

Evolutionary epidemiology in the 21st century

Data integration and modelling strategies

Luiz Max Carvalho



THE UNIVERSITY
of EDINBURGH

Acknowledgements



Andrew Rambaut
University of
Edinburgh



Leo Santos
CEMADEN

Who's this guy?

- ⊙ Have a degree in Biology, can't tell an insect from a spider;
- ⊙ PhD student in Evolutionary Biology (!) at the University of Edinburgh;
- ⊙ Work mainly in **Quantitative Biology**;
- ⊙ Interests include **Markov chain Monte Carlo, complex networks** and **risk modelling**.

Evolutionary epidemiology

Concepts and tools.

Evolutionary epidemiology

Concepts and tools.

Challenges and opportunities

Methodological issues, data collection and handling.

Evolutionary epidemiology

Concepts and tools.

Challenges and opportunities

Methodological issues, data collection and handling.

The role of mathematics

Which areas of mathematics are more heavily involved.

Evolutionary epidemiology

Concepts and tools.

Challenges and opportunities

Methodological issues, data collection and handling.

The role of mathematics

Which areas of mathematics are more heavily involved.

Much more work is needed

We should prepare for an era of plenty.

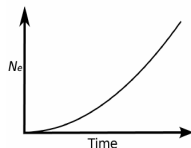
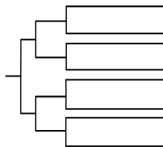
Phylogenetics of fast-evolving viruses

Inferring spatial and temporal dynamics from genomic data:

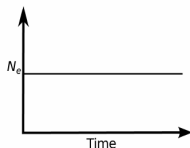
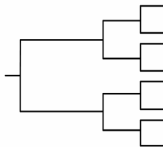
Phylogenies*!

* plus complicated models

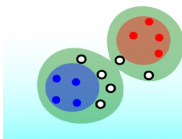
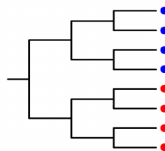
A
Exponential Growth



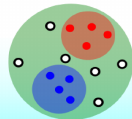
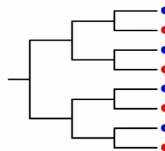
B
Constant Population Size



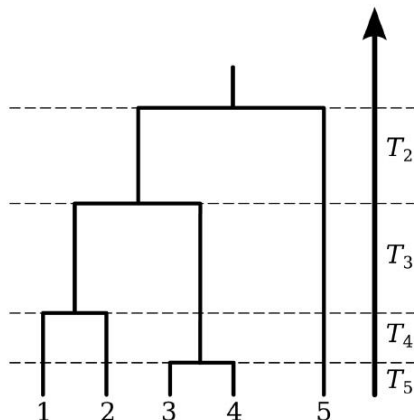
A
Structured Host Population



B
Unstructured Host Population



Trees and the coalescent



Let T_n denote the time for n lineages to *coalesce*, i.e., merge into one ancestral lineage, in a population of size N_e . Then:

$$\Pr(T_n = t) = \lambda_n e^{-\lambda_n t}$$

$$\lambda_n = \binom{n}{2} \frac{1}{N_e} = \binom{n}{2} \frac{1}{N_e \tau}$$

where N_e is the effective population size and τ is the generation time. Let T_{mrca} denote the age of the most recent common ancestor:

$$\begin{aligned}\mathbb{E}[T_{\text{mrca}}] &= \mathbb{E}[T_n] + \mathbb{E}[T_{n-1}] + \dots + \mathbb{E}[T_2] \\ &= 1/\lambda_n + 1/\lambda_{n-1} + \dots + 1/\lambda_2 \\ &= 2N_e \left(1 - \frac{1}{n}\right)\end{aligned}$$

Figure: Figure 4 from Volz et al. (2013).

Data Integration I: Ebola epidemics in West Africa

[animation]

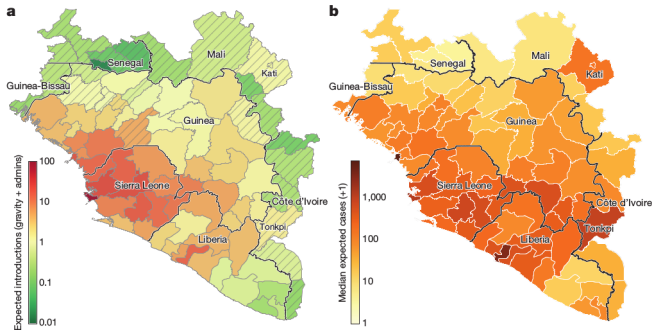
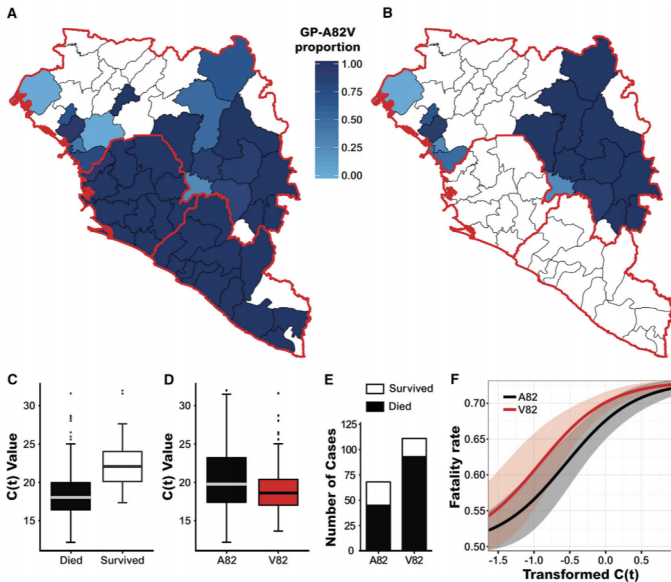


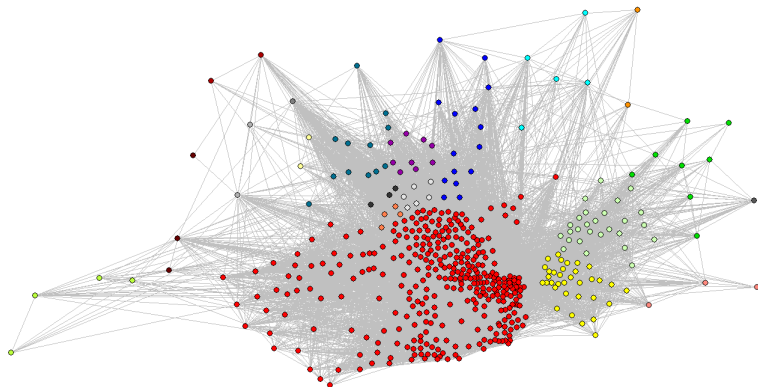
Figure 4 | Predicted destination consequences of viral dispersal
number of EBOV imports into 63 regions in Guinea, Sierra Leone, Liberia (including 7 without records in Guinea) and the surrounding of the neighbouring countries of Bissau, Senegal, Mali and Côte d'Ivoire. **a**, Expected number of EBOV exports from locations in the phylogeographic imports to any location were calculated on the basis of the phylogeographic estimates and associated predicted extended to apparently EVD-free (Supplementary Methods). **b**, Predicted cluster sizes from the Bayesian case data.

Data Integration II: GPA82V mutation and mortality



Potential applications

- ⊙ Human mobility + case data = epidemiological models of spread and maintenance;
- ⊙ Genomic data + environmental data = predictions of flow and case counts (e.g. Leptospirosis).



Phylodynamics is a powerful tool

DNA sequences from pathogens +
environmental/socio-economic data can give us insight

Phylodynamics is a powerful tool

DNA sequences from pathogens +
environmental/socio-economic data can give us insight

Searching trees is **hard**

Developing better statistical models and computational tools
is crucial.

Phylogenetics is a powerful tool

DNA sequences from pathogens +
environmental/socio-economic data can give us insight

Searching trees is **hard**

Developing better statistical models and computational tools
is crucial.

Data integration is crucial

We need more and better data.

Take home

Phylogenetics is a powerful tool

DNA sequences from pathogens +
environmental/socio-economic data can give us insight

Searching trees is **hard**

Developing better statistical models and computational tools
is crucial.

Data integration is crucial

We need more and better data.

Nature is complicated

We need better models to go along.

THE
END