Population dynamics in phylogeographic analyses

Denise Kühnert
Computational Evolution Group
University of Auckland, New Zealand

November 5, 2010
Motivation

Analyse data from different locations.
Reconstruct history of sampled individuals based on DNA sequences. Use knowledge of sampling date and location. Lemey et al 2009: Model the process of migration by CTMC where the state space contains all locations from where sequences have been sampled. (Bayesian inference.)
Consider population dynamics

Overview

Motivation

Phylogenetics vs / and Epidemiology

A small simulation study
Epidemiological viewpoint

- Different type of data: Incidence / case data
- Answering different questions
- Anticipating future developments

Under which circumstances does the introduction of a pathogen into a population cause an outbreak of the disease? When to expect an epidemic? How efficient are measures to contain disease outbreak? Vaccination? Isolation? ...
SIR model

Transmission of the disease to susceptibles leads to a period of illness until recovery, which implies immunity.

\( S, I \) and \( R \): Fractions of susceptible, infected and recovered individuals in host population

\[
\begin{align*}
\frac{dS}{dt} &= \mu - \beta SI - \mu S \\
\frac{dI}{dt} &= \beta SI - \gamma I - \mu I \\
\frac{dR}{dt} &= \gamma I - \mu R
\end{align*}
\]
Overview

Motivation

Phylogenetics vs / and Epidemiology

A small simulation study

Phylogenetics vs. Epidemiology

Genetic data
(e.g. DNA sequences)

Reconstruct phylogenetic history

Infer past infection dynamics

Incidence data
(e.g. case counts)

Fit data to epidemiological model

Anticipate future dynamics
Recent combinations of phylogenetic and epidemiological approaches:

- Genetic data used as "back up" for epidemiological studies
- Phylogenetic analysis of the basic reproduction ratio $R_0$
- Individual-based dynamical models enable sample-based approaches

Examples..
Overview
Motivation
Phylogenetics vs / and Epidemiology
A small simulation study

Genetic Diversity in the SIR Model of Pathogen Evolution

Isabel Gordo¹*, M. Gabriela M. Gomes²,³, Daniel G. Reis¹, Paulo R. A. Campos³

1 Instituto Gulbenkian de Ciência, Oeiras, Portugal, 2 Centro de Matemática e Aplicações Fundamentais, Universidade de Lisboa, Lisboa, Portugal, 3 Departamento de Física, Universidade Federal Rural de Pernambuco, Recife-PE, Brazil

Abstract
We introduce a model for assessing the levels and patterns of genetic diversity in pathogen populations, whose epidemiology follows a susceptible-infected-recovered model (SIR). We model the population of pathogens as a metapopulation composed of subpopulations (infected hosts). Where pathogens replicate and mutate. Hosts transmit pathogens to uninfected hosts. We show that the level of pathogen variation is well predicted by analytical expressions, such that pathogen neutral molecular variation is bounded by the level of infection and increases with the duration of infection. We then introduce selection in the model and study the invasion probability of a new pathogenic strain whose fitness (R₀(1+s)) is higher than the fitness of the resident strain (R₀). We show that this invasion probability is given by the relative increment in R₀ of the new pathogen (s). By analyzing the patterns of genetic diversity in this framework, we identify the molecular signatures during the replacement and compare these with those observed in sequences of influenza A.


Editors: Robert DeSalle, American Museum of Natural History, United States of America
Received: October 1, 2008; Accepted: January 21, 2009; Published: March 16, 2009

Copyright: © 2009 Gordo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: We thank the Portuguese Research Council (FCT) for financial support through PTDC/BAE-BOE/66276/2006. We thank FCT of Portugal grants PTDC/BAE-BOE/66276/2006, and Fundação Calouste Gulbenkian for the financial support of hermes, a High Performance Computing Centre, hosted in Instituto Gulbenkian de Ciência, Oeiras. PRAC is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Ciência e Tecnologia do Estado do Pernambuco (FACEPE) and program PRONEX-MCT-CNpq-FACEPE. MGS and DGR are supported by FCT and the European Commission grant MEXT-CT-2004-14338. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: igordo@igc.gulbenkian.pt
Overview
Motivation
Phylogenetics vs / and
Epidemiology
A small simulation study

**Phylogenetic analysis of the basic reproduction ratio** $R_0$

---

**Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings**

Christophe Fraser,1,* Christi A. Donnelly,1,* Simon Cauchemez,1 William P. Hanage,1
María D. Van Kerkhove,2† T. Déandre Hellingworth,3 Jamie Griffin,2 Rebecca F. Baggsley,2
Helen E. Jenkins,3 Emily J. Lyons,3 Thibault Jombart,5 Wes R. Hinsley,1 Nicholas C. Grassly,1
Francois Balloux,1 Aara C. Ghami,3 Neil M. Ferguson1‡;
Andrew Rambaut,1 Oliver G. Pybus3;
Hugo Lopez-Gatell,6 Celia M. Alpuente-Aranda,6 Letza Bojorquez Chapela,6 Ethel Palacios Zavala6;
Dulce Ma. Espejo Guevara6;
Francesco Checchi7, Erika Garcia7, Stephane Hugonnet7, Cathy Roth7

The WHO Rapid Pandemic Assessment Collaboration‡

A novel influenza A (H1N1) virus has spread rapidly across the globe. Judging its pandemic potential is difficult with limited data, but nevertheless essential to inform appropriate health responses. By analyzing the outbreak in Mexico, early data on international spread, and viral genetic diversity, we make an early assessment of transmissibility and severity. Our estimates suggest that 23,000 (range 6000 to 32,000) individuals had been infected in Mexico by late April, giving an estimated case fatality ratio (CFR) of 0.4% (range: 0.3 to 1.8%) based on confirmed and suspected deaths reported to that time. In a community outbreak in the small community of La Gloria, Veracruz, no deaths were attributed to infection, giving an upper 95% bound on CFR of 0.6%. Thus, although substantial uncertainty remains, clinical severity appears less than that seen in the 1918 influenza pandemic but comparable with that seen in the 1957 pandemic. Clinical attack rates in children in La Gloria were twice that in adults (<15 years of age: 61%, ≥15 years: 29%). Three different epidemiological analyses gave basic reproduction number ($R_0$) estimates in the range of 1.4 to 1.6, whereas a genetic analysis gave a central estimate of 1.2. This range of values is consistent with 14 to 23 generations of human-to-human transmission having occurred in Mexico to late April. Transmissibility is therefore substantially higher than that of seasonal flu, and comparable with lower estimates of $R_0$ obtained from previous influenza pandemics.

On 29 April 2009, the World Health Organization (WHO) announced that the rapid travels from that country. There are uncertainties about all aspects of this outbreak, including the those over 60 years of age (3), and this could result in an underestimate of overall morbidity. Right censoring of mortality data, which occurs when additional deaths subsequently arise among cases already included in surveillance data, can also bias estimates of the true case fatality ratio (4). Finally, suspected deaths may not all have been caused by infection with the novel virus. These uncertainties necessarily affect any estimate of the case fatality ratio (CFR).

On the basis of international travel patterns, we would expect a proportion of cases of any infection spreading widely in Mexico to be exported by travelers (5). Owing to intense surveillance for influenza-like illness in those returning from Mexico, ascertainment of early cases in newly affected countries was almost certainly more complete and rapid than local surveillance of mild cases in Mexico. Airline passenger flow out of Mexico shows a significant correlation with the frequency of detected confirmed cases worldwide (Spearman correlation coefficient: 0.56, $p = 0.004$) (Fig. 1A and B). We thus use data on cases among travelers and backcalculation methods to estimate the total number of people infected in Mexico. Key underlying assumptions in this analysis are that population mixing in Mexico is equally likely between Mexican residents and tourists, and tourists and Mexican residents are at equal risk of infection (despite demographic and other differences). If infections are concentrated away from traveler destinations (Fig. 1C) presents the spatial distribution, by state, of cases within Mexico by 5 May, the number of people infected in Mexico will be underestimated, and conversely will be overestimated if the epidemic has disproportionately affected geographical zones visited by travelers. Under the assumption that reporting of infections in travelers was complete, we estimated
Simulation of Influenza epidemic under a stochastic SIR model with migration

Simulate the disease dynamics when a new strain comes into a completely susceptible population.

- Simple structured population: 3 locations with symmetric migration
- At every (random) time step one of the following events happens:
  1. New Infection at rate $\beta$
  2. Migration at rates $m_{ik}$ for $i \neq k \in \{0, 1, 2\}$
  3. Recovery of an infected individual at rate $\gamma$
  3. Birth of a susceptible individual at rate $\mu$
  3. Death of an individual at rate $\mu$
Deterministic representation

\[ \frac{dS_k}{dt} = \mu_k - \beta_k S_k I_k - \mu_k S_k + \sum_{l \neq k} (m_{lk} S_l - m_{kl} S_k) \]

\[ \frac{dI_k}{dt} = \beta_k S_k I_k - \gamma_k I_k - \mu_k I_k + \sum_{l \neq k} (m_{lk} I_l - m_{kl} I_k) \]

\[ \frac{dR_k}{dt} = \gamma_k I_k - \mu_k R_k + \sum_{l \neq k} (m_{lk} R_l - m_{kl} R_k) \]
Outlook

- Simulate different scenarios of disease outbreak and test how well they can be reconstructed.
- How important are population dynamics for ancestral reconstruction?
- Does inclusion of genetic data in epidemiological analyses change recommended actions on disease prevention? Does it tell us more about long-term effects?
Thank you.