

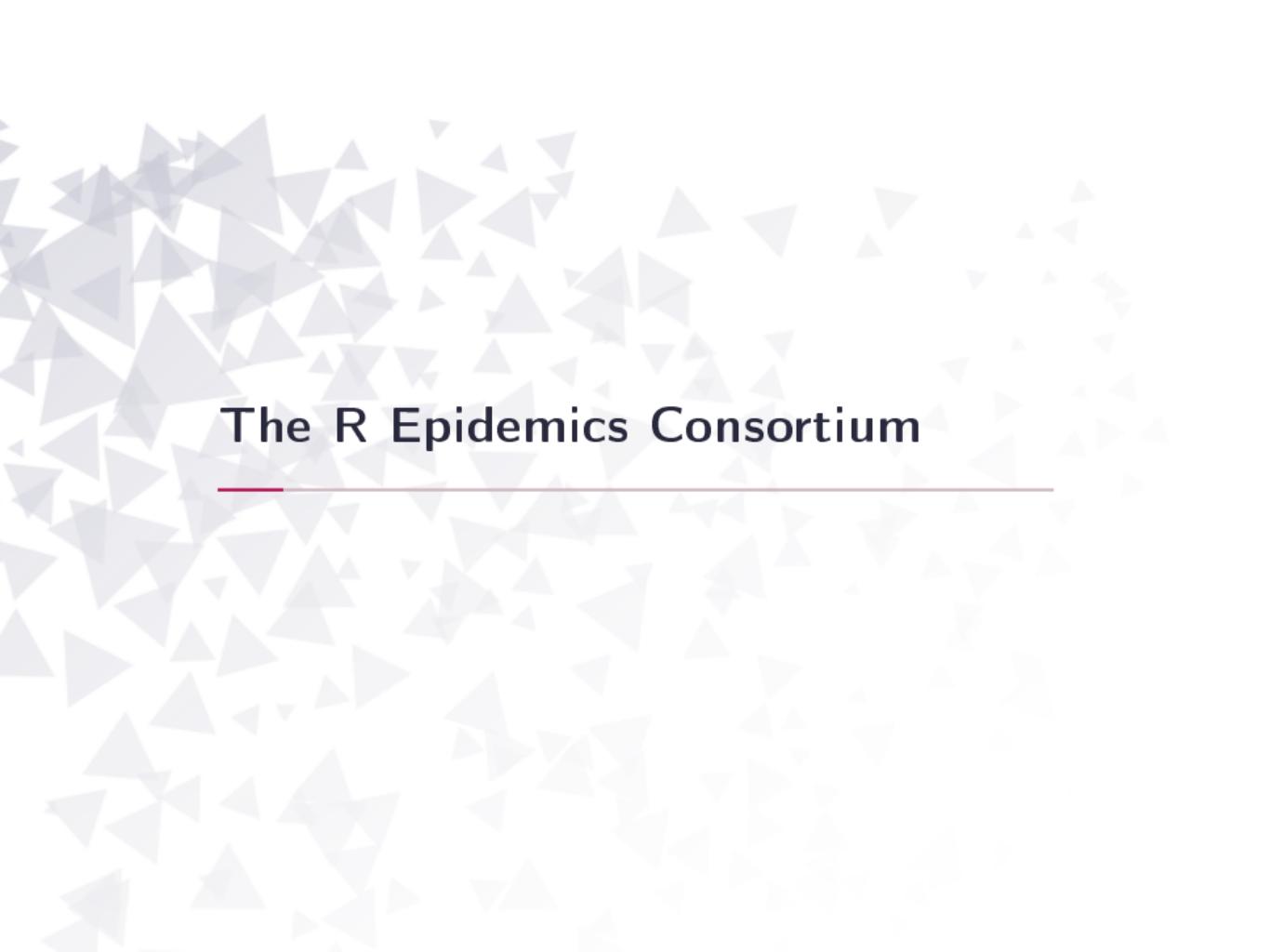
RECON

Building the next generation of statistical tools for outbreak response using 

Thibaut Jombart

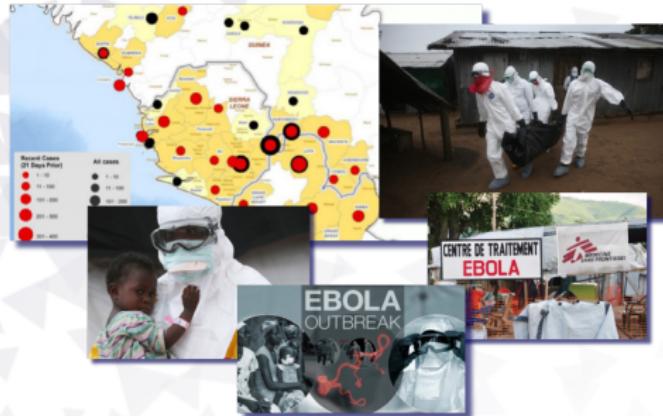
Epidemics6 - EPID2017_0085

Imperial College London
MRC Centre for Outbreak Analysis and Modelling

The background of the slide features a large number of small, light-gray triangles of various sizes scattered across the frame, creating a sense of depth and movement.

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:

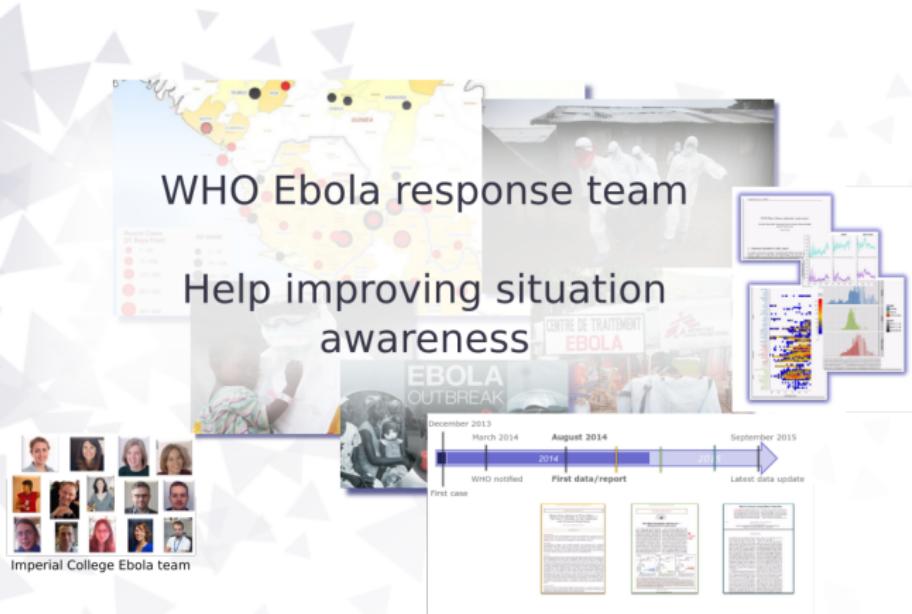
- December 2013: First case
- March 2014: WHO notified
- August 2014: First data/report
- September 2015: Latest data update

Centre de Traitement EBOLA

EBOLA OUTBREAK

WHO Ebola Outbreak Response 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



From a hack to a pack



Hackout 3, summer 2016, Berkeley

functional
incubation
userfriendly secure dictionary
systems testing automated continuous
collection series repository
rpp efficiency number fast
secured bias outbreaks
parsing code integration
reporting gui
unit data delay
epidemiology security peak
situation anonymised
opensource contact
epiinfo delay
clean compiled
outbreaker interface tree
symptoms lineelist fellow
tracing shiny
automation cdc
epicontacts edc
ggplot cleaning
dashboard clusters rates
parallel reliable
parameters contacttracing
epidemics genomics
distribution

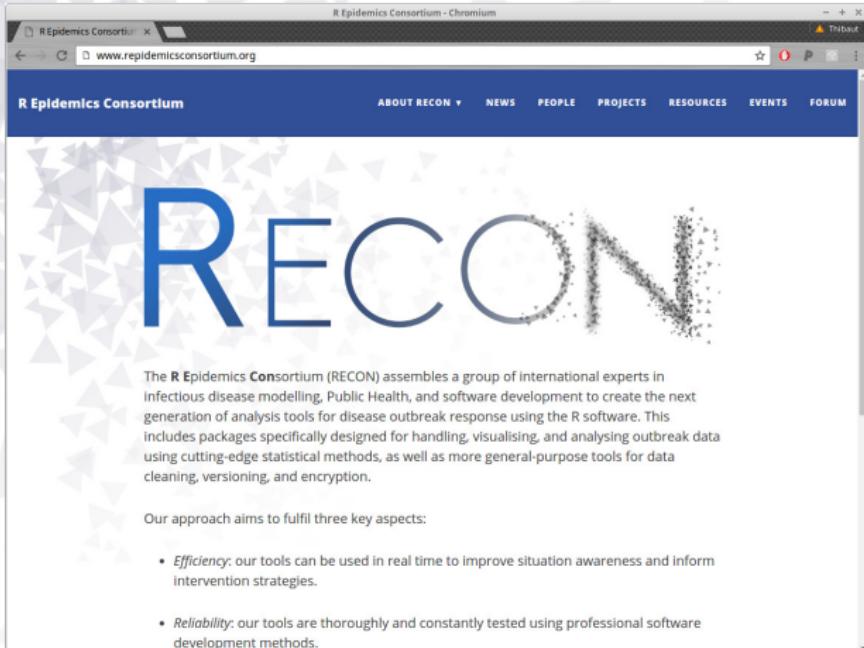
RECON
[The R Epidemics Consortium]

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 dots and lines, suggesting a network or data structure. Below the title is a descriptive paragraph about the consortium's purpose and approach. A bulleted list at the bottom outlines three key aspects of their work.

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

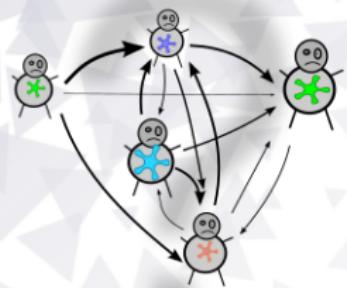
Translation

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

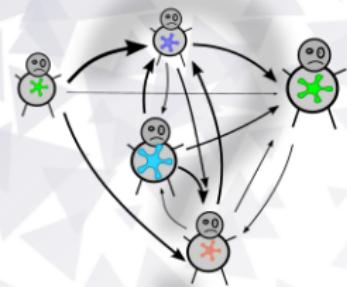
The background of the slide features a large number of small, light-gray triangles of various sizes scattered across the entire area, creating a subtle geometric pattern.

Current activities

Who infects whom? Many answers for a single question



Who infects whom? Many answers for a single question

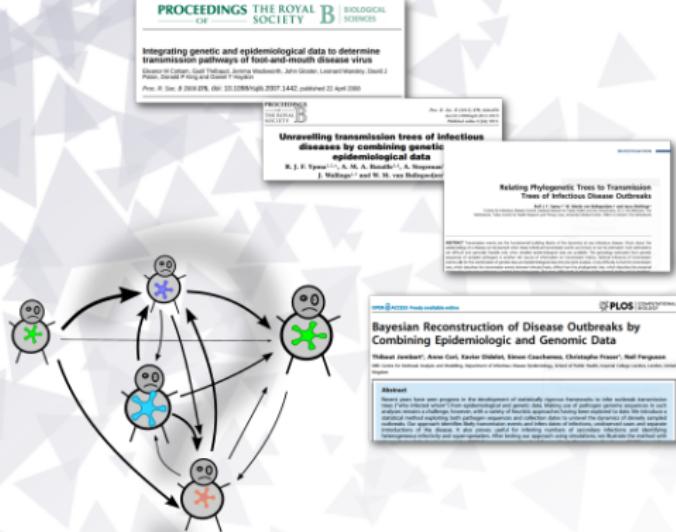


Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data

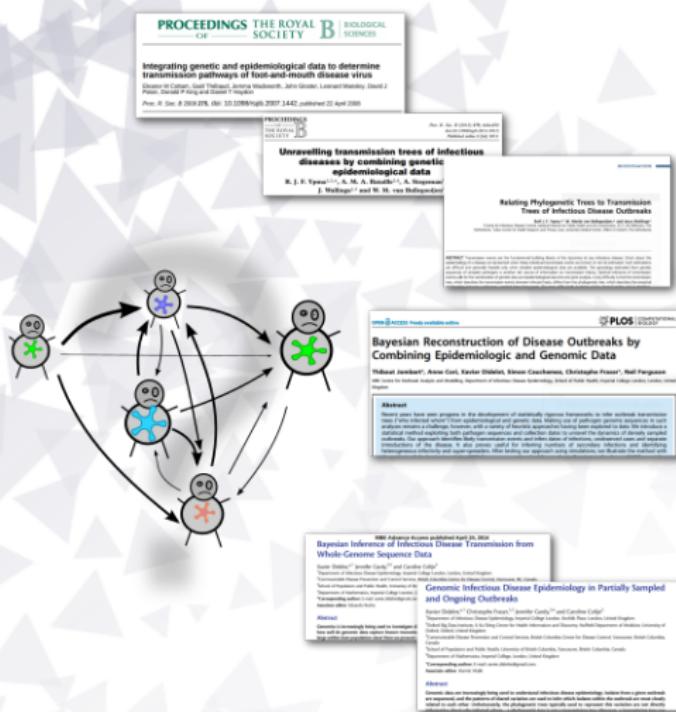
R. J. E. Verwoerd¹, S. M. A. Bouwink¹, J. A. Nijhuis², J. H. Hulst² & W. H. van den Brink¹



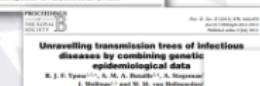
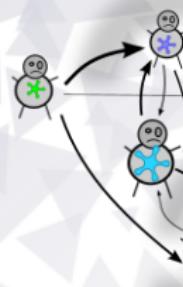
Who infects whom? Many answers for a single question



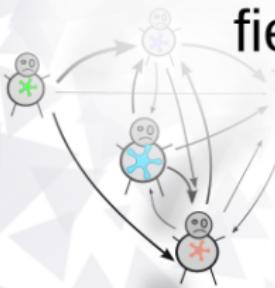
Who infects whom? Many answers for a single question



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!

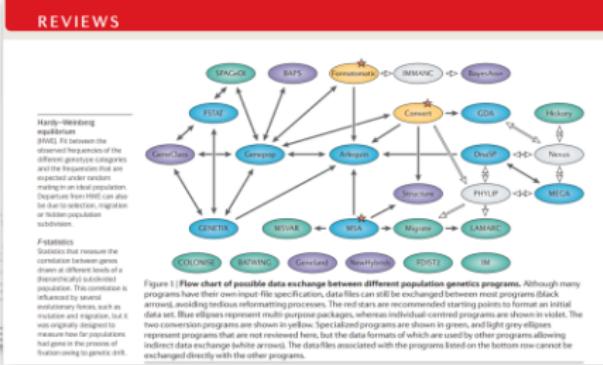
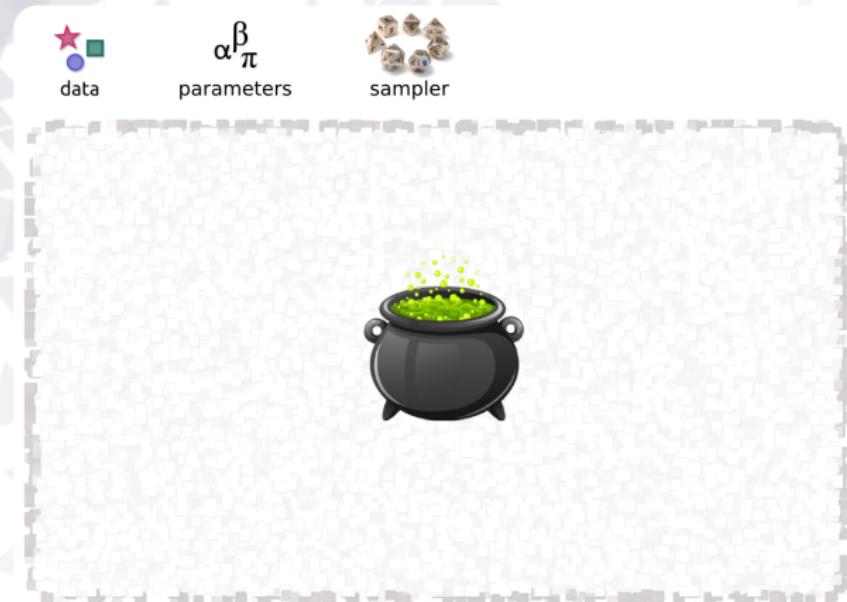


Figure 1 | Flow chart of possible data exchange between different population genetics programs. Although many programs have their own input–file specifications, data files can still be exchanged between most programs (black arrows), avoiding tedious reformatting processes. The red stars are recommended starting points to format an initial data set. Blue ellipses represent multi-purpose packages, while the individual programs shown in violet, The programs shown in grey are those that are no longer maintained. The programs are shown in groups of light grey to represent programs that are not reviewed here, but the data formats of which are used by other programs allowing indirect data exchange (bulky arrows). The data files associated with the programs listed on the bottom row cannot be exchanged directly with the other programs.

Excoffier and Heckel 2006, Nature Reviews Genetics

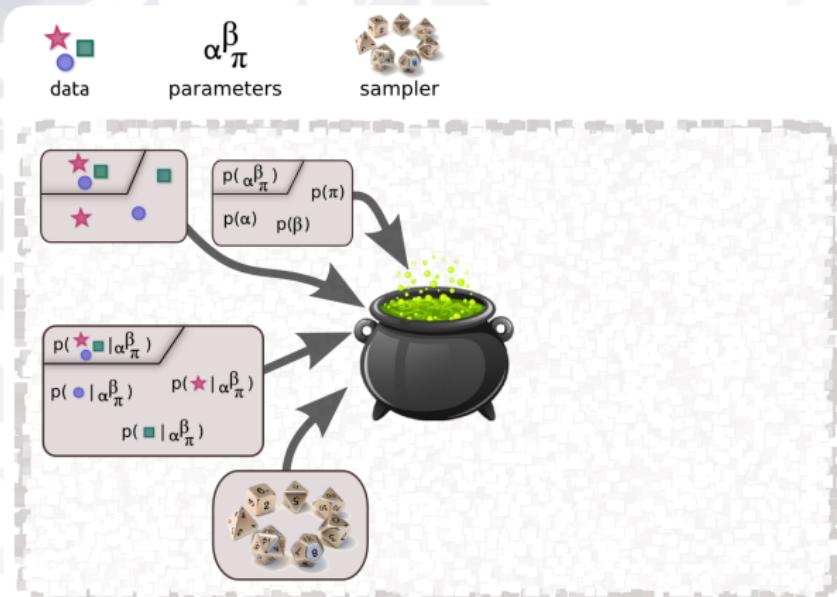
outbreaker2: a general tool for outbreak reconstruction

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



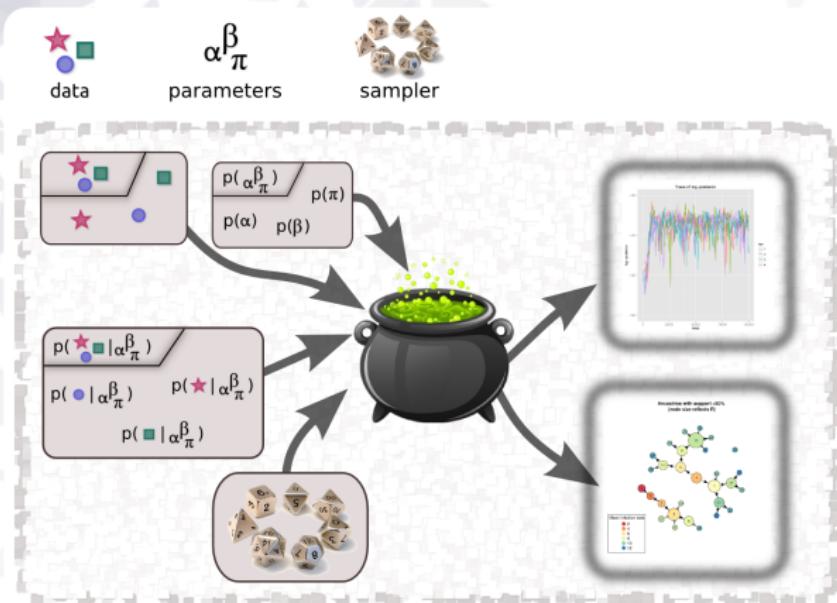
outbreaker2: a general tool for outbreak reconstruction

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

- Inferring "who infected whom" in an infectious disease outbreak provides valuable epidemiological insights which can improve infection response.
- Most outbreak reconstruction tools rely on highly informative WGS to infer likely transmission chains between cases [2]. To test if this approach is feasible, we first simulated outbreaks using synthetic sequence evolution for 10 major pathogens and attempted to reconstruct these [3].
- Here we describe first a simulation study to evaluate the performance of outbreak reconstruction tools, and secondly, RECON [1], 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 feasible, we first simulated outbreaks using synthetic sequence evolution for 10 major pathogens and attempted to reconstruct these [3].
- 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.

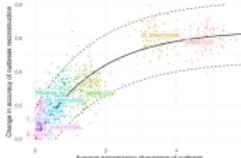
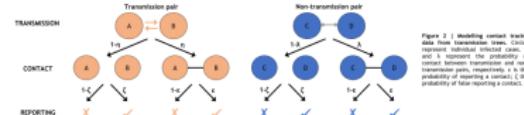


Figure 1 | Implementation of accuracy of outbreak reconstruction incorporating WGS data. We simulated outbreaks using synthetic sequence evolution for 10 major pathogens and reconstructed transmission pathways generated from transmission pairs. Outbreaks were simulated with increasing average transmission divergence. Shaded areas around the lines represent 95% quantile intervals of parameter values obtained by 100 runs and reconstructed using outbreak2.

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.reconimperial.ac.uk/outbreaker2>
- outbreaker2 development: <https://github.com/reconimperial/outbreaker2>
- [1] Campbell F, Ferguson N, Corl A, Jombart T. *PLoS One*. 2014; 9:e93610.
- [2] Ferguson NM, Riley S, Edmunds JW, et al. *PLoS Med*. 2003; 1:e12.
- [3] Campbell F, Ferguson NM, Riley S, Corl A, Jombart T. *PLOS Comput Biol*. 2011; (Submitted).



RECON

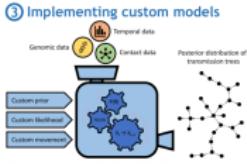


Figure 3 | Outbreaker2 is a modular and customizable tool for outbreak reconstruction.

- outbreaker2 is an R package and inference tool that was developed in a modular manner and allows users to implement their own models (Figure 3) [4]. By combining prior knowledge with contact and genomic data, more complex, sophisticated algorithms can be achieved with relative ease (Figure 4).
- The generic architecture facilitates model comparison and allows focus on methodological developments rather than implementation.

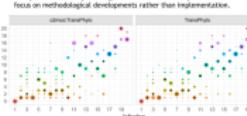


Figure 4 | Evaluating the TruePath algorithm. Outbreaker2 can increase high-level pathogen behaviour. Transmission pairs are represented using the original TruePath algorithm [5] as well as the outbreaker2 algorithm. The true posterior distribution of infections for a given case is shown as a solid black line. Black crosses represent true ancestors.

try outbreak2

- 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.

Non-infectious mixing

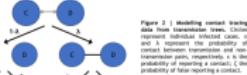


Figure 5 | Modelling contact mixing data. Note: transmission pairs are not necessarily unique, i.e. one transmission pair may correspond to many contact pairs. Contact mixing is modelled by a transmission pair (i, j) representing the probability of transmission pairs i and j representing the probability of non-infectious mixing between cases i and j . The probability of reporting a contact π_c is the probability of reporting a contact i , i.e. the probability of reporting a contact i .

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 extensible framework, outbreaker2 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

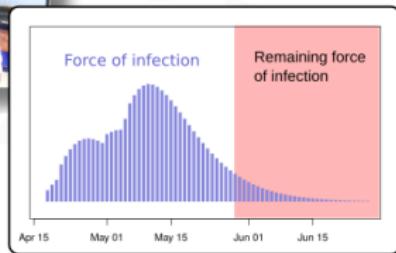
- phylogenetic module

- this afternoon 17:40-19:10

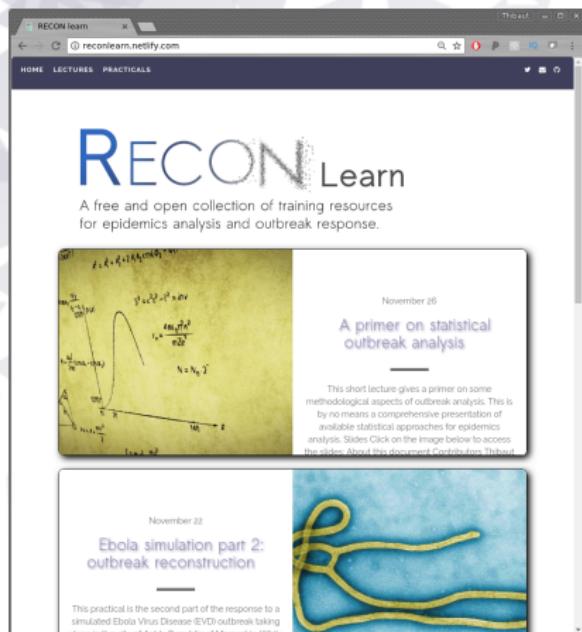
Ebola outbreak, Likati (DRC) 2017



- 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

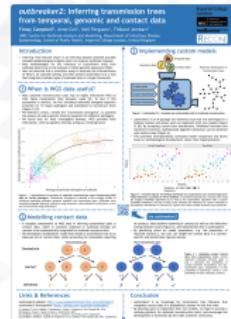


reconlearn.netlify.com

- 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, ...

Thanks to:

- **conference organisers**
- **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.