

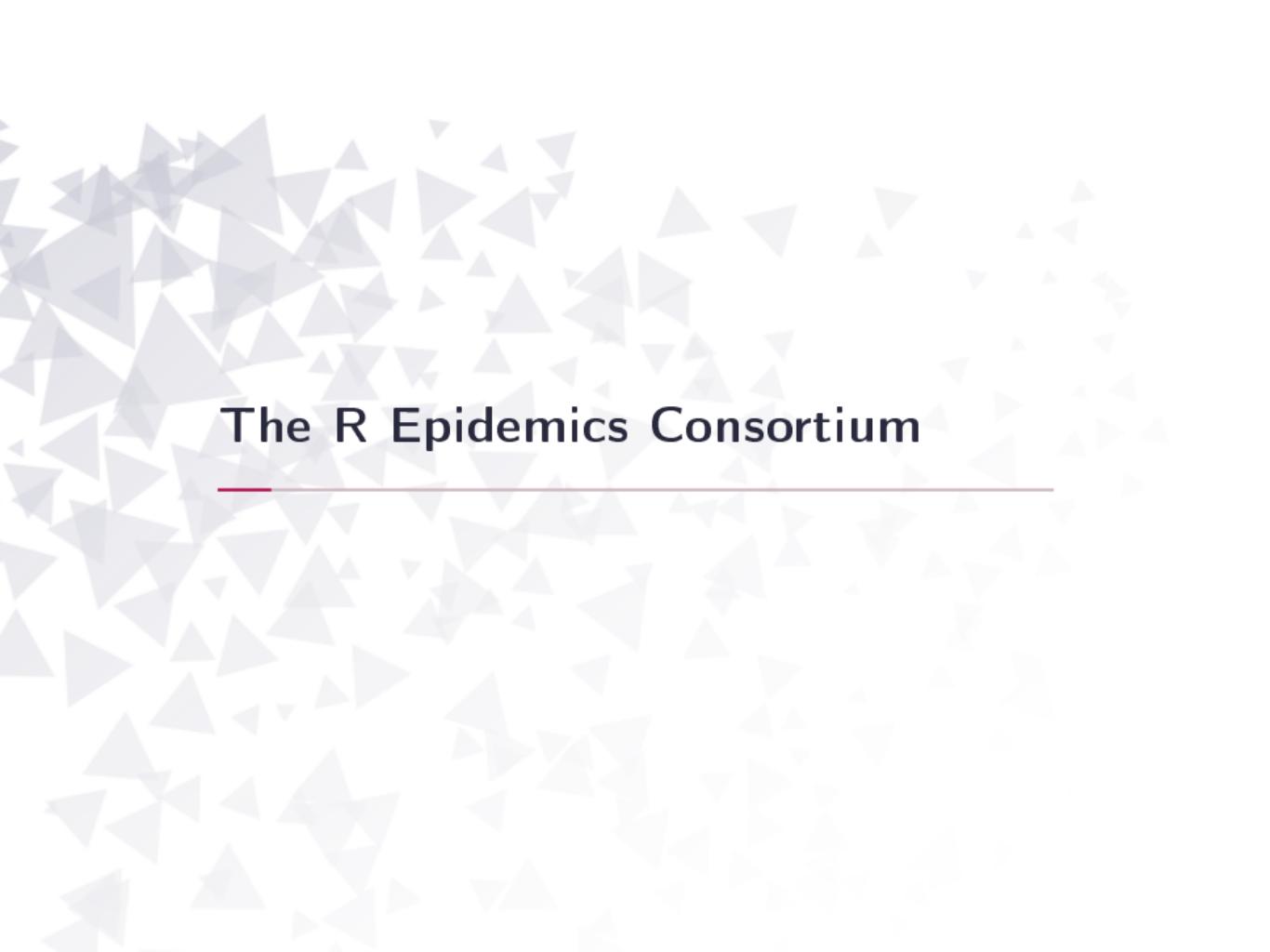
RECON

Building the next generation of statistical tools for outbreak response using 

Thibaut Jombart

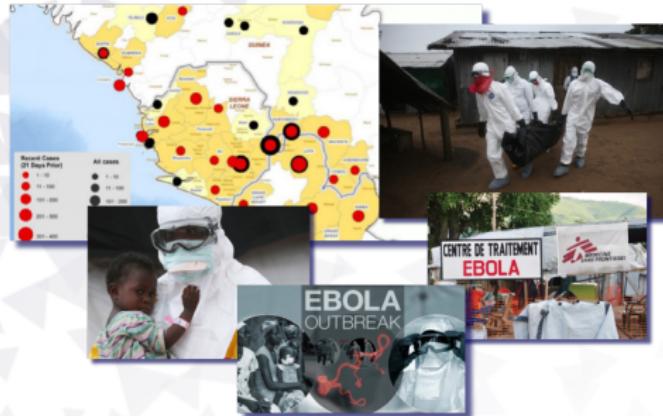
Epidemics6 - EPID2017_0085

Imperial College London
MRC Centre for Outbreak Analysis and Modelling



The R Epidemics Consortium

Lessons learnt from the Ebola response



Lessons learnt from the Ebola response



Lessons learnt from the Ebola response

WHO Ebola response team

Help improving situation awareness

Imperial College Ebola team

Timeline:

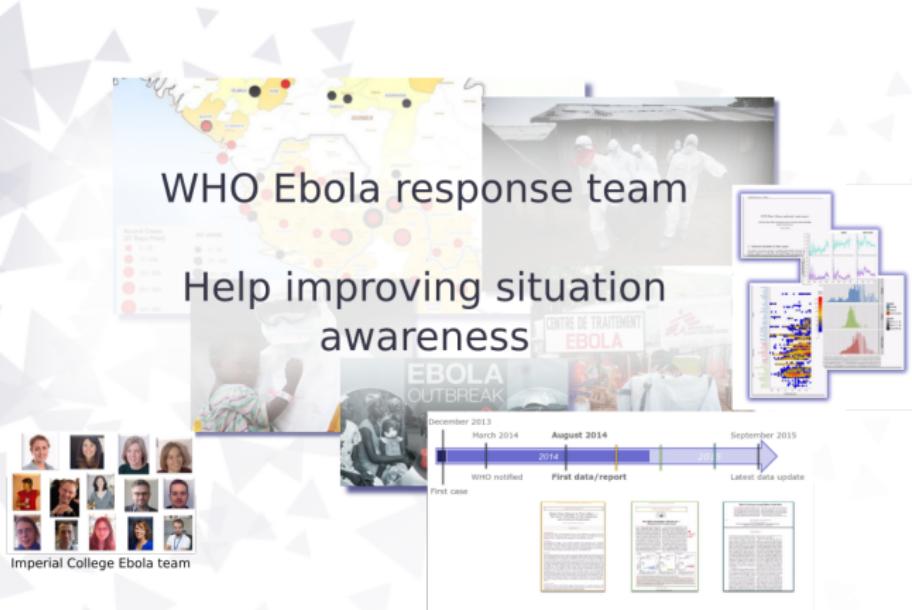
- December 2013: First case
- March 2014: WHO notified
- August 2014: First data/report
- September 2015: Latest data update

Centre de Traitement EBOLA

EBOLA OUTBREAK

WHO Ebola response team dashboard

Lessons learnt from the Ebola response



Most statistical/modelling tools for situation awareness were missing.

Who do we need to develop these tools?



Who do we need to develop these tools?

Public Health



Who do we need to develop these tools?



Who do we need to develop these tools?



From a hack to a pack



Hackout 3, summer 2016, Berkeley

From a hack to a pack



Hackout 3, summer 2016, Berkeley



A word cloud illustrating the evolution of the project from a hackathon to a comprehensive platform. The words are arranged in a grid-like structure, with larger words indicating higher frequency or importance.

Key terms include: functional, incubation, userfriendly, secure, dictionary, systems, testing, automated, continuous, collection, series, repository, tools, fast, outbreaks, parsing, secured, bias, reporting, integration, gui, vhl, integration, reporting, delay, interface tree, cdc, site, report, transmission, censoring, reproducible, package, reproduction, period, exposure, mutations, linelist, situations, synchronised, anonymised, epistemic, contact, epiinfo, compiled, clean, time, interface tree, fellow, symptoms, outbreaker, interface tree, follow, tracing, shiny, automation, incidence, cdc, cleaning, cdo, edc, bayesian, sitereport, estimation, curation, model, ggplot, clusters, rates, reliable, contacttracing, parallel, parameters, epidemics, genomics, distribution, assessments, encrypted, annotations, annotations, annotations.

From a hack to a pack



Hackout 3, summer 2016, Berkeley

A large cluster of words representing the tools and concepts used in the project, including:
functional incubation
userfriendly secure dictionary
systems testing automated continuous
collection series repository
rpp efficiency number fast
secured bias outbreaks
parsing code integration
reporting gui vhl
unit data delay interface tree
epidemiologist contact follow
epilist compiled
outbreaker symptoms
tracing lineelist shiny
automation automation cdc
epicontacts ggplot edic
dashboard clusters rates
parallel reliable contacttracing
parameters epidemics genomics
distribution

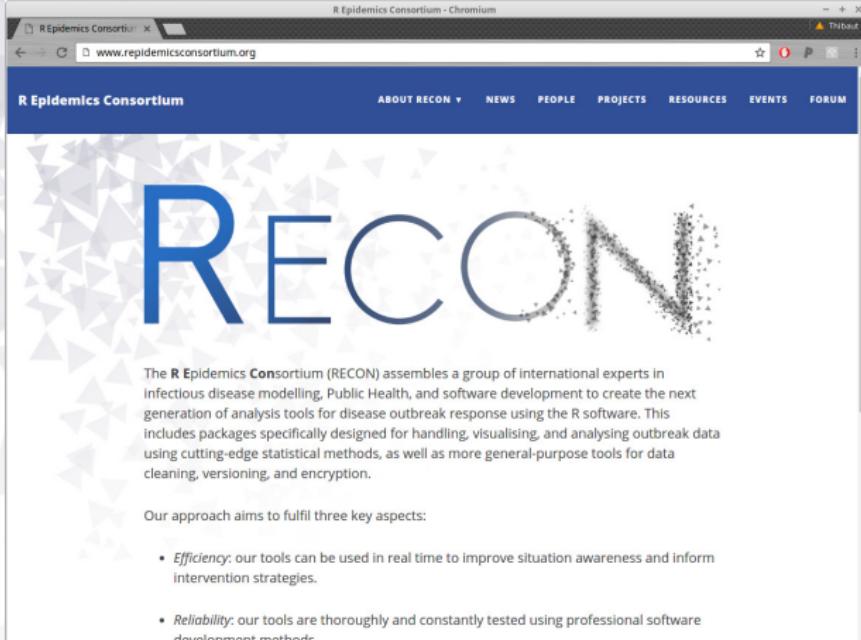
RECON
[The R Epidemics Consortium]

From a hack to a pack



RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in .



The screenshot shows a web browser window for the "R Epidemics Consortium - Chromium" at www.repidemicsconsortium.org. The page features a large, stylized title "RECON" where the letters are composed of small, dark grey dots. Below the title is a paragraph of text describing the consortium's purpose and approach. At the bottom, there is a bulleted list of three key aspects.

R Epidemics Consortium

ABOUT RECON NEWS PEOPLE PROJECTS RESOURCES EVENTS FORUM

RECON

The **R Epidemics Consortium** (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption.

Our approach aims to fulfil three key aspects:

- *Efficiency*: our tools can be used in real time to improve situation awareness and inform intervention strategies.
- *Reliability*: our tools are thoroughly and constantly tested using professional software development methods.

www.repidemicsconsortium.org

RECON

www.repidemicsconsortium.org

- started 6th September 2016
- ~75 members
- 20 countries, > 50 institutions
- 7 packages/projects released, 20 under development
- public forum, blog, online resources

Statistical software

- **efficiency**: useful for improving situation awareness in real time
- **reliability**: outputs can be trusted
- **accessibility**: widely available, easy learning curve

RECON: activities

Statistical software

- **efficiency**: useful for improving situation awareness in real time
- **reliability**: outputs can be trusted
- **accessibility**: widely available, easy learning curve

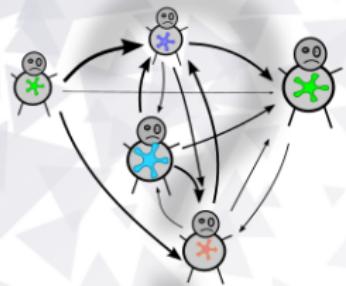
Translation

- **disseminating knowledge**: free online training material, involvement with FETPs, workshops
- **supporting response**: portable data analysis environment, deployment to the field

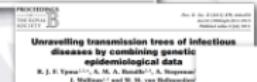
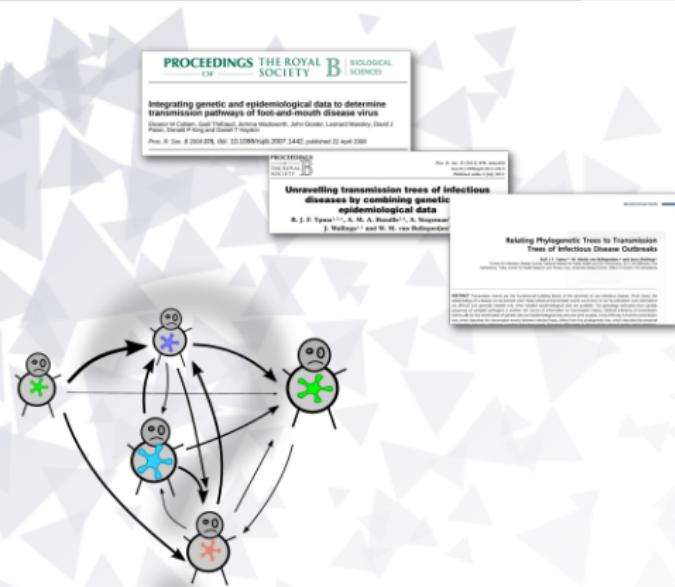
The background of the slide features a large number of small, light-gray triangles of various sizes scattered across the entire area, creating a subtle geometric pattern.

Current activities

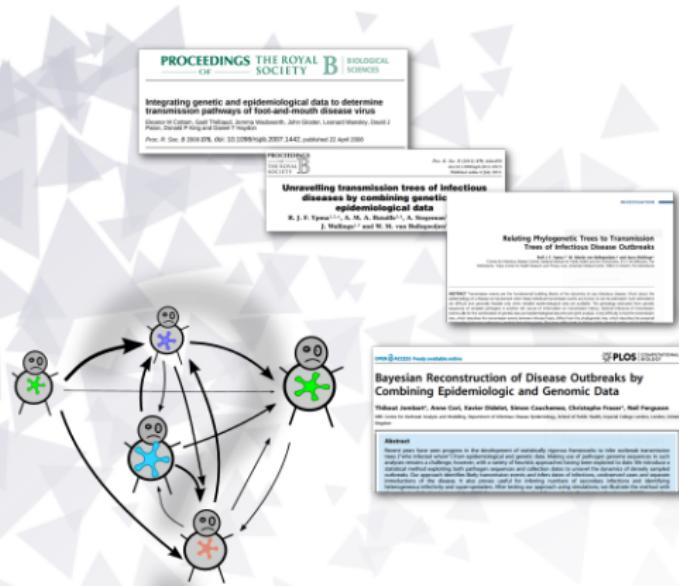
Who infects whom? Many answers for a single question



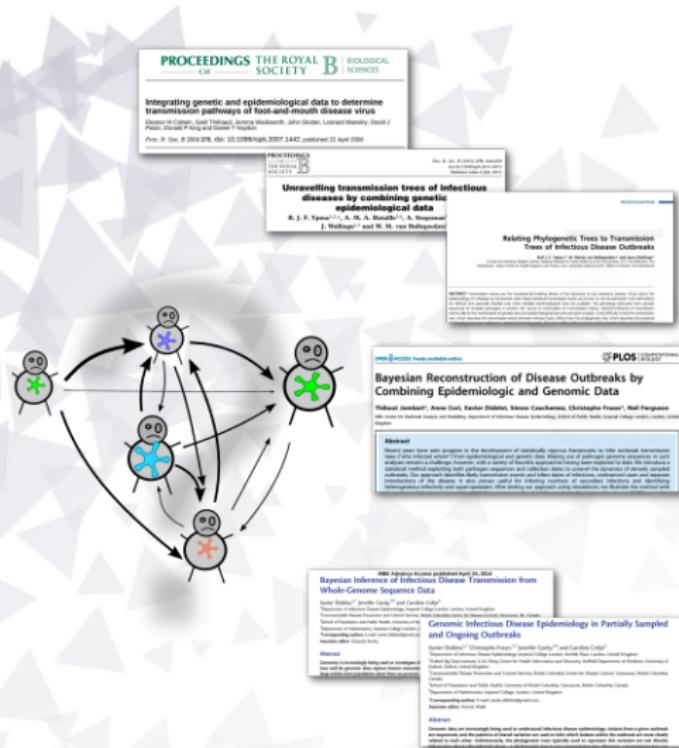
Who infects whom? Many answers for a single question



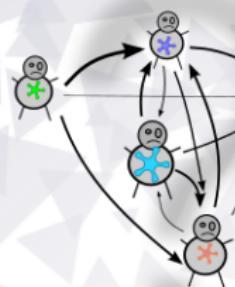
Who infects whom? Many answers for a single question



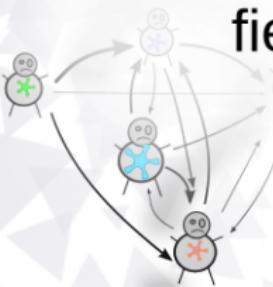
Who infects whom? Many answers for a single question



Who infects whom? Many answers for a single question



Who infects whom? Many answers for a single question



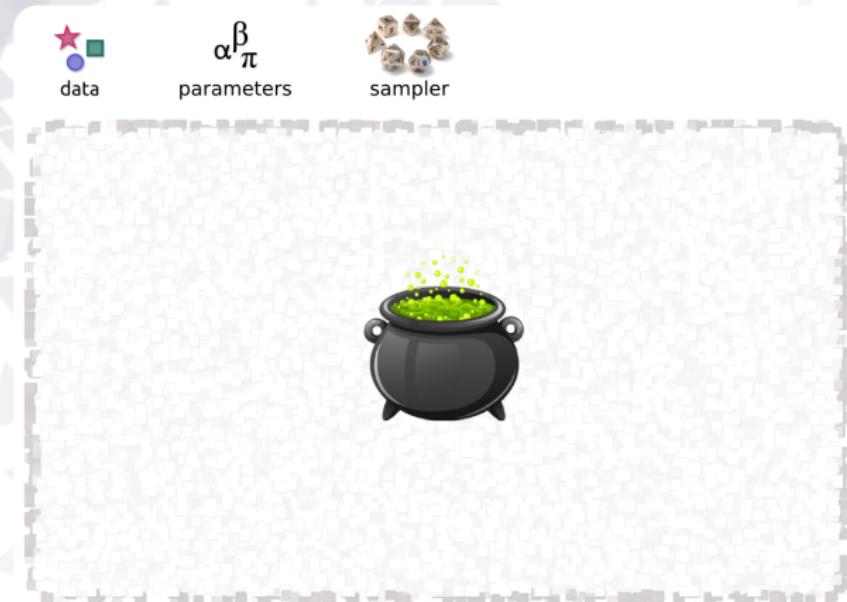
But fast growing methodological fields can get messy!



Excoffier and Heckel 2006, Nature Reviews Genetics

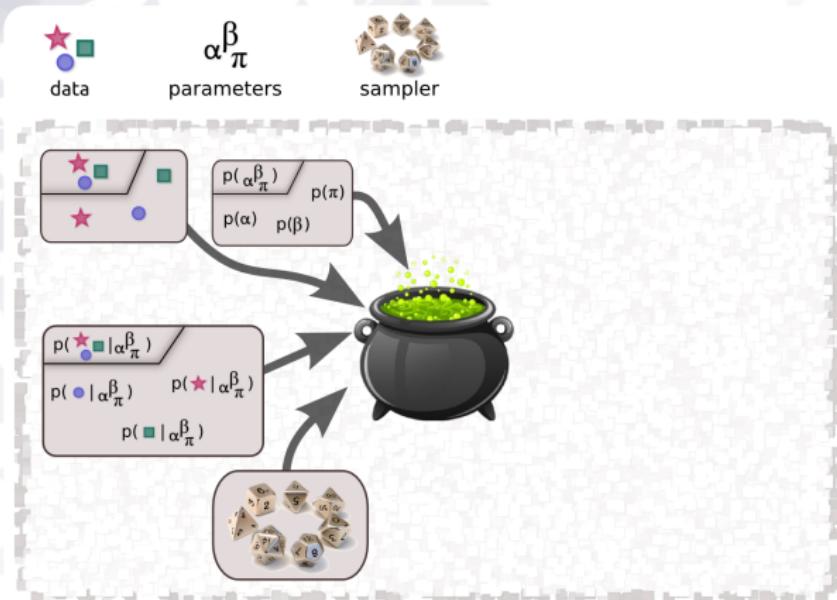
outbreaker2: a general tool for outbreak reconstruction

A **modular** framework for outbreak reconstruction: bring your own data, prior, likelihood, MCMC.



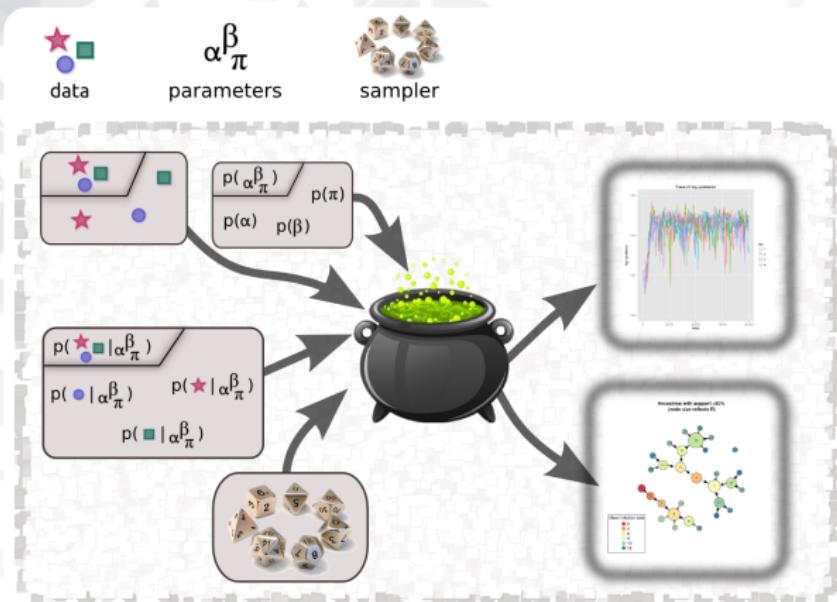
outbreaker2: a general tool for outbreak reconstruction

A **modular** framework for outbreak reconstruction: bring your own data, prior, likelihood, MCMC.



outbreaker2: a general tool for outbreak reconstruction

A **modular** framework for outbreak reconstruction: bring your own data, prior, likelihood, MCMC.



outbreaker2: check poster 3.036 by Finlay Campbell

outbreaker2: Inferring transmission trees from temporal, genomic and contact data

Finlay Campbell¹, Anne Corl¹, Neil Ferguson¹, Thibaut Jombart¹
MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, United Kingdom

Introduction

- Infering "who infected whom" in an infectious disease outbreak provides valuable epidemiological insights which can improve infection response.
- Most outbreak reconstruction tools rely on whole genome sequencing (WGS) data to infer likely transmission chains, but this approach is slow and expensive.
- Here we describe first a simulation study to evaluate the performance of outbreak reconstruction tools, and secondly the development of RECON¹, a tool that integrates multiple types of outbreak data in a single framework.

① When is WGS data useful?

- Most outbreak reconstruction tools rely on highly informative WGS to infer likely transmission chains between cases [2]. To test if this approach is beneficial, we first simulated outbreaks with known transmission events for 10 major pathogens and attempted to reconstruct these using outbreak reconstruction tools.
- We defined a metric, namely the "transmission divergence", to quantify the amount of genetic diversity expected for different pathogens. In general, for most pathogens, disease WGS provided little information, with low genetic diversity acting as a limiting factor.

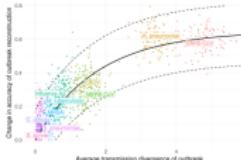
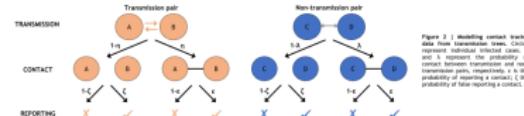


Figure 1 | Implementation of accuracy of outbreak reconstruction now incorporating WGS data. We simulated outbreaks with known transmission events for 10 major pathogens and attempted to reconstruct these using outbreak reconstruction tools. Disease WGS provided little information, with low genetic diversity acting as a limiting factor.

② Modelling contact data

- A valuable complement to WGS data in inferring transmission pairs is contact data, which is routinely collected in outbreak settings yet remains underutilised in outbreak reconstruction tools.
- We developed a probabilistic model that relates a transmission tree to an observed set of contact data, while accounting for incomplete reporting.



Links & References

- outbreaker2 website: <http://www.recongenomics.org/outbreaker2>
- outbreaker2 development: <https://github.com/recongenomics/outbreaker2>
- 1. Ferguson N, Riley S, Edmunds W. Transmission potential and generation intervals in human infectious diseases. *Philos Trans R Soc Lond B Biol Sci*. 2011;366:288–297.
- 2. Ferguson NM, Riley S, Ferguson NM, et al. A Bayesian approach to inferring transmission parameters for emerging infectious disease outbreaks. *J Clin Microbiol*. 2003;41:1599–1606.
- 3. Campbell F, Stirling K, Phillips R, Ferguson N, Corl A, Jombart T. *PLoS Comput Biol*. 2011; (Submitted).



RECON

③ Implementing custom models

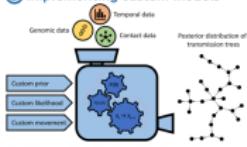


Figure 3 | Outbreaker2 is a modular and customizable tool for outbreak reconstruction. By combining different data types (genomic, contact, temporal), users can implement their own models (Figure 3) [4]. By combining a prior (e.g., a phylogenetic tree) with a likelihood function, posterior distributions can be calculated. By combining different likelihood functions, sophisticated algorithms can be achieved with relative ease (Figure 4).

This generates a new paradigm, facilitates model comparison and allows focus on methodological developments rather than implementation.

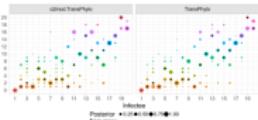


Figure 4 | Implementing the TimetreePhy likelihood into Outbreaker2 to increase high-level disease behaviour. Outbreaker2 can now implement the TimetreePhy likelihood, represented using the original TimetreePhy [2] as well as the Outbreaker2 algorithm with a custom likelihood function. This allows users to implement their own models (Figure 3) [4].

The TimetreePhy likelihood allows users to incorporate time into the likelihood calculation, as well as the posterior distribution of infections for a given case. Black circles represent true ancestors.

try outbreaker2

- of contacts, false-positive reporting of contacts as well as non-infectious mixing between cases (Figure 2), and implements this in Outbreaker2.
- the user can weight the contact data by context-specific parameters, e.g. the probability of reporting a contact π_c , the user can weight the contact data in a context-specific and statistically rigorous manner.

New transmission

- Transmission data (Figure 2) from laboratory tests, clinical records, and other sources can be used to inform the transmission process. π_{trans} represents the probability of transmission given a contact pair, π_{rec} is the probability of becoming infected given a contact, π_{rep} is the probability of false reporting a contact.

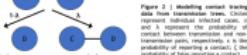


Figure 5 | Modeling new transmission data from laboratory tests, clinical records, and other sources can be used to inform the transmission process. π_{trans} represents the probability of transmission given a contact pair, π_{rec} is the probability of becoming infected given a contact, π_{rep} is the probability of false reporting a contact.

Conclusion

- outbreaker2 is an R package for transmission tree inference that integrates contact data in a probabilistic manner for the first time.
- by providing a modular and customizable platform, we hope to provide a unified platform for outbreak reconstruction tools, and encourage the development of extensions by the wider scientific community.

- modular platform for outbreak reconstruction

- limits of how informative WGS are

- model for contact data

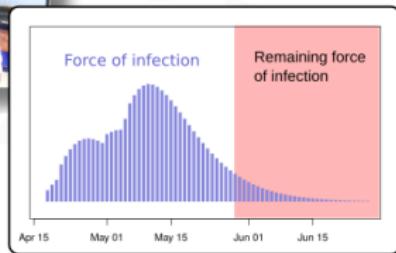
- phylogenetic module

- this afternoon 17:40-19:10

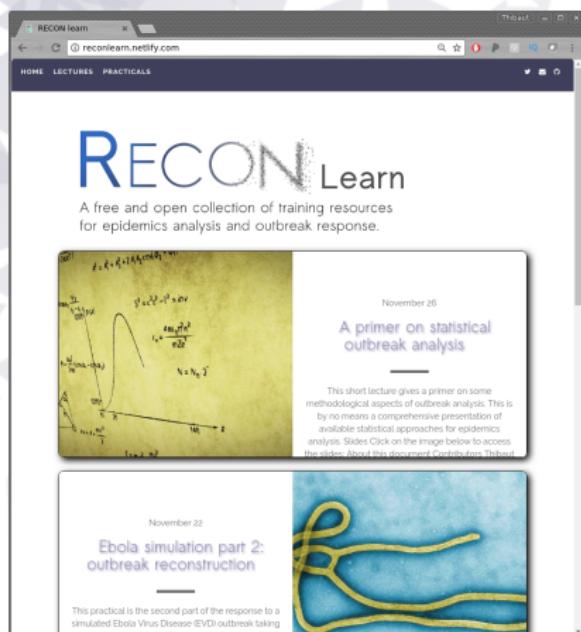
Ebola outbreak, Likati (DRC) 2017



- EVD outbreak May 2017
- contact data visualisation tools used in contact tracing
- simple model informed response (scaling)
- end: 2nd July 2017



RECON learn: training resources for epidemics analysis



reconlearn.netlify.com

- repository for free, open training material
- lectures, practicals, case studies, code gists
- emphasis on community contributions
- workshops in 2018: CDC, MSF, WHO, EPIET Alumni Network, ...

The background of the slide features a large number of small, light-gray triangles of various sizes scattered across the entire area, creating a subtle geometric pattern.

To conclude

A role beyond modelling

- modelling plays a central role in **advising public health policies**

A role beyond modelling

- modelling plays a central role in **advising** public health policies
- methodology needs to be **open** and **transparent**

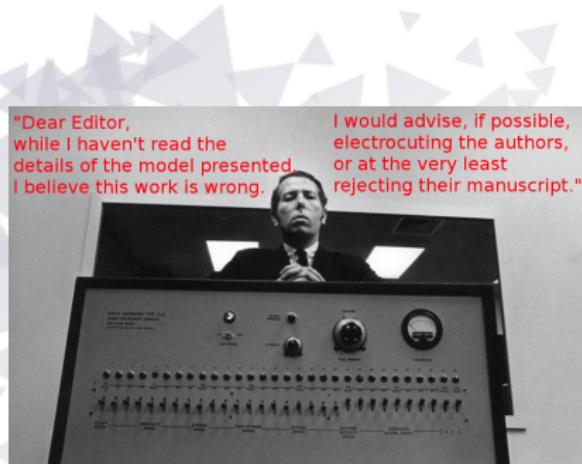
A role beyond modelling

- modelling plays a central role in **advising public health policies**
- methodology needs to be **open** and **transparent**
- RECON promotes the development of efficient and reliable tools

A role beyond modelling

- modelling plays a central role in **advising public health policies**
- methodology needs to be **open** and **transparent**
- RECON promotes the development of efficient and reliable tools
- (ideally free) training is essential to ensure knowledge dissemination

Open science, closed publishing?



[Milgram experiment, 1963]

Determinants of abusive behaviours

- position of power
- impunity / transfer of responsibility
- having been a victim of abuse
- personal gain

Open science, closed publishing?



[Milgram experiment, 1963]

Determinants of abusive behaviours

- position of power
- impunity / transfer of responsibility
- having been a victim of abuse
- personal gain

All these feature our current reviewing system.

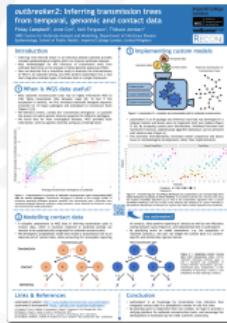
A simple solution



“Everyone’s worried about stopping bad reviewing practices. Well, there’s really an easy way: Stop participating in it.”

Thanks to:

- **colleagues:** Finlay Campbell, Amrish Baidjoe, Anne Cori, Rich Fitzjohn, Xavier Didelot, James Hayward, Neil Ferguson
- **groups:** RECON members, WHO Ebola Likati Response Team, GOARN
- **funding:** HPRU-NIHR, MRC
- **poster:**



Poster 3.036, today, 17:40-19:10.