_{j=1}^n \sigma_{ij} I_j + \gamma(N - S_i - I_j)$$

$$\frac{dI_i}{dt} = \beta_t \frac{S_i}{N} I_i - (\mu + \nu) I_i$$

- specify only tempo with which variants emerge

$$\frac{dI_i}{dt} = \beta_t \frac{S_i}{N} I_i - (\mu + \nu) I_i + h(\text{age}_i) I_i$$

$$h(a) = \kappa / \lambda (a / \lambda)^{\kappa - 1}$$

- simulate strains of each variant

Antigenic tempo model

- MCMC using epidemiological and immunological summaries

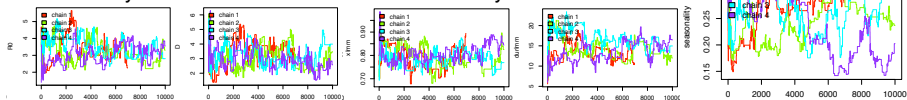
fix: demography, birth/death rate, linear aging, low migration; not shown: λ , report prob

Antigenic tempo model

- MCMC using epidemiological and immunological summaries

fix: demography, birth/death rate, linear aging, low migration; not shown: λ , report prob

⇒ intrinsic dynamics contribute to overall seasonality

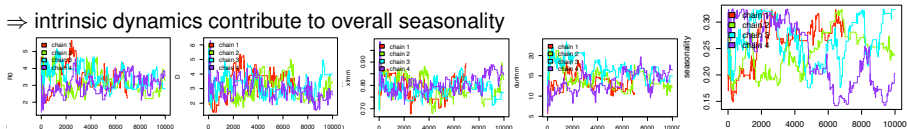


Antigenic tempo model

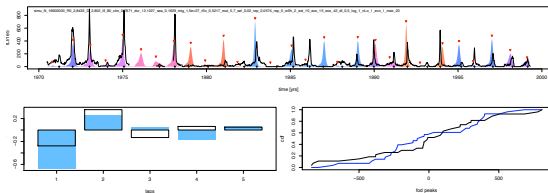
- MCMC using epidemiological and immunological summaries

fix: demography, birth/death rate, linear aging, low migration; not shown: λ , report prob

⇒ intrinsic dynamics contribute to overall seasonality



⇒ consistent with observed summaries



Antigenic tempo model with genetic simulations

- MCMC using also genetic summaries

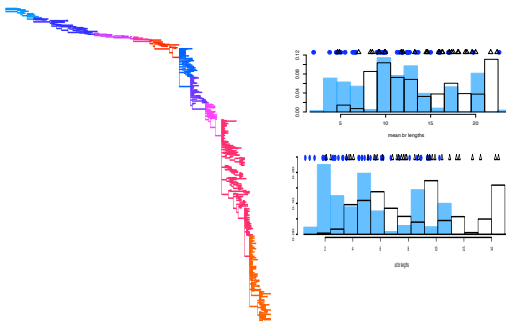
fix: HA nucl mut rate 5.7×10^{-3} /site/yr, low seasonality (mild bottleneck)

Antigenic tempo model with genetic simulations

- MCMC using also genetic summaries

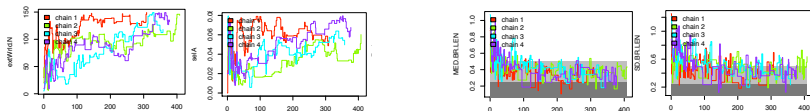
fix: HA nucl mut rate 5.7×10^{-3} /site/yr, low seasonality (mild bottleneck)

⇒ in principle, model of punctuated antigenic change can reproduce limited diversity



Antigenic tempo model with genetic simulations

- MCMC using also genetic summaries of strains sampled from northern EU (1968-2009)
fix: HA nucl mut rate 5.7×10^{-3} /site/yr, low seasonality (mild bottleneck)
- ⇒ **However**, given summer trough, Dutch population ($N \approx 15m$) too small to create diversity scale system by constant e to see how big the population should be



To match avg expected diversity and variation in diversity across seasons within 1.5-fold,
 $e > 100$ or $eN > 1500m$

Take home

Methodological:

- ABC enables the analysis of influenza dynamics with epidemiological, genetic and immunogenic data

Epidemiological:

- SIRS fails to reproduce influenza A (H3N2)'s irregular seasonality
- modeling abrupt changes in herd immunity within H3N2:
 - excite dynamics that match H3N2's irregular seasonality
 - in principle limit genetic diversity to observed levels
- pop size required suggests spatial model component necessary

Thank you!

and the Wellcome Trust for funding through a Sir Henry Wellcome fellowship