

Expanding the toolbox of outbreak analytics

When are genomic data useful?

Thibaut Jombart

27 November 2018

HPRU away day



@teebzr

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.

Context

RECON is now an NGO!

RECON 2.0: going official - Chromium

RECON 2.0: going official https://www.repidemicsconsortium.org/2018-10-18-RECON-NGO/ Apps gmail LSHTM mail Mails Calendar LSHTM desk LSHTM printing ICIS Print Paperpile Facebook Other bookmarks

R Epidemics Consortium ABOUT RECON NEWS PEOPLE PROJECTS RESOURCES EVENTS FORUM

RECON 2.0: going official
RECON is now an NGO!
Posted on October 18, 2018

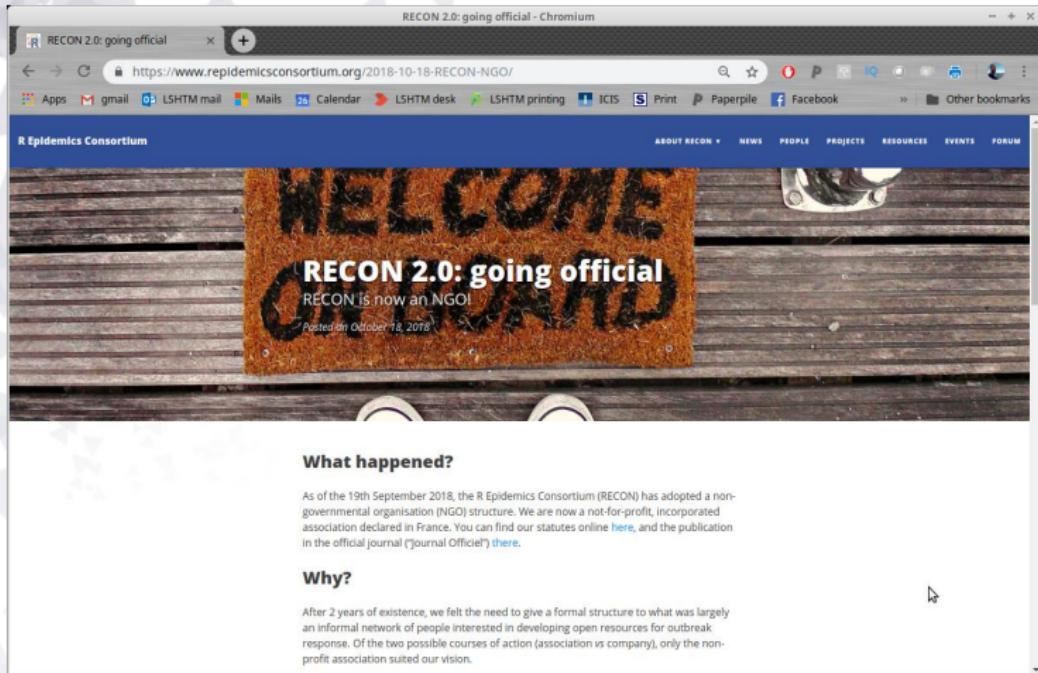
What happened?

As of the 19th September 2018, the R Epidemics Consortium (RECON) has adopted a non-governmental organisation (NGO) structure. We are now a not-for-profit, incorporated association declared in France. You can find our statutes online [here](#), and the publication in the official journal ("Journal Officiel") [there](#).

Why?

After 2 years of existence, we felt the need to give a formal structure to what was largely an informal network of people interested in developing open resources for outbreak response. Of the two possible courses of action (association vs company), only the non-profit association suited our vision.

RECON is now an NGO!



The screenshot shows a web browser window titled "RECON 2.0: going official - Chromium". The URL in the address bar is <https://www.repidemicsconsortium.org/2018-10-18-RECON-NGO/>. The page content is from the "R Epidemics Consortium" website. It features a large image of a doormat with the words "WELCOME" and "RECON 2.0: going official" on it. Below the image, the text reads "RECON is now an NGO!" and "Posted on October 18, 2018". The page has a blue header with links for "ABOUT RECON", "NEWS", "PEOPLE", "PROJECTS", "RESOURCES", "EVENTS", and "FORUM".

What happened?

As of the 19th September 2018, the R Epidemics Consortium (RECON) has adopted a non-governmental organisation (NGO) structure. We are now a not-for-profit, incorporated association declared in France. You can find our statutes online [here](#), and the publication in the official journal ("Journal Officiel") [there](#).

Why?

After 2 years of existence, we felt the need to give a formal structure to what was largely an informal network of people interested in developing open resources for outbreak response. Of the two possible courses of action (association vs company), only the non-profit association suited our vision.

Long-lasting consequence of our HPRU.

The RECON toolbox keeps growing

Released projects and packages

These projects are in a usable form. Packages have been developed following RECON's standards, are fully functional, documented and tested, and have been released on CRAN.

Icon	Name	Description
	discrete	Discretized probability distributions.
	earlyR	Estimation of infectiousness at the early stage of an outbreak
	epicontacts	Handling, visualisation and analysis of epidemiological contacts
	epitrix	Small helpers and tricks for epidemics analysis
	Incidence	Computation, handling, visualization and simple modeling of incidence
	outbreaker2	Modular framework for outbreak reconstruction
	outbreaks	Collection of outbreak data
	projections	Projections of future incidence
	RECON learn	Open training platform for epidemics analysis
	spiflow	Visualization and analysis of passenger flows

Up-and-coming packages

The RECON packages listed below are still in development. They may be functional already, but a stable version has yet to be released.

Icon	Name	Description
	branchr	R Estimation from Cluster Sizes
	clear	Rationalised and reproducible data cleaning
	drat	Install dev versions of RECON packages by default
	deployer	Portable environment for epidemics analysis
	dibbler	Investigation of food-borne disease outbreaks
	decker	Docker images for RECON packages
	epicontacts.ul	Graphical user interface for epicontacts
	epimaps	Helpers and wrappers for mapping diseases
	epimatch	Finding matching patient records across tabular data sets
	glitterstaid	Tutorials and code gists for mapping infectious diseases
	Incidence.ad	Graphical user interface for incidence
	listlist	Tools to Import and Tidy Case List Data

The RECON toolbox keeps growing

The image displays two side-by-side screenshots of the R Epidemics Consortium website, illustrating the growth of the RECON toolbox.

Released projects and packages: This section lists 13 projects, each represented by a circular icon with gears and a brief description:

- discrete**: Discretized probability distributions.
- earlyR**: Estimation of infectiousness at the early stage of an outbreak.
- epicontacts**: Handling, visualization and analysis of epidemiological contacts.
- epitrix**: Small helpers and tricks for epidemics analysis.
- Incidence**: Computation, handling, visualization and simple modeling of incidence.
- outbreaker2**: Modular framework for outbreak reconstruction.
- outbreaks**: Collection of outbreak data.
- projections**: Projections of future incidence.
- RECON learn**: Open training platform for epidemics analysis.
- spiflow**: Visualization and analysis of passenger flows.

Up-and-coming packages: This section lists 12 packages, each represented by a circular icon with gears and a brief description:

- branchr**: R estimation from Cluster Sizes.
- clear**: Rationalized and reproducible data cleaning.
- drat**: Install dev versions of RECON packages by default.
- deployer**: Portable environment for epidemics analysis.
- dibbler**: Investigation of food-borne disease outbreaks.
- decker**: Docker images for RECON packages.
- epicontacts.ul**: Graphic user interface for epicontacts.
- epimaps**: Helpers and wrappers for mapping diseases.
- epimatch**: Finding matching patient records across tabular data sets.
- glitterstd**: Tutorials and code gists for mapping infectious diseases.
- Incidence.ad**: Graphical user interface for incidence.
- listlist**: Tools to Import and Tidy Case Listwise Data.

What is the place of genomics tools?

Role of Whole Genome Sequences (WGS) in outbreak analytics: a controversial topic?

Nick Loman on Twitter: "Thank goodness I am on my way to #ESCAIDE to demonstrate in my plenary that genomics is the one and only true answer to all epidemiological questions."

Thibaut Jombart @TeelbzX @Virology_Bonn on detecting viruses early: surveillance and training, not genomics, is where efforts should be spent. Can't agree more. @ESCAIDE

11:33 AM - 21 Nov 2018
12 Retweets 74 Likes

Nick Loman Retweeted
ESCAIDE @ESCAIDE - Nov 21
"We can have all that rich information in one test - #wholegenomesequencing • 10
me its a no brainer"

@pathogenomicnick defends the superiority of genomics in answering our
epidemiological conundrums #ESCAIDE Plenary B

Role of Whole Genome Sequences (WGS) in outbreak analytics: a controversial topic?

Nick Loman on Twitter: "Thank goodness I am on my way to #ESCAIDE to demonstrate in my plenary that genomics is the one and only true answer to all epidemiological questions."

Thibaut Jombart @Teelbzr @Virology_Bonn on detecting viruses early: surveillance and training, not genomics, is where efforts should be spent. Can't agree more. #ESCAIDE

11:33 AM - 21 Nov 2018

12 Retweets 74 Likes

Nick Loman Retweeted
ESCAIDE @ESCAIDE - Nov 21
"We can have all that rich information in one test - #wholegenomesequencing • 10
me its a no brainer"

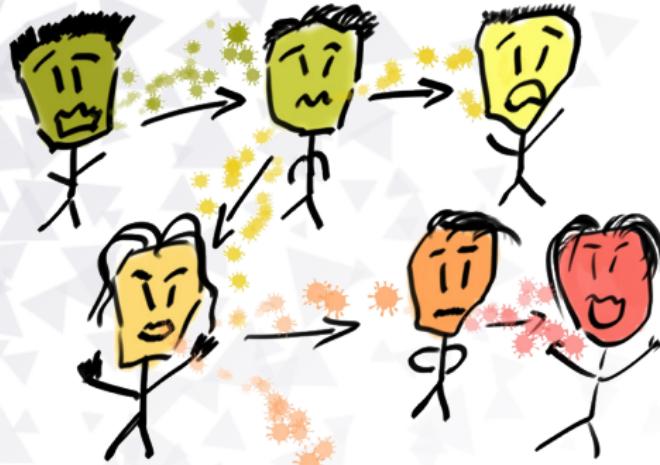
@pathogenomeric defends the superiority of #genomics in answering our
#epidemiological conundrums #ESCAIDE Plenary B

Nick Loman Retweeted
Karen Buttigieg @karen_buttigieg - Nov 21
"We throw away a lot of potentially useful host information in

What is the role of WGS in outbreak response?

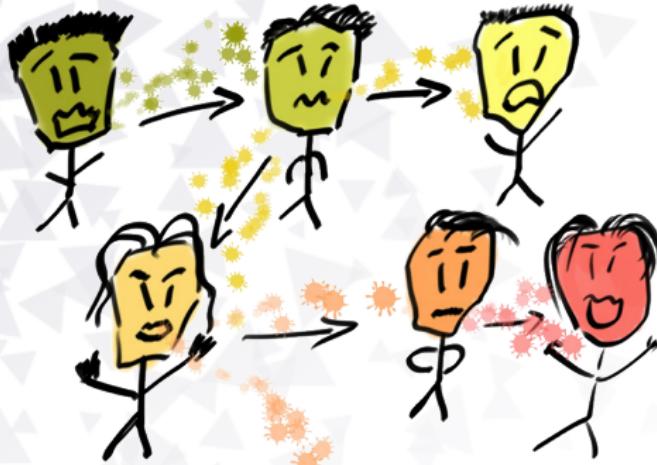
Reconstructing transmission trees

Using WGS to infer who infected whom



Mutations accumulate along transmission chains.

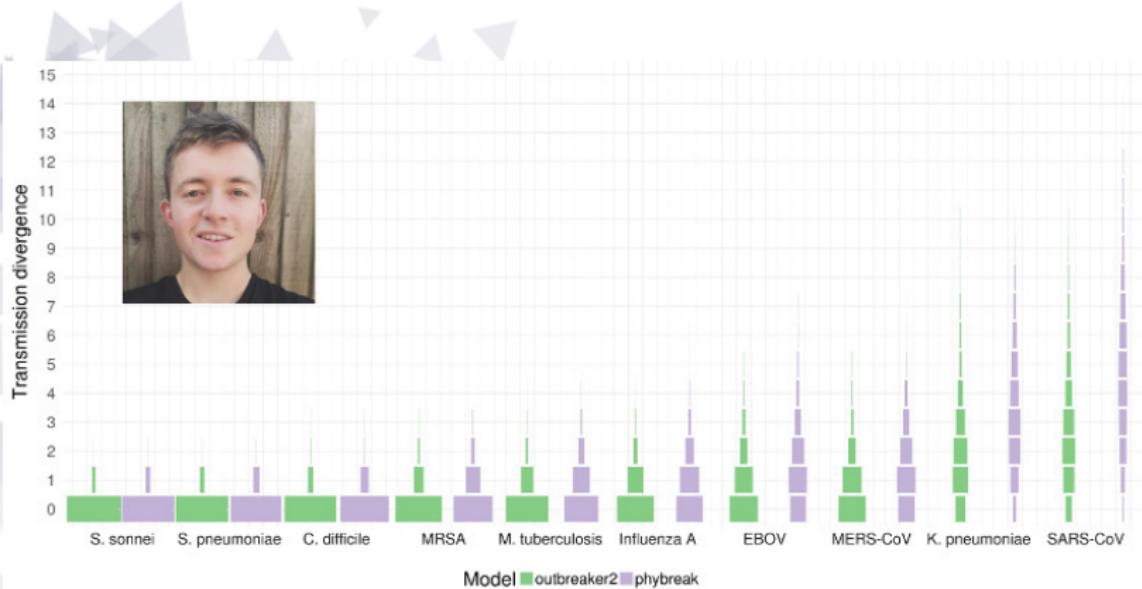
Using WGS to infer who infected whom



Mutations accumulate along transmission chains.

Can be used to reconstruct transmission trees.

How informative are whole genome sequences?



[Campbell *et al.* (2018) PLoS Pathogens]

Insufficient diversity for most diseases.

Evidence synthesis approach to outbreak reconstruction



Combine different data to shrink the set of plausible trees.

outbreaker2: evidence synthesis framework for outbreak reconstruction

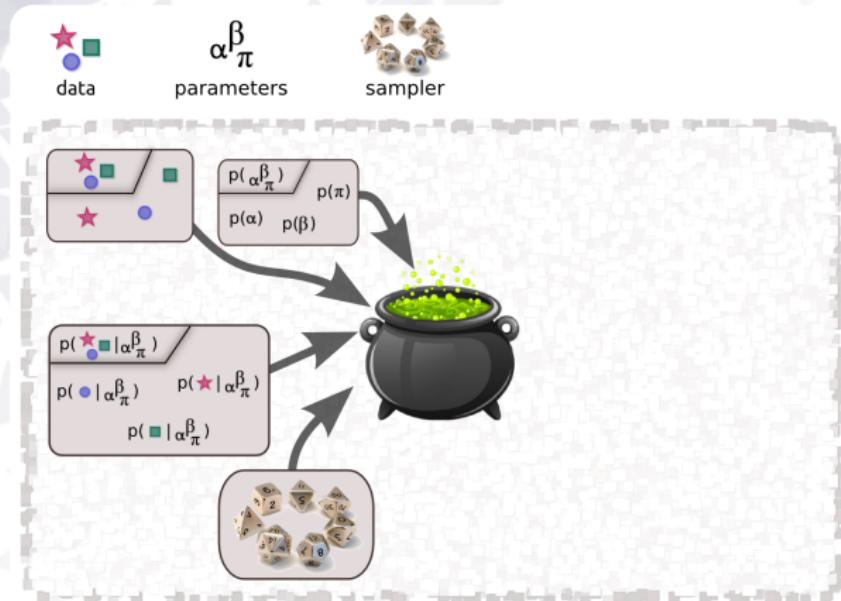
Modularity: customise data, prior, likelihood, MCMC.



[Campbell *et al.* (2018) BMC Bioinformatics]

outbreaker2: evidence synthesis framework for outbreak reconstruction

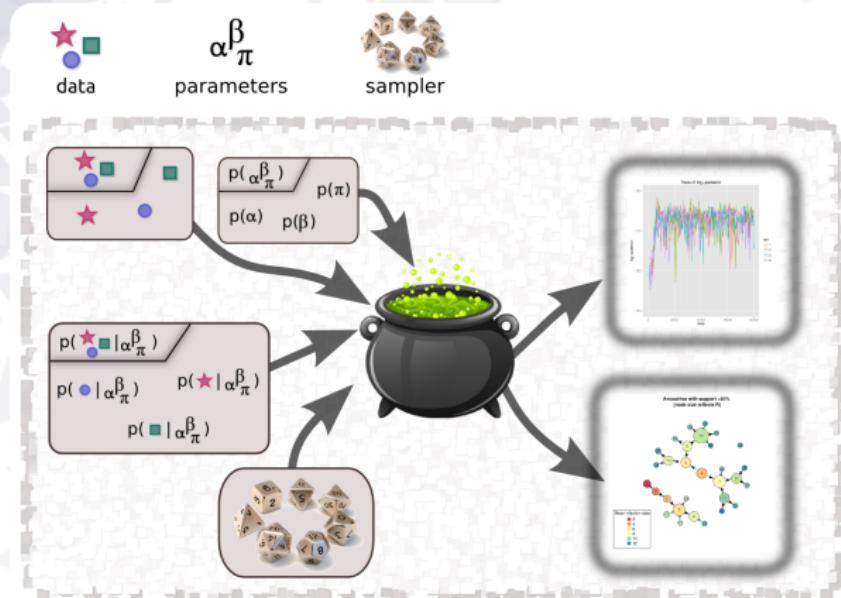
Modularity: customise data, prior, likelihood, MCMC.



[Campbell *et al.* (2018) BMC Bioinformatics]

outbreaker2: evidence synthesis framework for outbreak reconstruction

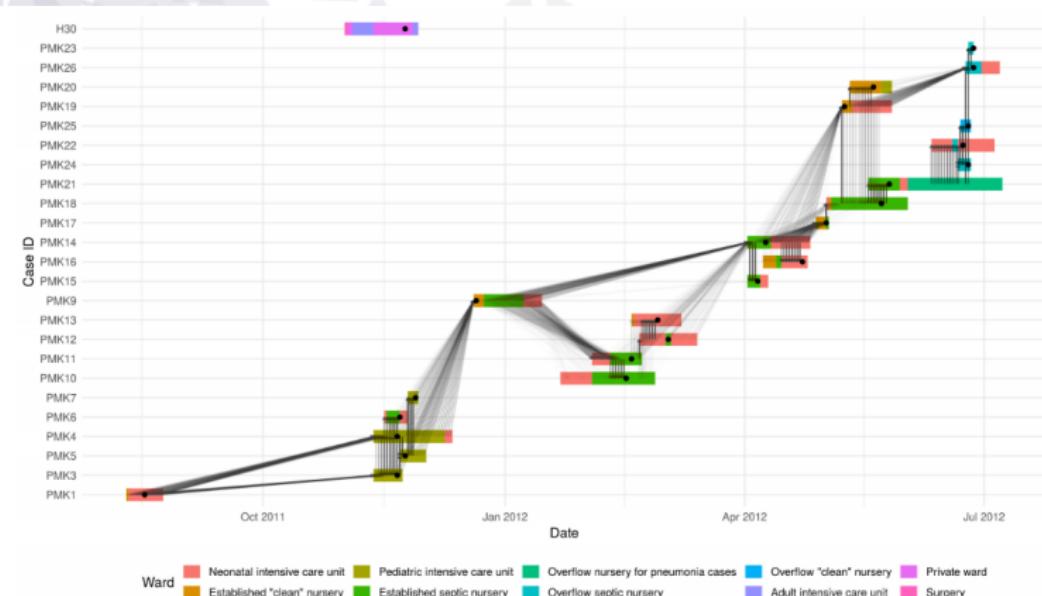
Modularity: customise data, prior, likelihood, MCMC.



[Campbell *et al.* (2018) BMC Bioinformatics]

New modules and ongoing work

Contact data, nosocomial transmission, haplotype model, spatial model, ...



[Campbell et al. (in prep)]

Looking ahead



- **common platform**: good opportunity for increased collaborations and comparisons of methods

Looking ahead



- **common platform**: good opportunity for increased collaborations and comparisons of methods
- **complex methods**: especially important to use continuous integration / extensive testing

Looking ahead



- **common platform**: good opportunity for increased collaborations and comparisons of methods
- **complex methods**: especially important to use continuous integration / extensive testing
- **within-host diversity**: what do we actually know about within-host evolution?

Looking ahead



- **common platform**: good opportunity for increased collaborations and comparisons of methods
- **complex methods**: especially important to use continuous integration / extensive testing
- **within-host diversity**: what do we actually know about within-host evolution?
- **transmission trees vs transmission clusters**

Reconstructing Ebola outbreaks: we are ready

Ebola simulation part 2: outbreak reconstruction

```
graph TD
    A["[1-26]"] --> B["[21-30]"]
    A --> C["[21-25]"]
    A --> D["[21-26]"]
    A --> E["[21-26]"]
    B --> F["[21-26]"]
    B --> G["[21-26]"]
    B --> H["[21-26]"]
    B --> I["[21-26]"]
    C --> J["[21-26]"]
    C --> K["[21-26]"]
    C --> L["[21-26]"]
    C --> M["[21-26]"]
    D --> N["[21-26]"]
    D --> O["[21-26]"]
    D --> P["[21-26]"]
    D --> Q["[21-26]"]
    E --> R["[21-26]"]
    E --> S["[21-26]"]
    E --> T["[21-26]"]
    F --> U["[21-26]"]
    F --> V["[21-26]"]
    F --> W["[21-26]"]
    G --> X["[21-26]"]
    G --> Y["[21-26]"]
    H --> Z["[21-26]"]
    I --> AA["[21-26]"]
    J --> BB["[21-26]"]
    K --> CC["[21-26]"]
    L --> DD["[21-26]"]
    M --> EE["[21-26]"]
    N --> FF["[21-26]"]
    O --> GG["[21-26]"]
    P --> HH["[21-26]"]
    Q --> II["[21-26]"]
    R --> JJ["[21-26]"]
    S --> KK["[21-26]"]
    T --> LL["[21-26]"]
    U --> MM["[21-26]"]
    V --> NN["[21-26]"]
    W --> OO["[21-26]"]
    X --> PP["[21-26]"]
    Y --> QQ["[21-26]"]
    Z --> RR["[21-26]"]
    AA --> SS["[21-26]"]
    BB --> TT["[21-26]"]
    CC --> UU["[21-26]"]
    DD --> VV["[21-26]"]
    EE --> WW["[21-26]"]
    FF --> XX["[21-26]"]
    GG --> YY["[21-26]"]
    HH --> ZZ["[21-26]"]
    II --> AA["[21-26]"]
    JJ --> BB["[21-26]"]
    KK --> CC["[21-26]"]
    LL --> DD["[21-26]"]
    MM --> EE["[21-26]"]
    NN --> FF["[21-26]"]
    OO --> GG["[21-26]"]
    PP --> HH["[21-26]"]
    QQ --> GG["[21-26]"]
    RR --> HH["[21-26]"]
    SS --> BB["[21-26]"]
    TT --> CC["[21-26]"]
    UU --> DD["[21-26]"]
    VV --> EE["[21-26]"]
    WW --> FF["[21-26]"]
    XX --> GG["[21-26]"]
    YY --> HH["[21-26]"]
    ZZ --> HH["[21-26]"]
    AA --> BB["[21-26]"]
    BB --> CC["[21-26]"]
    CC --> DD["[21-26]"]
    DD --> EE["[21-26]"]
    EE --> FF["[21-26]"]
    FF --> GG["[21-26]"]
    GG --> HH["[21-26]"]
    HH --> HH["[21-26]"]
    
```

Looking carefully at the documentation of `viz.genealogies`, try to reproduce the final consensus tree below:

Ebola simulation part 2: outbreak reconstruction

Ebola
A collection of 4 posts

Ebola simulation part 1: early outbreak assessment

Introduction: understanding Ebola outbreaks

This practical simulates the early assessment of an Ebola outbreak in a city of 100,000 inhabitants. It introduces various aspects of analysis at the early stage of an outbreak, including contact tracing and specimen collection, global network estimation, and how these can be used to estimate the basic reproduction number (R_0). Following on from early assessments of transmissibility, this part explores more methodological options for estimating transmission potential and provides an introduction to outbreak reconstruction using epiNowcast 2.

An Ebola outbreak has been notified in the small city of Aïtou, Republic of Mopanezia. A new Ebola audience has been notified in the small city of Aïtou, Republic of Mopanezia. A new Ebola audience has been notified in the small city of Aïtou, Republic of Mopanezia. Public Health officials have been sent to update the health and contact network.

Ebola simulation part 2: outbreak reconstruction

Introduction: understanding Ebola outbreaks

This practical is the second step of the response to a simulated Ebola Virus Disease (EVD) outbreak taking place in the city of Aïtou, Republic of Mopanezia. Building on the work done in the first part of this practical, this part explores more methodological options for estimating transmission potential and provides an introduction to outbreak reconstruction using epiNowcast 2. An Ebola audience has been notified in the small city of Aïtou, Republic of Mopanezia. A new Ebola audience has been notified in the small city of Aïtou, Republic of Mopanezia. Public Health officials have been sent to update the health and contact network.

Compartmental model example for Ebola

This short exercise gives a primer on fitting compartmental models. It covers various aspects around epidemiological modelling such as how to estimate the basic reproduction number and reproduction number. The example of the dynamics of Ebola is used. Slides Click on the image link to access them. About this post: [Ebola: building few models for Ebola](#)

THOMAS CHAMBERS
PHILIP CAMPBELL
[practicals@polaris.ac](#)

THOMAS CHAMBERS
PHILIP CAMPBELL
[practicals@polaris.ac](#)

PIERRE NOUVELLET
[p.nouvellet@polaris.ac](#)

Ebola: building few models for Ebola

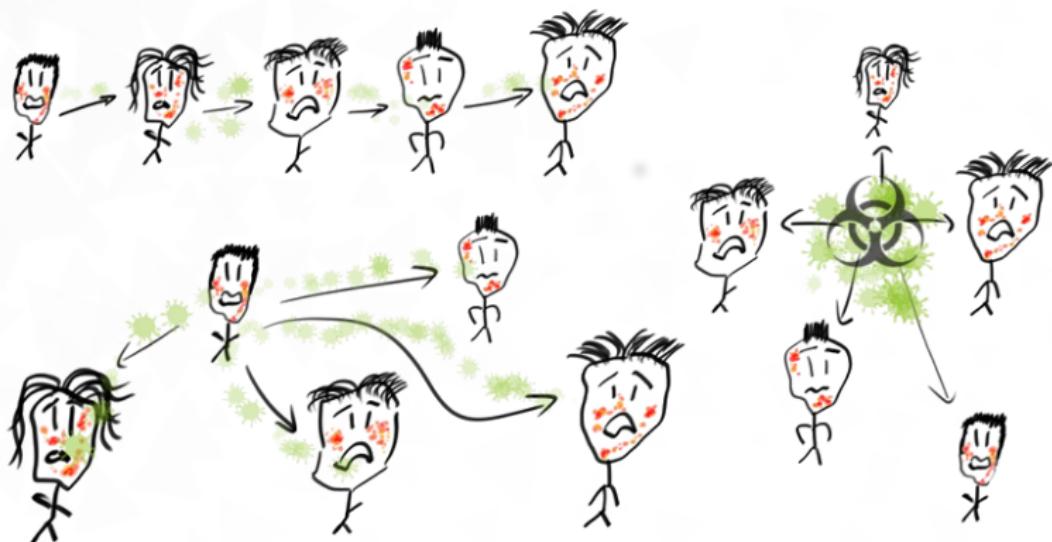
This practical aims to illustrate the basics of Ebola modelling using R, with an emphasis on how the methods work. We will start with a basic model for Ebola. We will then illustrate

Source: <https://reconlearn.org>

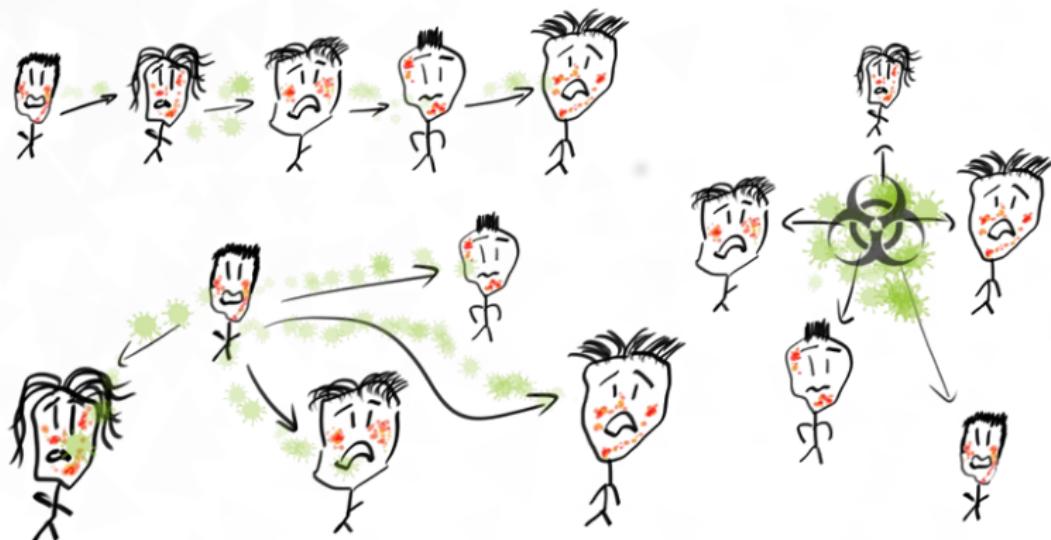
The background of the slide features a large number of small, light-gray triangles of various sizes scattered across the white space, creating a subtle geometric pattern.

Identifying transmission clusters

Sustained transmission vs repeated introductions



Sustained transmission vs repeated introductions

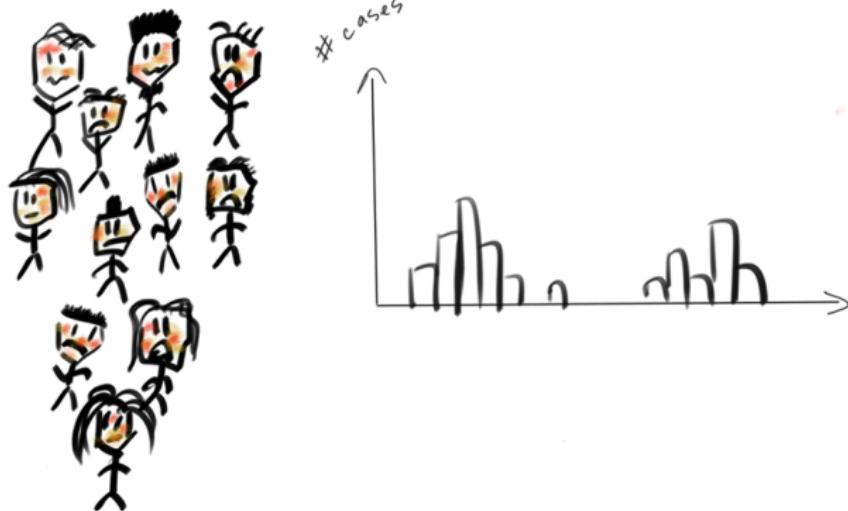


Different transmission patterns call for different interventions

One outbreak.. how many introductions?



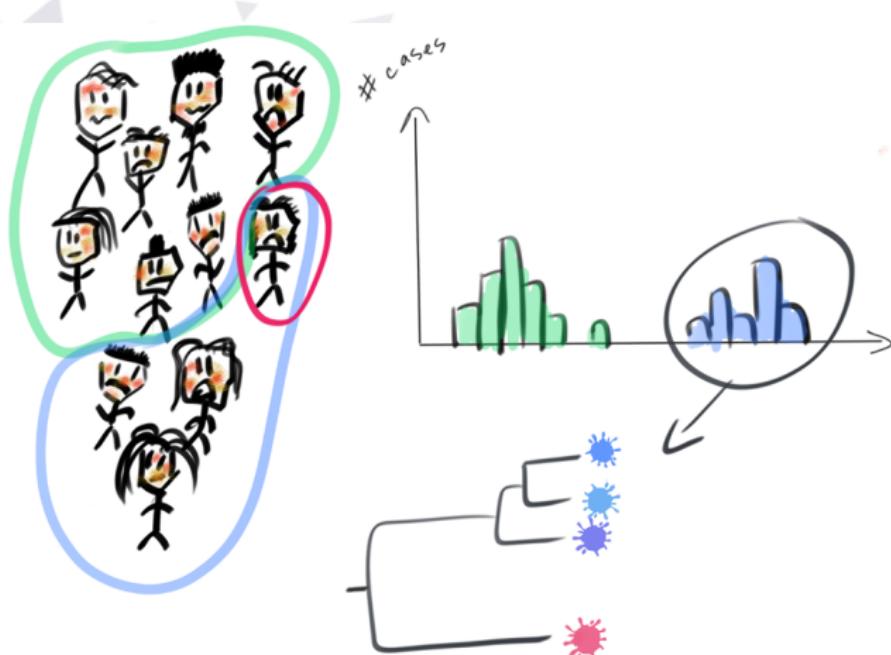
One outbreak.. how many introductions?



One outbreak.. how many introductions?



One outbreak.. how many introductions?



Combined data sources can detect **outbreak clusters**

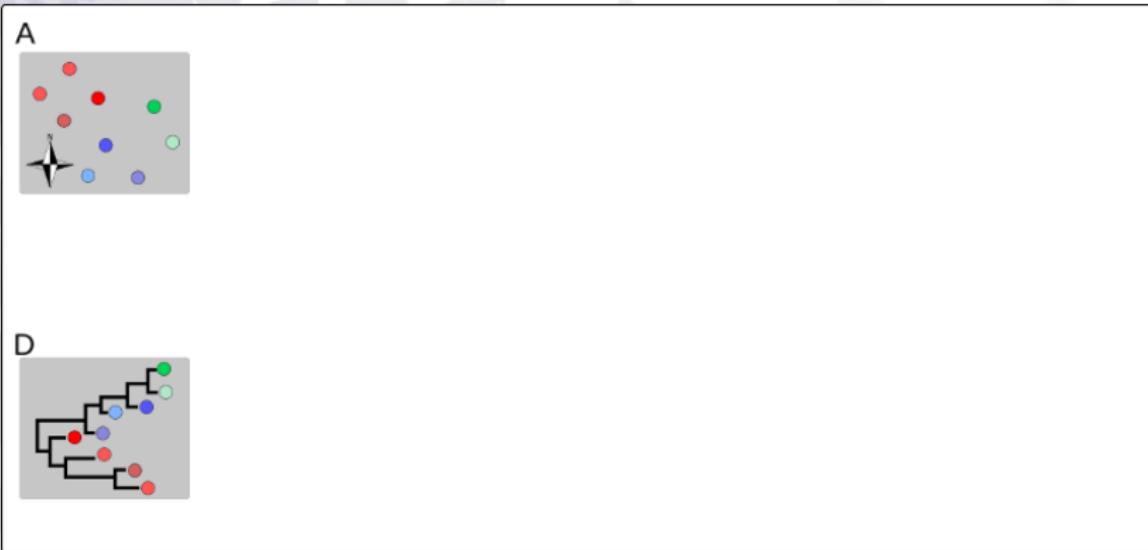


Aims: develop a new method which...

- **detects outbreak clusters:** cases stemming from same introduction (same transmission tree)
- **integrates different data:** temporal, spatial, genetic, etc.
- **works fast, scales well:** so that it can be used for real-time outbreak detection

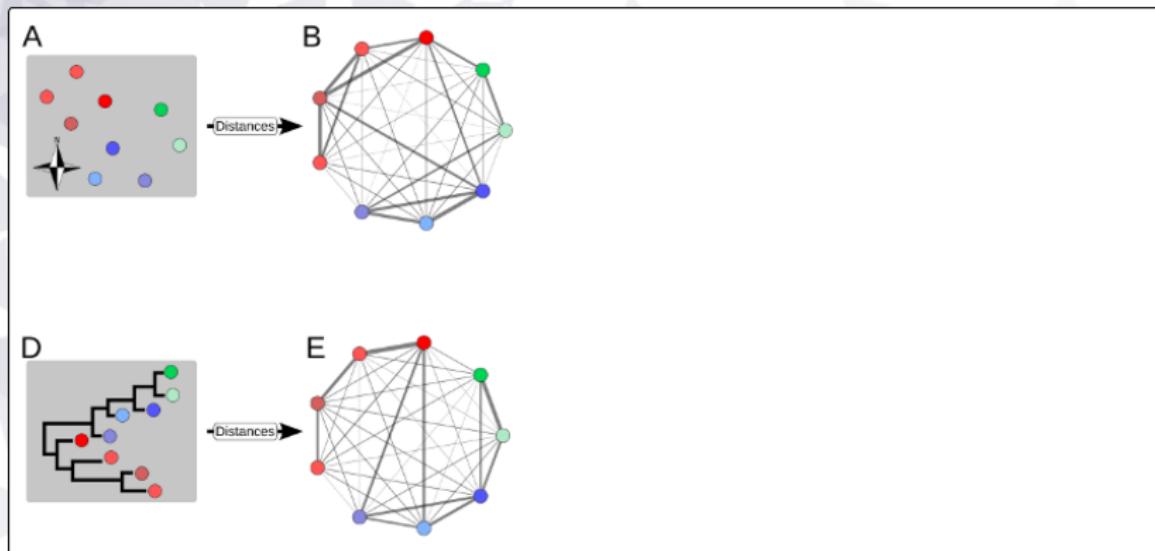
¹well, really, I made that up because I was reading 'Snuff' by Terry Pratchett at the time; incidentally, Pratchett was a huge fan of using long footnotes in his novels, which are often quite entertaining to read; well, this does not apply here: if you are still reading this, you probably missed what I just said

A graph-based evidence synthesis approach



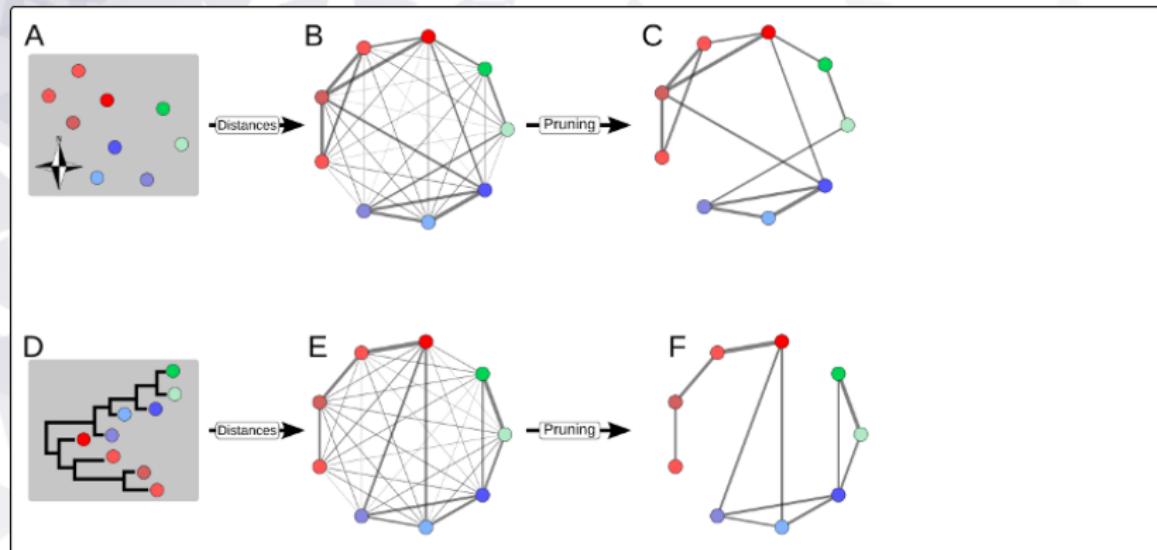
[Cori et al. (2018) PLoS Computational Biology]

A graph-based evidence synthesis approach



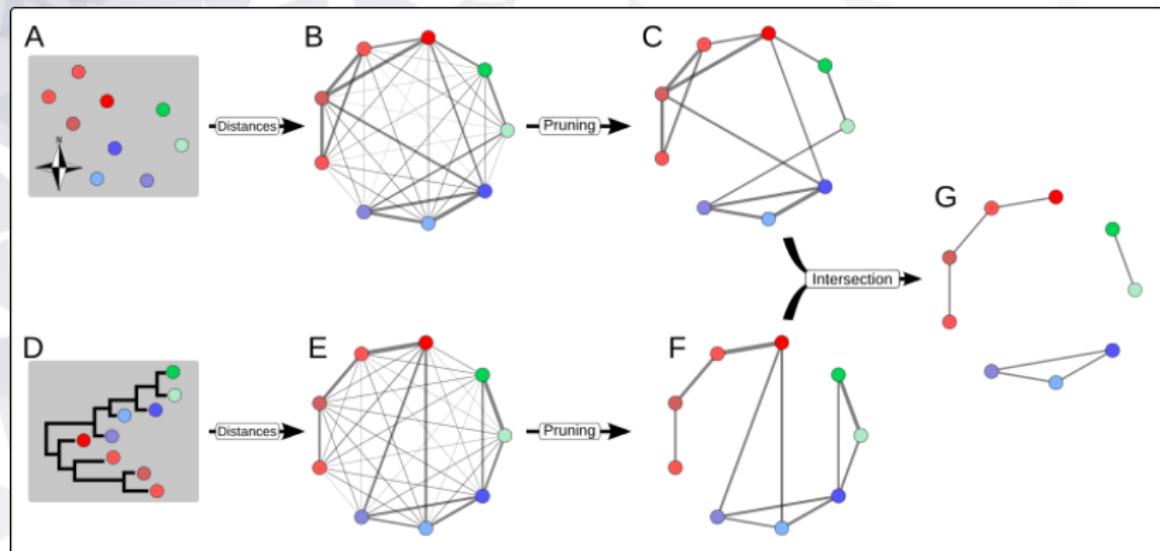
[Cori et al. (2018) PLoS Computational Biology]

A graph-based evidence synthesis approach



[Cori et al. (2018) PLoS Computational Biology]

A graph-based evidence synthesis approach



[Cori et al. (2018) PLoS Computational Biology]

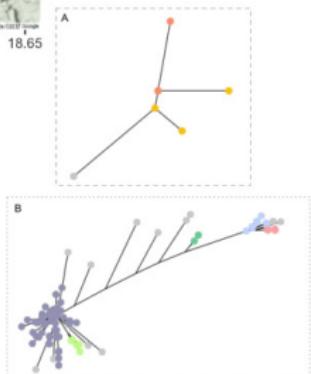
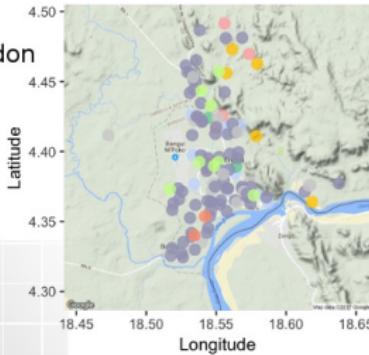
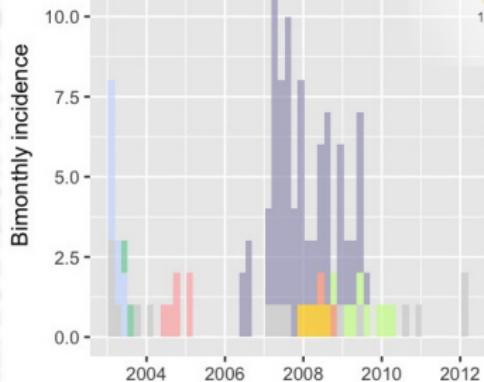
Application: dog rabies epidemics, Central African Republic



Anne Cori
Imperial College London



Pierre Nouvellet
University of Sussex



[Cori et al. (2018) PLoS Computational Biology]

Looking ahead



- **genomic surveillance**: integration in existing pipelines at PHE and Public Health Wales

Looking ahead



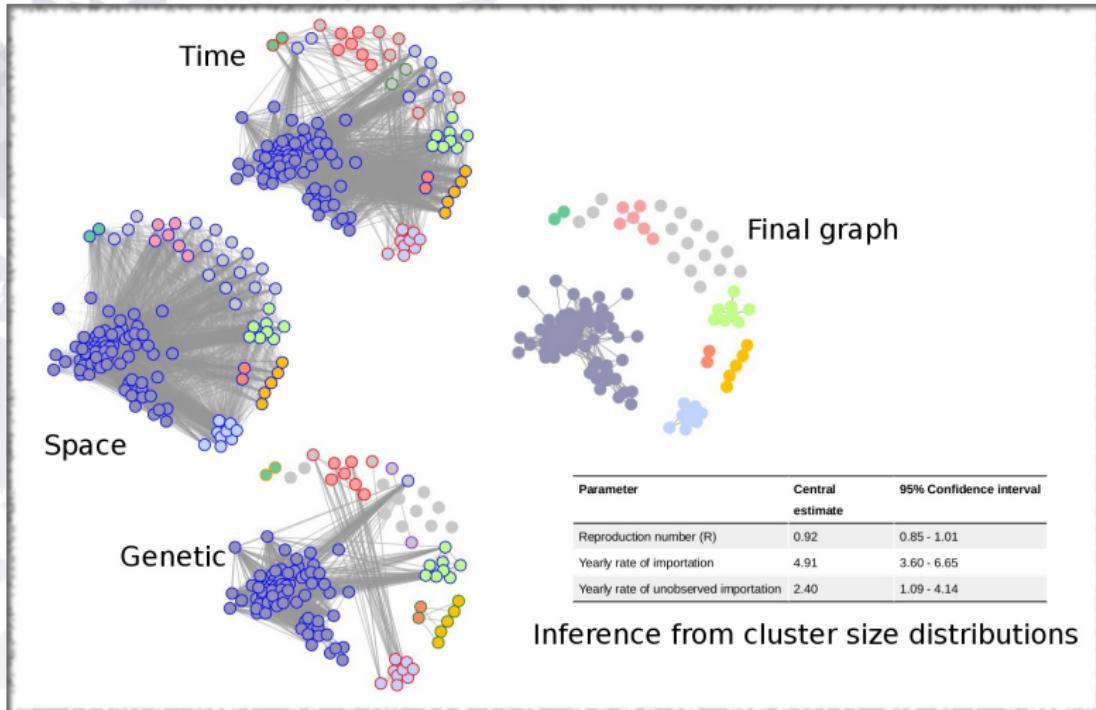
- **genomic surveillance**: integration in existing pipelines at PHE and Public Health Wales
- **refinements**: optimisation of the pruning method, improve algorithm complexity

Looking ahead

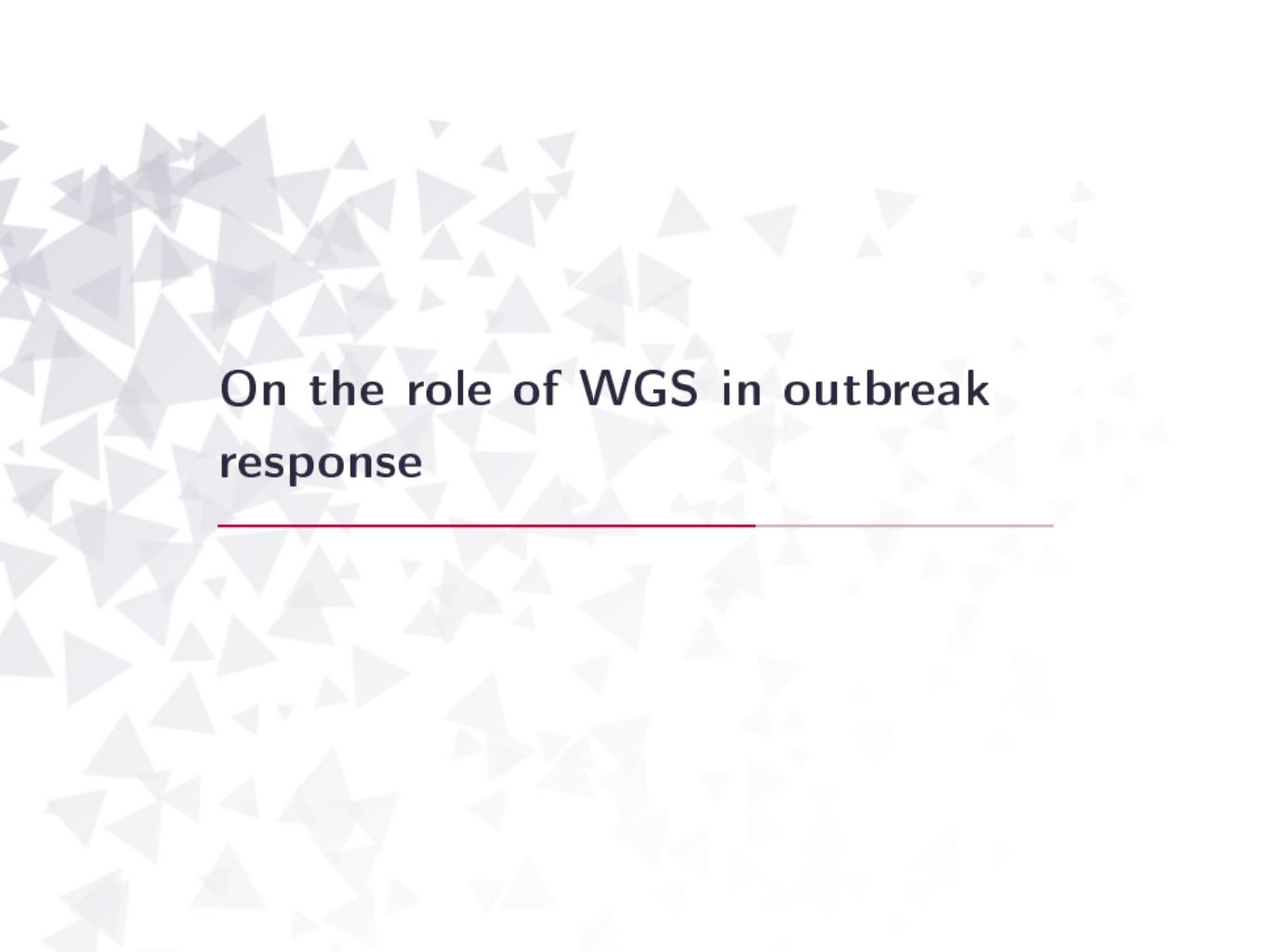


- **genomic surveillance**: integration in existing pipelines at PHE and Public Health Wales
- **refinements**: optimisation of the pruning method, improve algorithm complexity
- **further developments**: alternative methods using probabilistic framework

Results



[Cori et al. (2018) PLoS Computational Biology]



On the role of WGS in outbreak response

Pathogen genetics in outbreak response



- in general, not needed for forecasting or control

Pathogen genetics in outbreak response



- in general, **not needed** for forecasting or control
- useful to detect **multiple introductions** or **superspreading**

Pathogen genetics in outbreak response



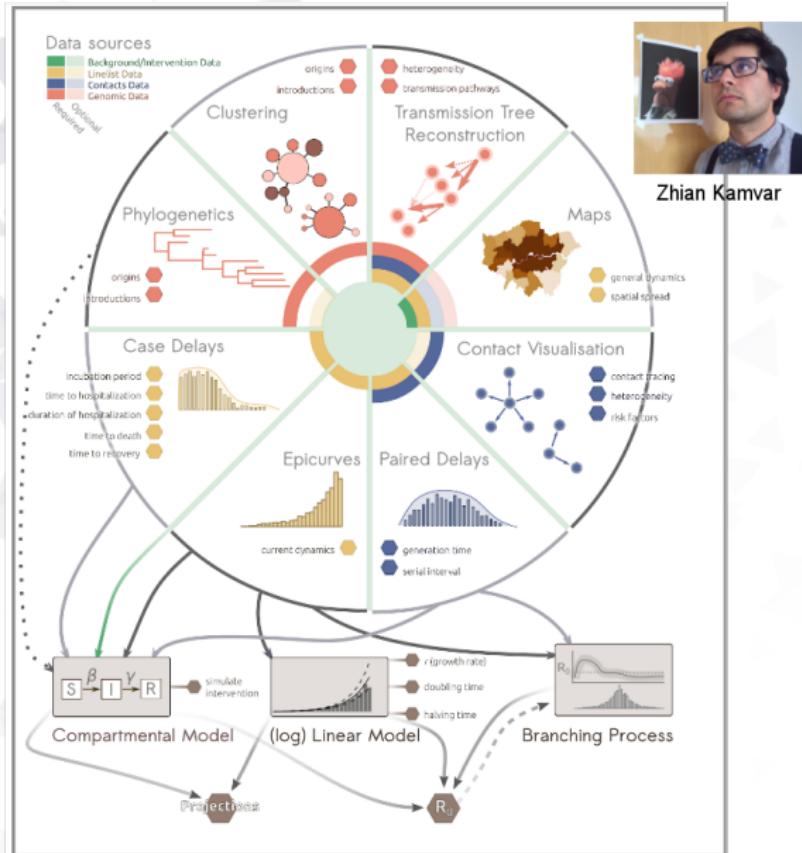
- in general, not needed for forecasting or control
- useful to detect multiple introductions or superspreading
- can complement contact tracing data

Pathogen genetics in outbreak response



- in general, not needed for forecasting or control
- useful to detect multiple introductions or superspreading
- can complement contact tracing data
- WGS are costly: is it worth it?

The outbreak analytics toolbox

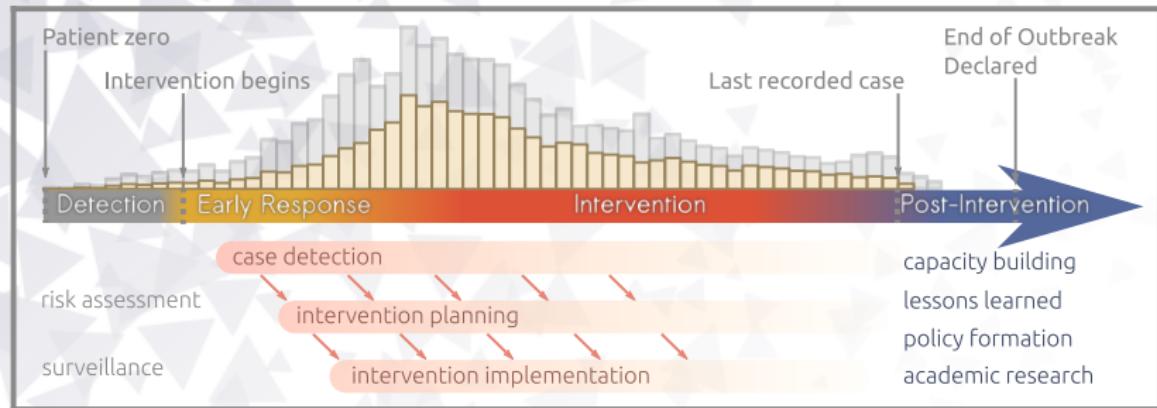


Thanks to:

- **Claire Thomson**
- **Collaborators:** Finlay Campbell, Anne Cori, Pierre Nouvellet, Zhian Kamvar, Amrish Baidjoe, Roz Eggo, Tini Garske, Hervé Bourhy, Emmanuel Nakouné, Jimmy Whitworth, Dan Bausch, Bayard Roberts, Neil Ferguson
- **Groups:** R Epidemics Consortium
- **Funding:** HPRU-NIHR, GCRF project RECAP (ES/P010873/1), UK PH RST

RECON

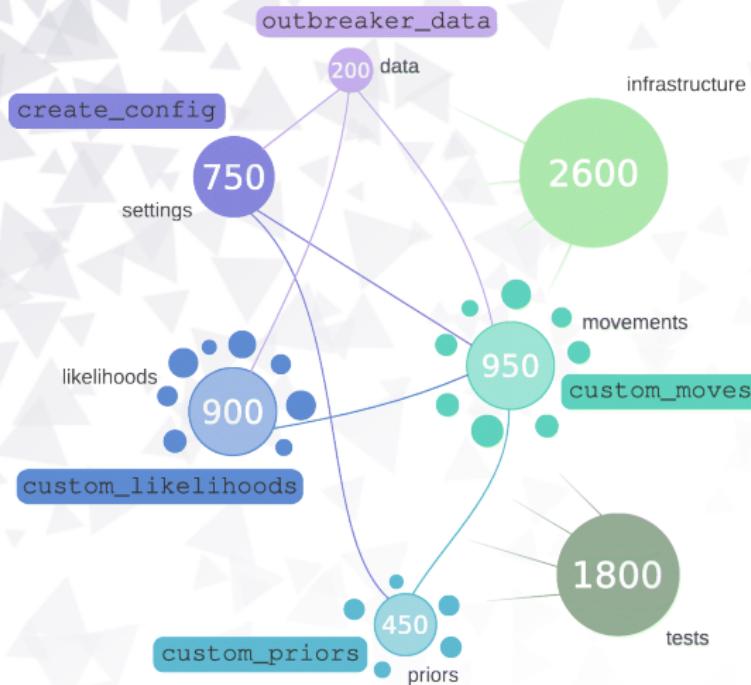
Timeline of an outbreak response



[Baidjoe et al. (minor revision) Phil Trans Roy Soc B]

What is inside the pot?

Module sizes in lines of code, and entry points:



Example: implementing *TransPhylo* in *outbreaker2*

outbreaker likelihood

- $p(s, t | \alpha, T^{inf}, \kappa, \mu, \pi) = p(T^{inf} | \alpha, \kappa) p(T^{inf} | t) p(s | \alpha, \kappa, \mu) p(\kappa | \pi)$
- i.e. *timing infection* \times *incubation* \times *genetic (simple)* \times *missing cases*

Example: implementing *TransPhylo* in *outbreaker2*

outbreaker likelihood

- $p(s, t | \alpha, T^{inf}, \kappa, \mu, \pi) = p(T^{inf} | \alpha, \kappa) p(T^{inf} | t) p(s | \alpha, \kappa, \mu) p(\kappa | \pi)$
- i.e. *timing infection* \times *incubation* \times *genetic (simple)* \times *missing cases*

TransPhylo likelihood

- $p(G | \beta, \gamma, N_{eg}, \alpha) = p(G | N_{eg}, \alpha) \times p(\alpha | \beta, \gamma)$
- i.e. *phylogeny (coalescent)* \times *SIR*

Example: implementing *TransPhylo* in *outbreaker2*

outbreaker likelihood

- $p(s, t | \alpha, T^{inf}, \kappa, \mu, \pi) = p(T^{inf} | \alpha, \kappa) p(T^{inf} | t) p(s | \alpha, \kappa, \mu) p(\kappa | \pi)$
- i.e. *timing infection* \times *incubation* \times *genetic (simple)* \times *missing cases*

TransPhylo likelihood

- $p(G | \beta, \gamma, Neg, \alpha) = p(G | Neg, \alpha) \times p(\alpha | \beta, \gamma)$
- i.e. *phylogeny (coalescent)* \times *SIR*

Can we combine the two models?

TransPhylo module for *outbreaker2*

3 Custom likelihood

In order to calculate the likelihood under the TransPhylo model, we need to (i) extract the transmission tree from the outbreaker2 parameter, (ii) combine this transmission tree with the phylogenetic tree to form a colored tree, and (iii) calculate the likelihood of this colored tree. Step (i) is easy since transmission tree are encoded almost in the same way in TransPhylo and outbreaker2. For step (ii) we have to write the `combine` function which is tedious but not especially interesting (this function is included in this Rnw file but its code is not shown in the pdf). For step (iii) we only need to call the appropriate function of the TransPhylo package which is `probPTreeGivenTree`. During step (ii) messages can arise indicating that the transmission tree and phylogenetic tree are in fact incompatible, in which case the likelihood is returned as -Inf.

```
lik_TransPhylo <- function(data, param) {
  ttree <- list(ttree = cbind(param$t.inf, data$dates, param$alpha),
    nam = data$ptree$nam)
  ttree$tree[which(is.na(ttree$ptree[,3])),3] <- 0
  txt <- capture.output(ctree <- combine(ttree,data$ptree))
  if (length(txt)>0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
  } else {
    prob <- -Inf
  }
  return(prob)
}

## Function to calculate the likelihood
## Function (date, param, i = NULL, custom_functions = NULL)
## NULL

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:date$N) {
    current_ll <- api$pp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(1:100),1,100,1)
    pos_ll <- sample(1:(date$N-1),1,100,1)
    new_ll <- api$pp_ll_all(data,param, i = NULL, list_custom_ll)
    if (log(modif[i]) > (new_ll - current_ll)) {
      param$ll_inf[[i]] <- param$ll_inf[[i]] + modif
    }
  }
  return(param)
}

new_moves <- custom_moves(t.inf = new_move_tinf)
new_moves

## 
## ////////////// outbreaker movement functions ///
## 
## class: outbreaker_moves list
## number of items: 8
## 
## // movement functions //
## $eu
```

[Campbell *et al.* (2018) BMC Bioinformatics]

TransPhylo module for *outbreaker2*

3 Custom likelihood

In order to calculate the likelihood under the TransPhylo model, we need to (i) extract the transmission tree from the outbreaker2 parameter, (ii) combine this transmission tree with the phylogenetic tree to form a colored tree, and (iii) calculate the likelihood of this colored tree. Step (i) is easy since transmission tree are encoded almost in the same way in TransPhylo and outbreaker2. For step (ii) we have to write the *combine* function which is tedious but not especially interesting (this function is included in this Rnw file but its code is not shown in the pdf). For step (iii) we only need to call the appropriate function of the TransPhylo package which is *probPTreeGivenTree*. During step (ii) messages can arise indicating that the transmission tree and phylogenetic tree are in fact incompatible, in which case the likelihood is returned as -Inf.

```
lik_TransPhylo <- function(data, param) {
  ttree <- list(ttree = cbind(param$t.inf, data$dates, param$alpha),
    nam = data$ptree$nam)
  ttree$tree[which(is.na(ttree$ptree[,3]))] <- 0
  txt <- capture.output(ctree <- combine(ttree, data$ptree))
  if (length(txt)>0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
  } else {
    prob <- -Inf
  }
  return(prob)
}

## function (date, param, i = NULL, custom_functions = NULL)
## NULL

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:date$N) {
    current_ll <- api$pp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(1:100),1:100, 1)
    pos_inf <- inf.list[[i]][modif]
    new_ll <- api$pp_ll_all(data,param, i = NULL, list_custom_ll)
    if (log(modif[i]) > (new_ll - current_ll)) {
      param$ll_inf[[i]] <- param$ll_inf[[i]] + modif
    }
  }
  return(param)
}

new_moves <- custom_moves(t.inf = new_move_tinf)
new_moves

## 
## //////////////////////////////////////////////////////////////////
## class: outbreaker_moves list
## number of items: 8
## 
## // movement functions //
## $eu
```

likelihood

[Campbell *et al.* (2018) BMC Bioinformatics]

TransPhylo module for *outbreaker2*

3 Custom likelihood

In order to calculate the likelihood under the TransPhylo model, we need to (i) extract the transmission tree from the outbreaker2 parameter, (ii) combine this transmission tree with the phylogenetic tree to form a colored tree, and (iii) calculate the likelihood of this colored tree. Step (i) is easy since transmission tree are encoded almost in the same way in TransPhylo and outbreaker2. For step (ii) we have to write the *combine* function which is tedious but not especially interesting (this function is included in this Rnw file but its code is not shown in the pdf). For step (iii) we only need to call the appropriate function of the TransPhylo package which is *probPTreeGivenTree*. During step (ii) messages can arise indicating that the transmission tree and phylogenetic tree are in fact incompatible, in which case the likelihood is returned as -Inf.

```
lik_TransPhylo <- function(data, param) {
  ttree <- list(ttree = cbind(param$t.inf, data$dates, param$alpha),
    nam = data$tree$nam)
  ttree$tree[which(is.na(ttree$tree[,3]))] <- 0
  txt <- capture.output(ctree <- combine(ttree, data$ptree))
  if (length(txt)>0) {
    prob <- probPTreeGivenTree(ctree, neg = 365 * 0.25)
  } else {
    prob <- -Inf
  }
  return(prob)
}
```

likelihood

movement function

```
args(api$cpp_ll_all)
## function (date, param, i = NULL, custom_functions = NULL)
## NULL

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:date$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(1:100),1,100,1)
    param$inf$inf[modif] <- param$inf$inf[modif] + 1
    new_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    if (logif(modif[i]) > (new_ll - current_ll)) {
      param$inf$inf[i] <- param$inf$inf[i] + modif
    }
  }
  return(param)
}

new_moves <- custom_moves(t.inf = new_move_tinf)
new_moves

##
## //////////////////////////////////////////////////////////////////
## class: outbreaker_moves list
## number of items: 8
## 
## // movement functions //
## $eu
```

[Campbell *et al.* (2018) BMC Bioinformatics]

TransPhylo module for *outbreaker2*

3 Custom likelihood

In order to calculate the likelihood under the TransPhylo model, we need to (i) extract the transmission tree from the outbreaker2 parameter, (ii) combine this transmission tree with the phylogenetic tree to form a colored tree, and (iii) calculate the likelihood of this colored tree. Step (i) is easy since transmission tree are encoded almost in the same way in TransPhylo and outbreaker2. For step (ii) we have to write the *combine* function which is tedious but not especially interesting (this function is included in this Rnw file but its code is not shown in the pdf). For step (iii) we only need to call the appropriate function of the TransPhylo package which is *probPTreeGivenTree*. During step (ii) messages can arise indicating that the transmission tree and phylogenetic tree are in fact incompatible, in which case the likelihood is returned as -Inf.

```
lik_TransPhylo <- function(data, param) {
  ttree <- list(ttree = cbind(param$t.inf, data$dates, param$alpha),
    nam = data$ptree$nam)
  ttree$tree[which(is.na(ttree$ptree[,3]))] <- 0
  txt <- capture.output(ctree <- combine(ttree, data$ptree))
  if (length(txt)>0) {
    prob <- probPTreeGivenTree(ctree, neg = 365 * 0.25)
  } else {
    prob <- -Inf
  }
  return(prob)
}
```

likelihood

Total: 25 lines of R

outbreaker2: 7,500 lines of R/C++

Code difference: 0.3%

movement function

```
args(api$cpp_ll_all)
## function (date, param, i = NULL, custom_functions = NULL)
## NULL

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:date$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(1:100),1,100,1)
    param$inf$inf$inf[[modif]] <- current_ll
    new_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    if (is.na(modif)) {
      new_ll[[i]] <- (new_ll[[i]] + current_ll[[i]])
    } else {
      param$inf$inf[[modif]] <- param$inf$inf[[modif]] + new_ll[[i]]
    }
  }
  return(param)
}

new_moves <- custom_moves(t.inf = new_move_tinf)
new_moves

## 
## //////////////////////////////////////////////////////////////////
## class: outbreaker_moves list
## number of items: 8
## 
## // movement functions //
## $na
```

[Campbell *et al.* (2018) BMC Bioinformatics]

TransPhylo module for *outbreaker2*

3 Custom likelihood

In order to calculate the likelihood under the TransPhylo model, we need to (i) extract the transmission tree from the *outbreaker2* parameter, (ii) combine this transmission tree with the phylogenetic tree to form a new tree, (iii) calculate the likelihood of this colored tree. Step (i) is easy since transmission tree are encoded as a list in *outbreaker2*. Step (ii) requires us to write some code to combine the two trees. This is where the new stuff in *TransPhylo* comes in. Step (iii) is also easy since the likelihood function in *outbreaker2* only need to call the appropriate function in *TransPhylo*. In practice, it is possible to get an Inf value as a result of step (iii). (iii) messages can arise indicating that the transmission tree and phylogenetic tree are in fact incompatible. In this case the likelihood is returned as -Inf.

```
lik_TransPhylo <- function(data, param) {
  ttree <- list(tree = chisq(param$`-inf`, data$data, param$alpha),
    max = data$tree$max)
  ttree$tree[which(is.na(ttree$tree[,3]))] <- 0
  txt <- capture.output(ctree <- combine(ttree, data$ptree))
  if (length(txt) > 0) {
    prob <- as.numeric(strsplit(txt, "\n")[[1]][25])
  } else {
    prob <- 1
  }
  return(prob)
}
```

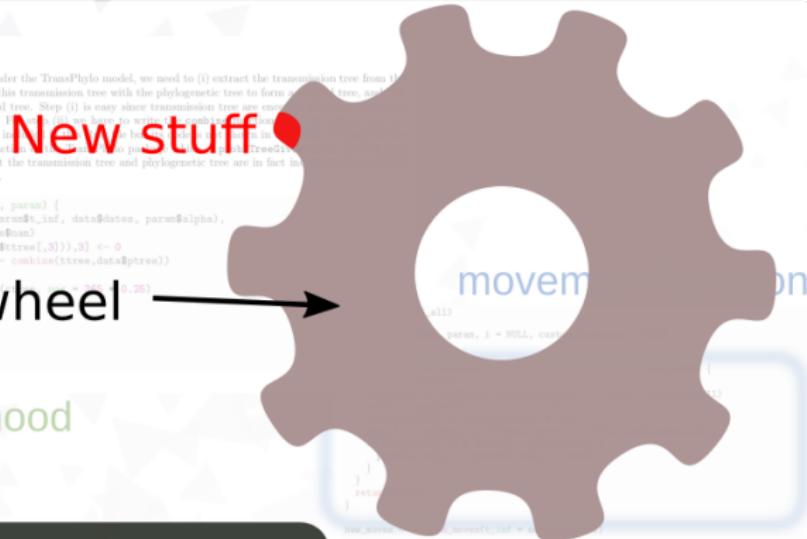
likelihood

Total: 25 lines of R

outbreaker2: 7,500 lines of R/C++

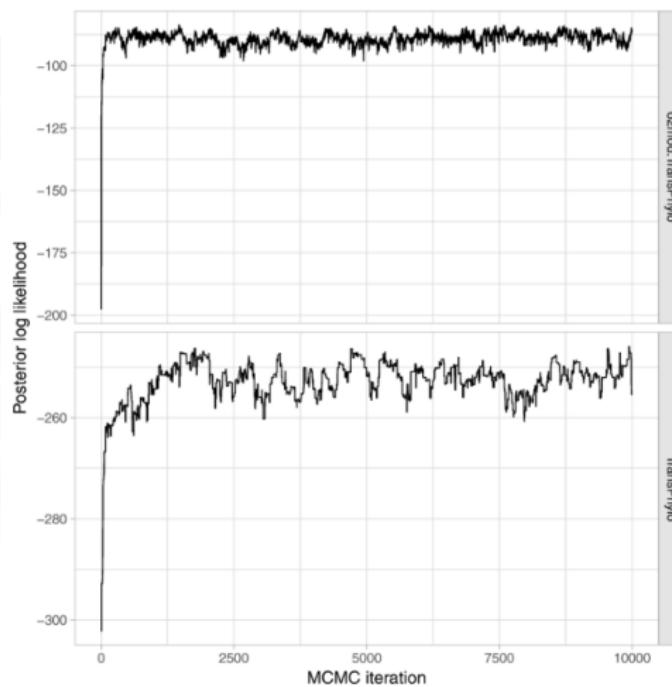
Code difference: 0.3%

[Campbell *et al.* (2018) BMC Bioinformatics]



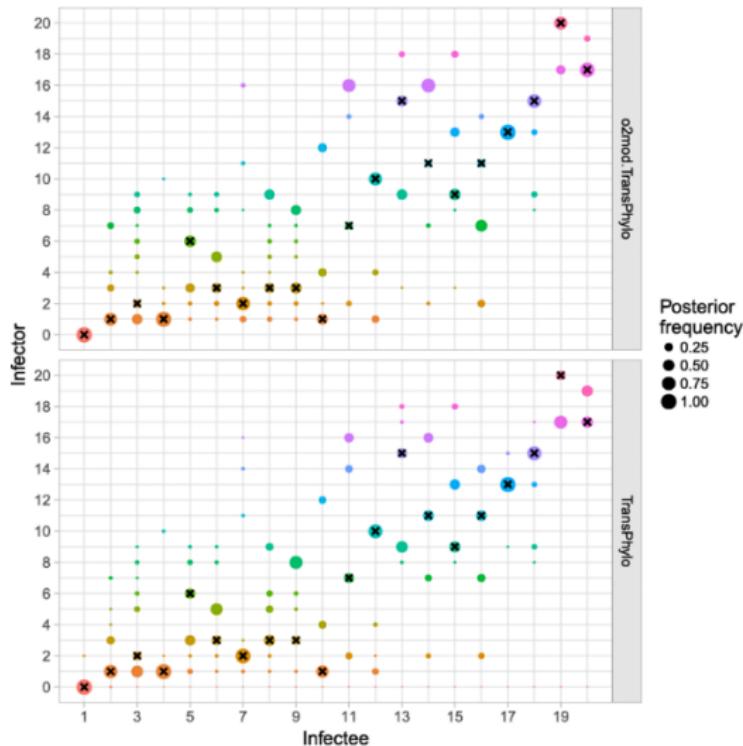
```
##  
## ////////////// outbreaker movement functions ///////////////  
##  
## class: outbreaker_movement_list  
## number of items: 8  
##  
## ////////////// movement functions ///////////////  
##  
##
```

Transphylo module results (1/2): convergence



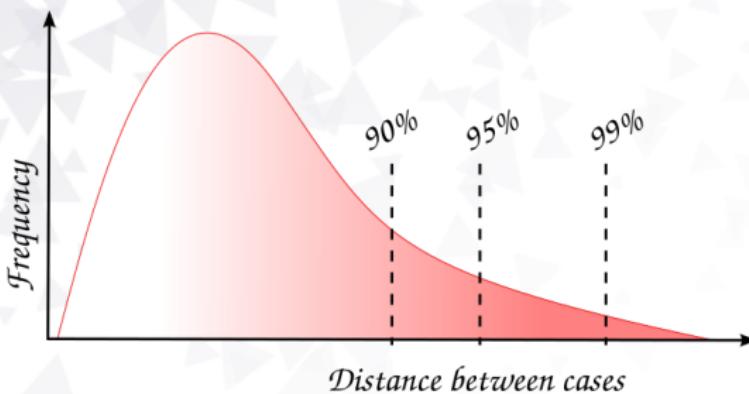
[Campbell *et al.* (2018) BMC Bioinformatics]

Transphylo module results (2/2): trees

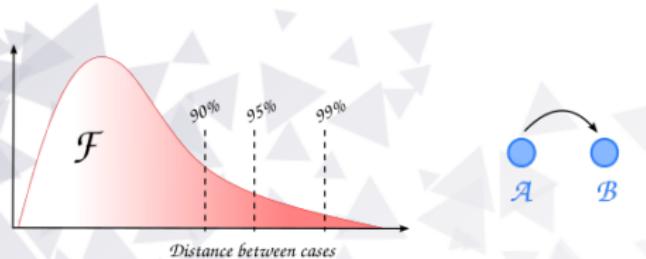


Pruning graphs: where to cut?

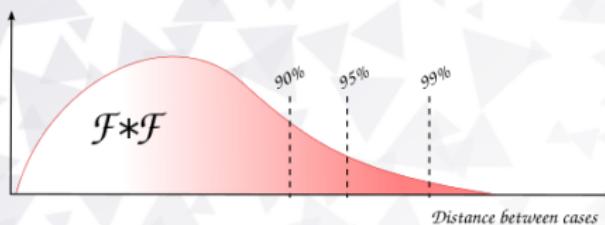
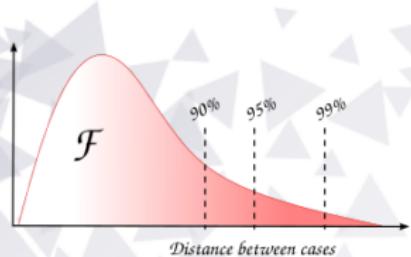
Assuming a known expected distribution between pairs of cases (e.g. serial interval, spatial kernel, molecular clock), different quantiles can be used:



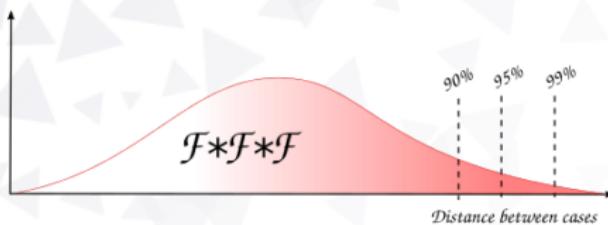
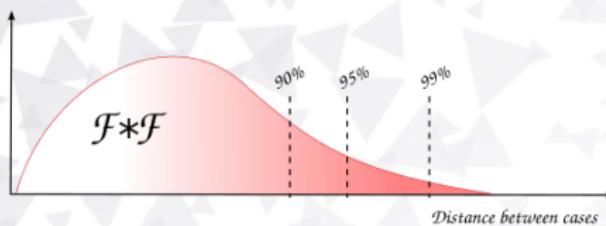
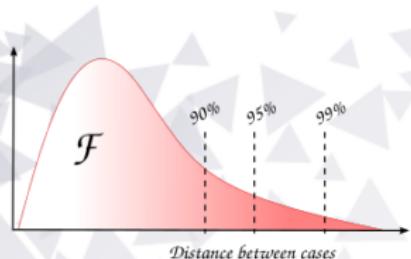
Pruning graphs: where to cut?



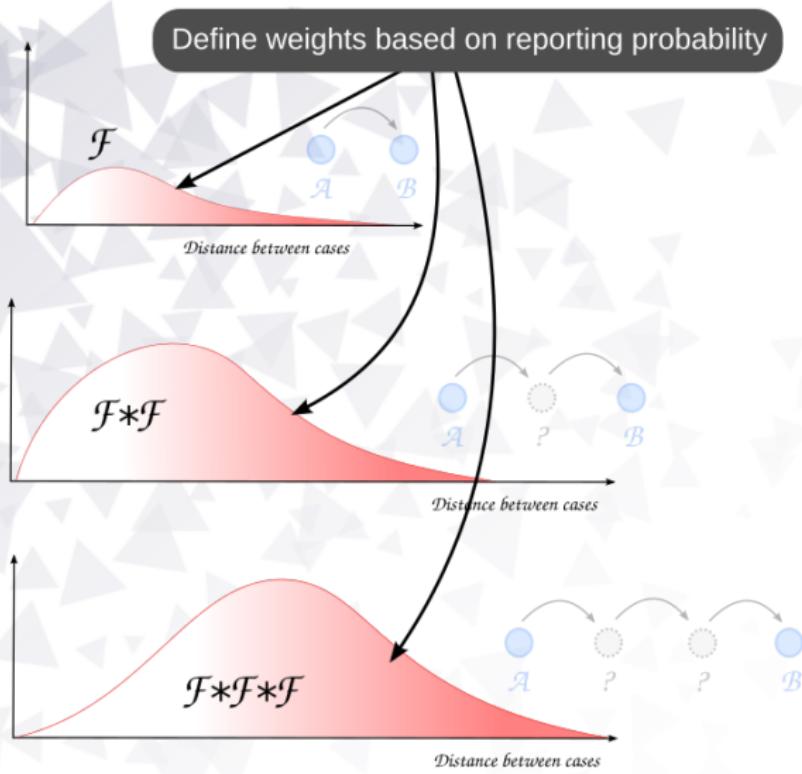
Pruning graphs: where to cut?



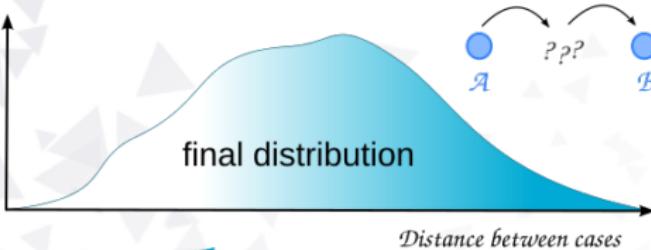
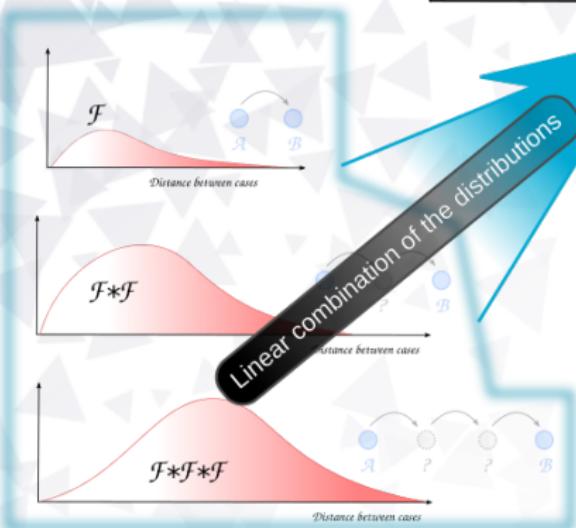
Pruning graphs: where to cut?



Pruning graphs: where to cut?



Pruning graphs: where to cut?



idea by
Pierre Nouvellet



analytical results by
Anne Cori