Outbreak response analytics

When are pathogen genomes useful?

Thibaut Jombart

5th November 2018

SMBE Satellite Workshop

London School of Hygiene and Tropical Medicine
Imperial College London

Emerging disease, early outbreak response context

- situational awareness urgently needed
- limited data available
- questions focus on delays, risk factors, transmissibility
- reproducibility and reliability » refinement and complexity

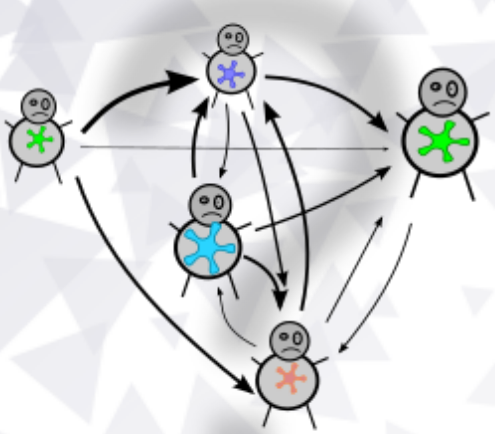
Emerging disease, early outbreak response context

- situational awareness urgently needed
- limited data available
- questions focus on delays, risk factors, transmissibility
- reproducibility and reliability » refinement and complexity

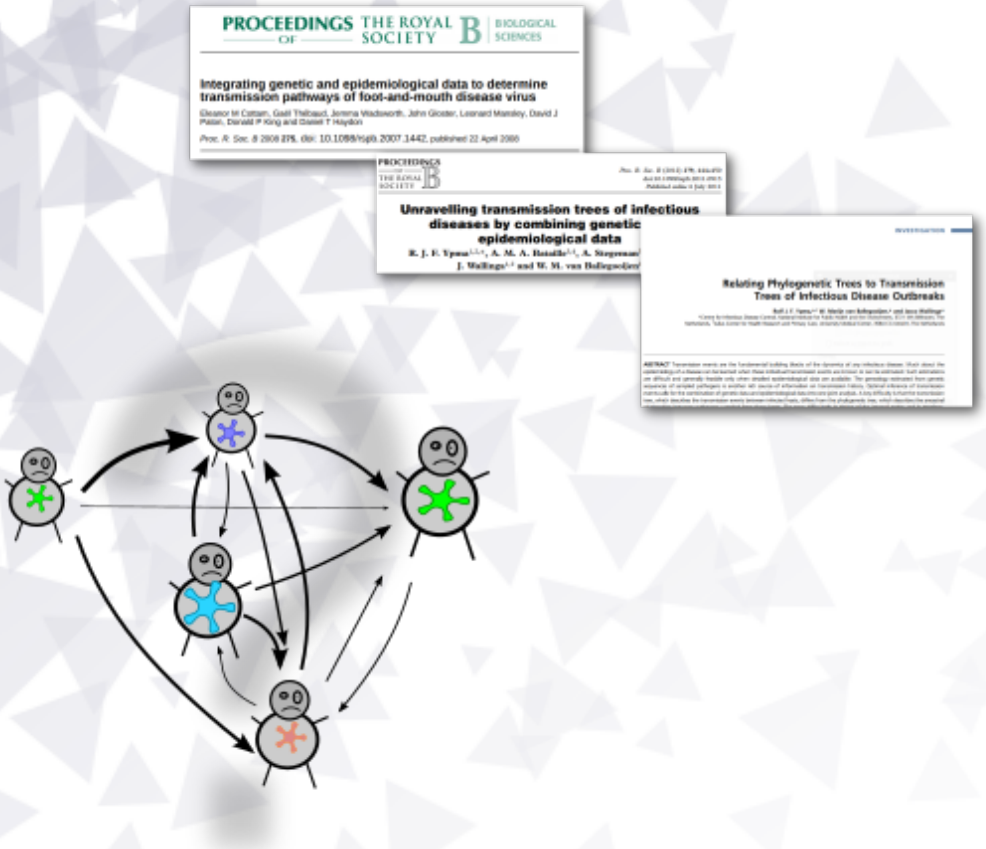
How can data analytics / modelling help?

Reconstructing transmission trees

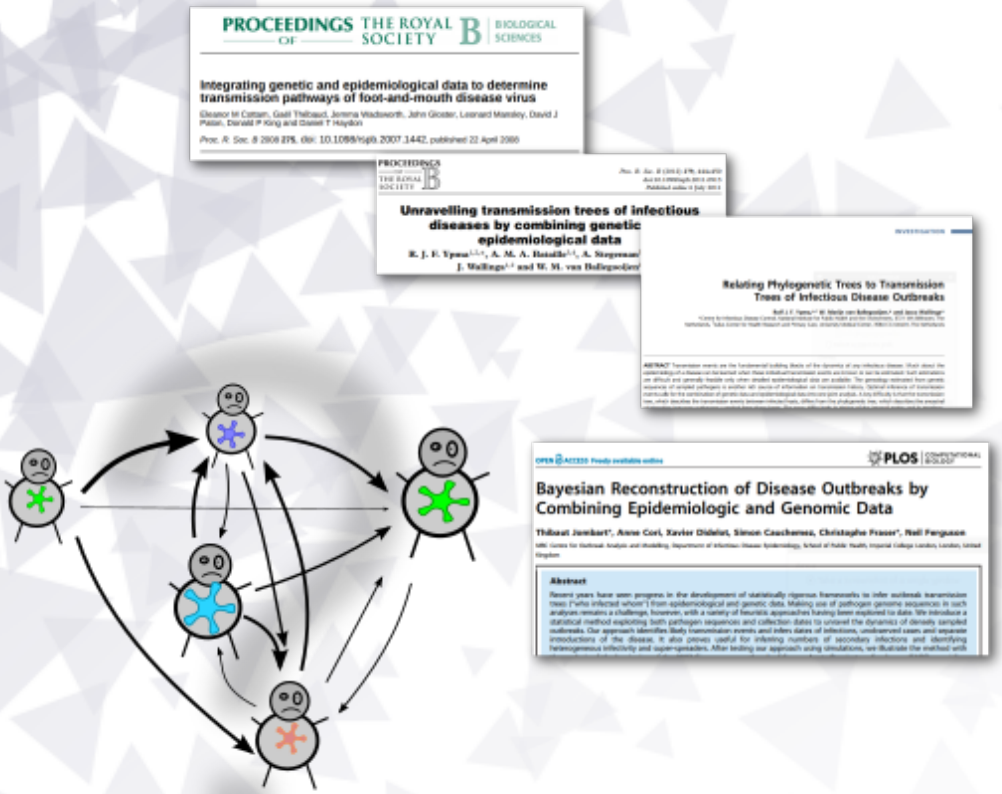
Using genomics to infer who infects whom?



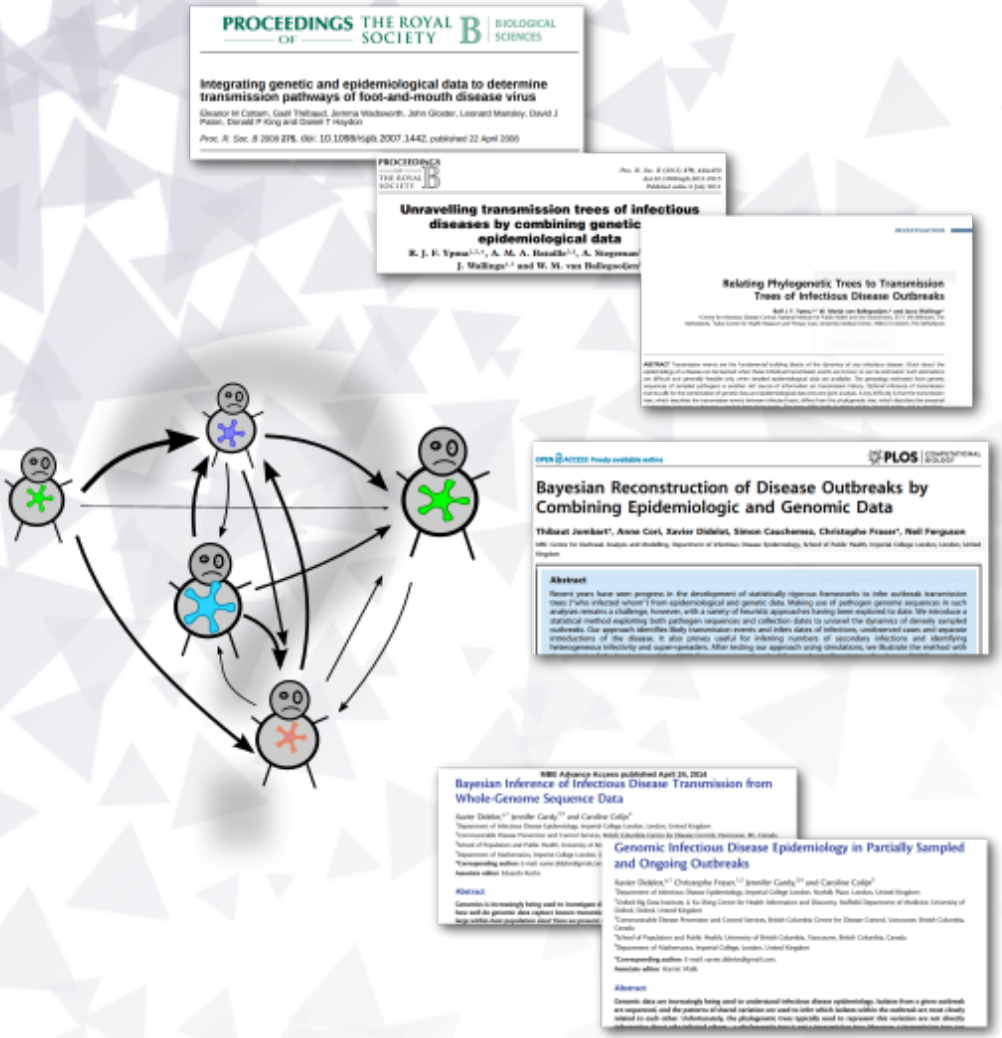
Using genomics to infer who infects whom?



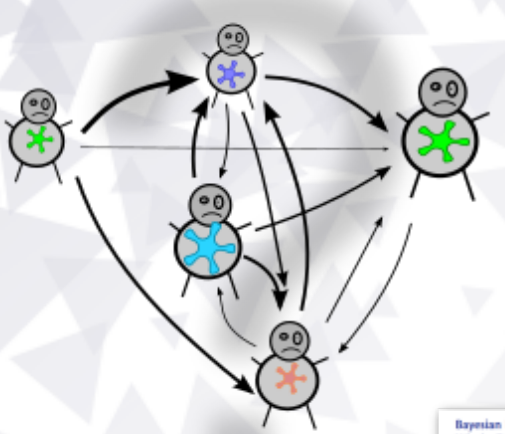
Using genomics to infer who infects whom?



Using genomics to infer who infects whom?



Using genomics to infer who infects whom?



PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus

Eleanor M Cottam, Gaël Thébaud, Jemma Wadsworth, John Gloster, Leonard Mansley, David J Paton, Donald P King and Daniel T Haydon

Proc. R. Soc. B 2008 **275**, doi: 10.1098/rspb.2007.1442, published 22 April 2008

PROCEEDINGS OF THE ROYAL SOCIETY B

Unravelling transmission trees of infectious diseases by combining genetic epidemiological data

R. J. F. Ypma, A. M. A. Bataille, A. Stegeman, J. Wallinga and W. M. van Ballegooijen

Relating Phylogenetic Trees to Transmission Trees of Infectious Disease Outbreaks

PLOS COMPUTATIONAL BIOLOGY

Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data

Thibaut Jombart, Anne Cori, Xavier Didelot, Simon Cauchemez, Christophe Fraser, Neil Ferguson

Bayesian Inference of Infectious Disease Transmission from Whole-Genome Sequence Data

Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks

HHS Public Access

Author manuscript

Ann Appl Stat. Author manuscript; available in PMC 2016 April 01.

Reconstructing transmission trees for communicable diseases using densely sampled genetic data

Colin J. Worby, Philip D. O'Neill, Theodore Kypraios, Julie V. Robotham, Daniela De Angelis, Edward J. P. Cartwright, Sharon J. Peacock, and Ben S. Cooper

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

A Systematic Bayesian Integration of Epidemiological and Genetic Data

Wai S. Y. Lau, Glenn Marion, George Streftaris, Gavin Gibson

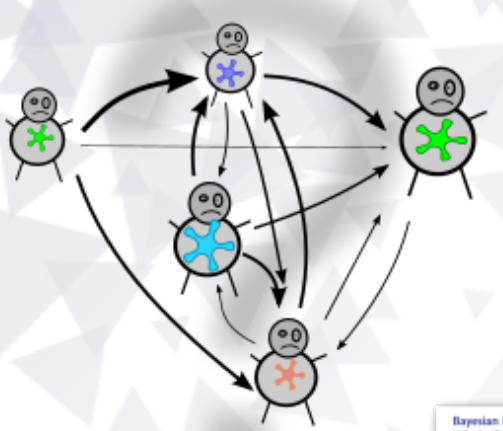
PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data

Nadius Mollentze, Louis H. Nel, Sunny Townsend, Kevin le Roux, Katie Hampson, Daniel T. Haydon and Samuel Soubeyrand

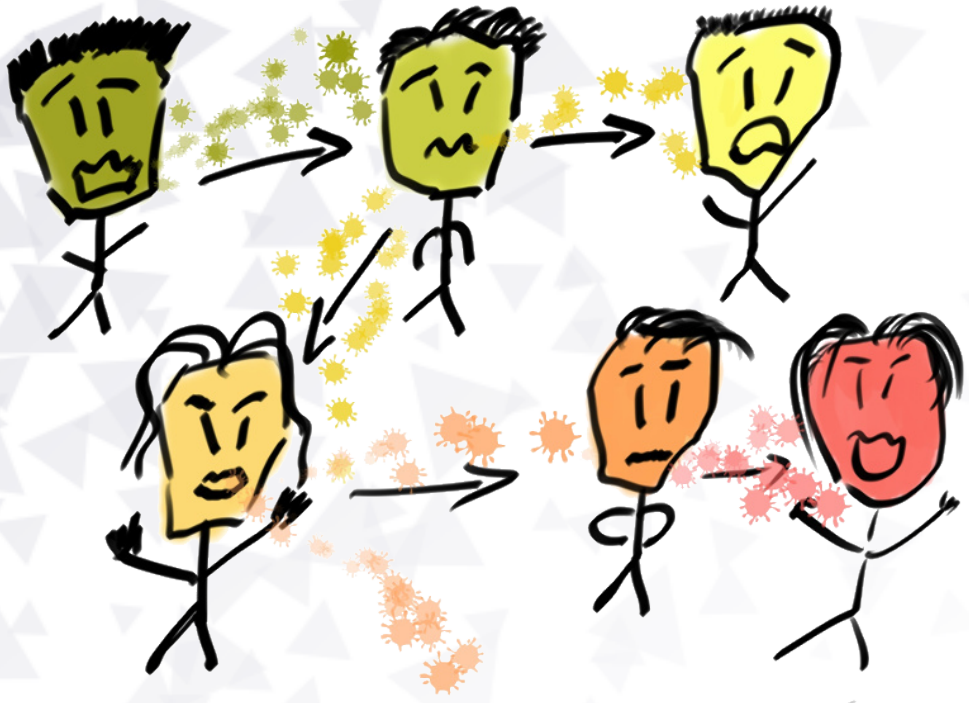
Proc. R. Soc. B 2014 **281**, 20133251, published 11 March 2014

Using genomics to infer who infects whom?



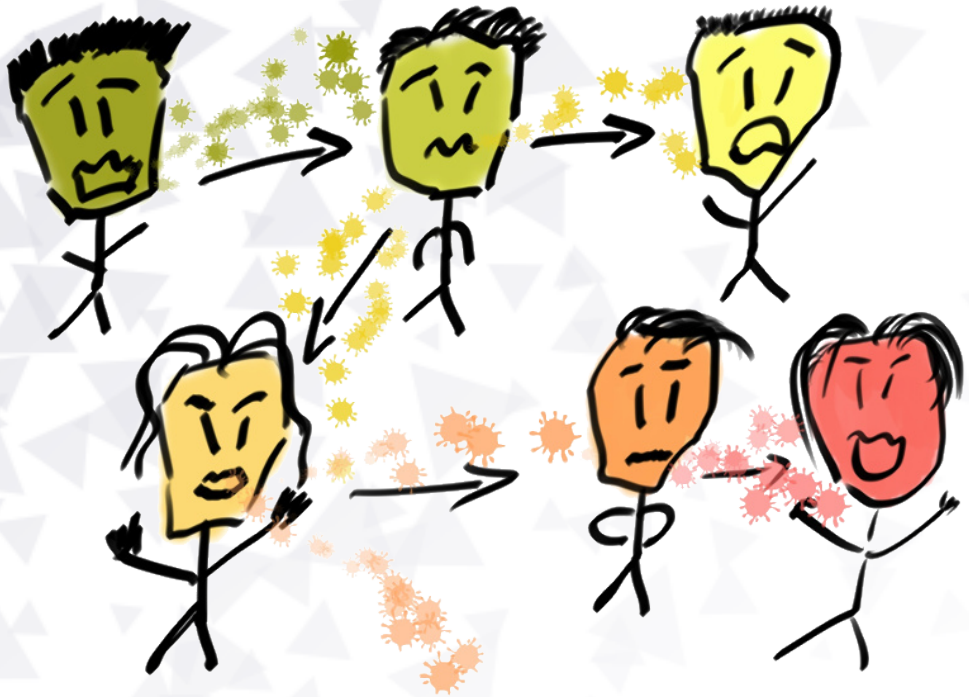
Methods heavily rely on whole genome sequence data

Using WGS to infer who infected whom



Mutations accumulate along transmission chains.

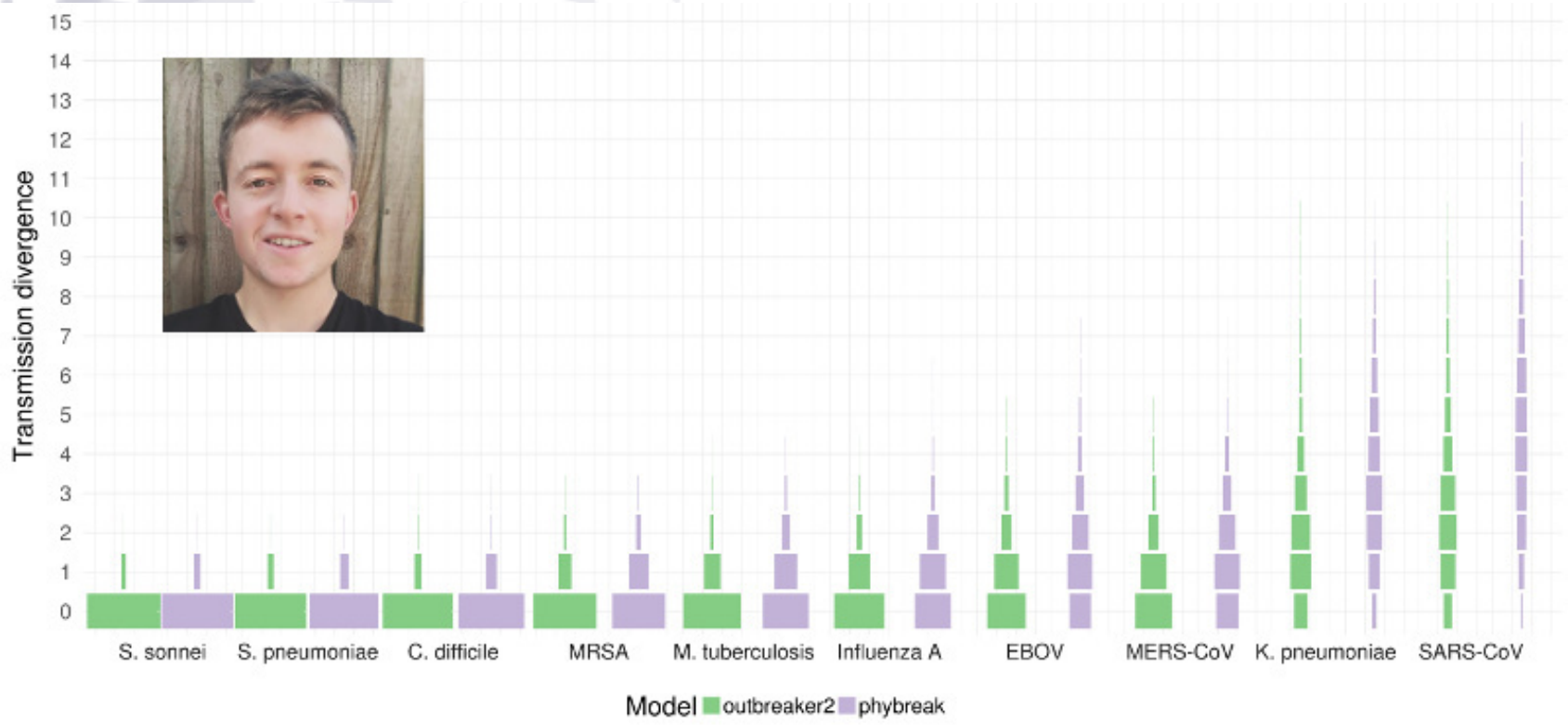
Using WGS to infer who infected whom



Mutations accumulate along transmission chains.

Can be used to reconstruct transmission trees.

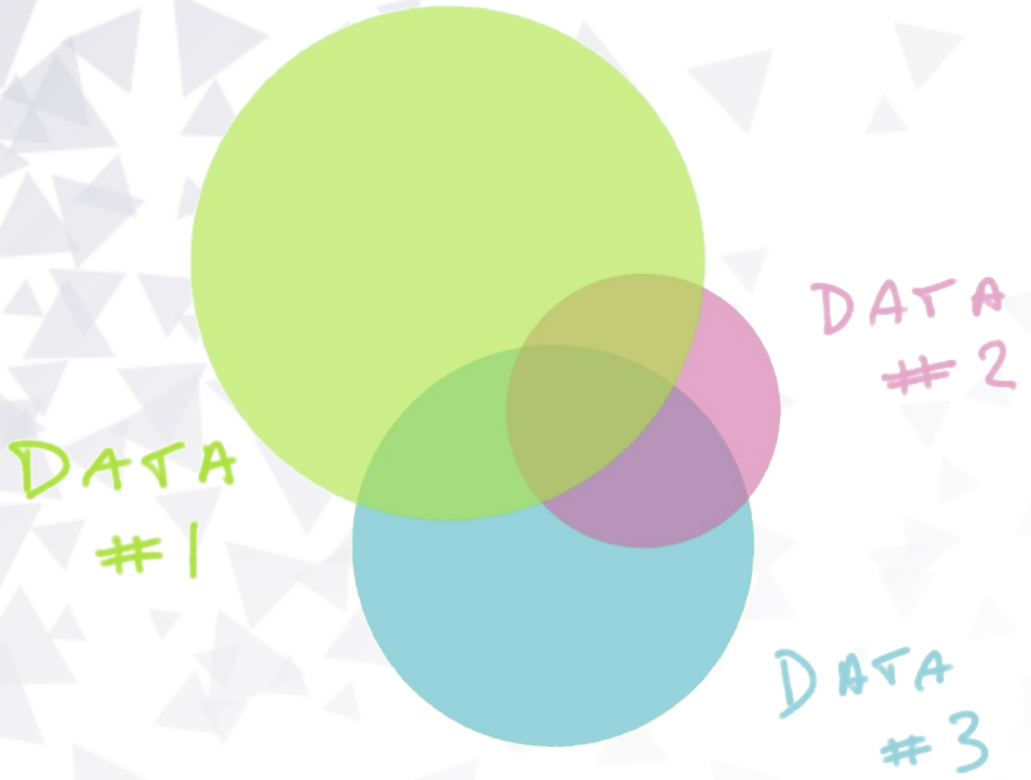
How informative are whole genome sequences?



[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.

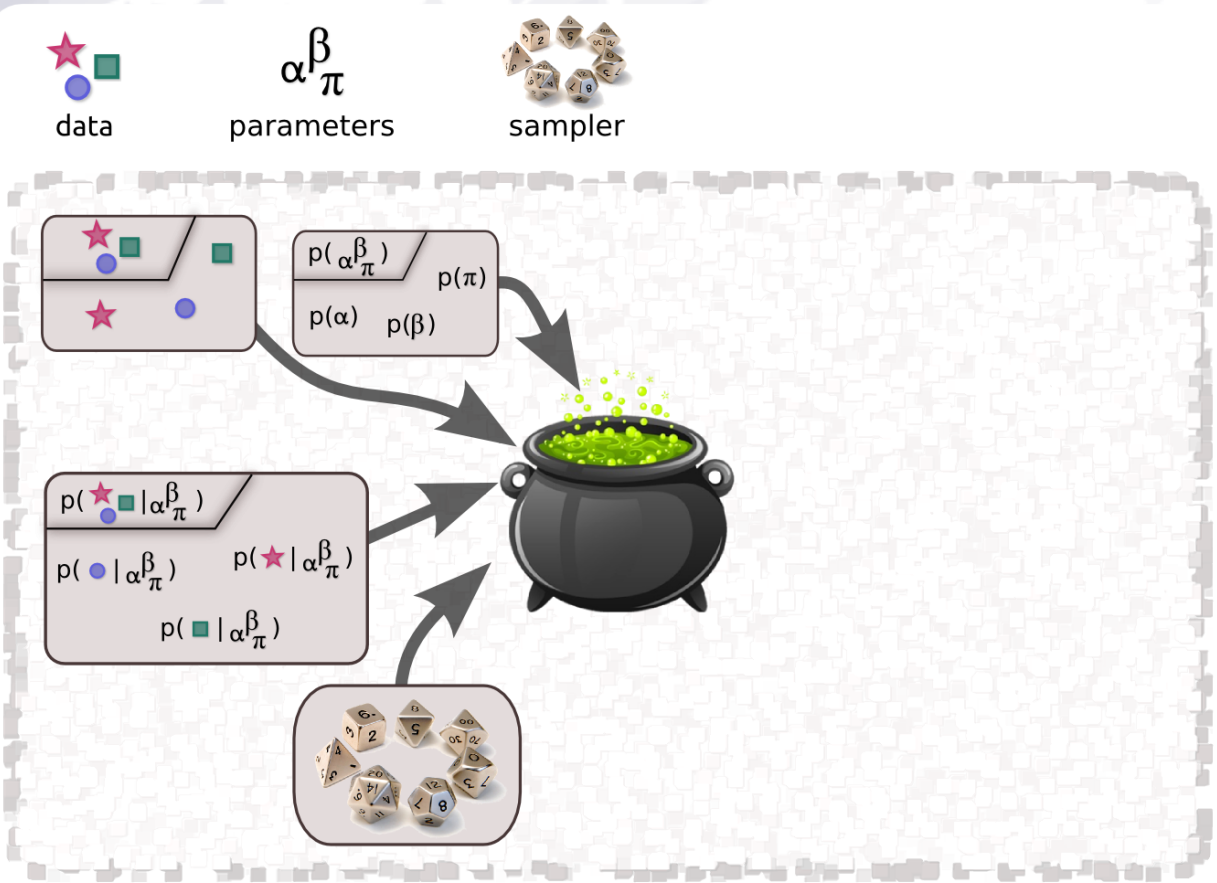
outbreaker2: evidence synthesis framework for outbreak reconstruction

Modularity: customise data, prior, likelihood, MCMC.

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

outbreaker2: evidence synthesis framework for outbreak reconstruction

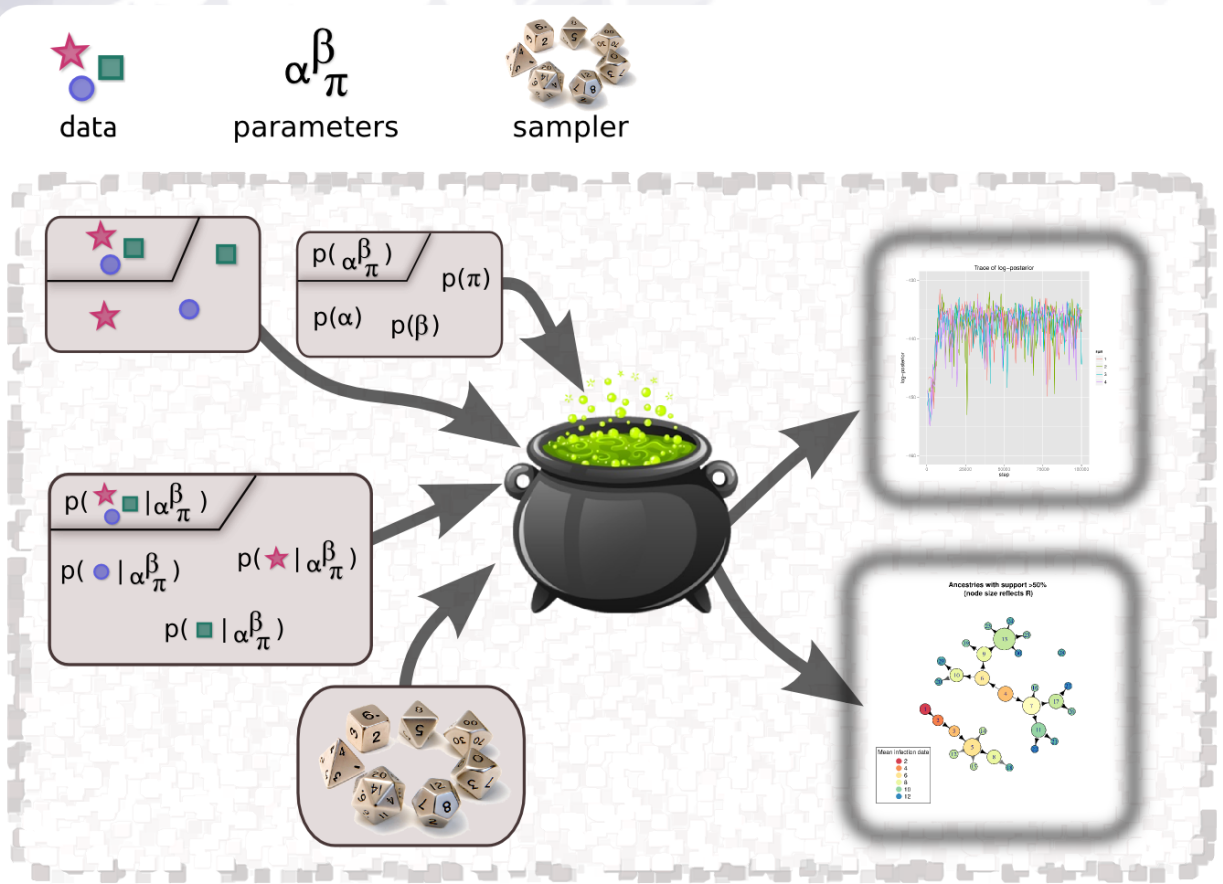
Modularity: customise data, prior, likelihood, MCMC.



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

outbreaker2: evidence synthesis framework for outbreak reconstruction

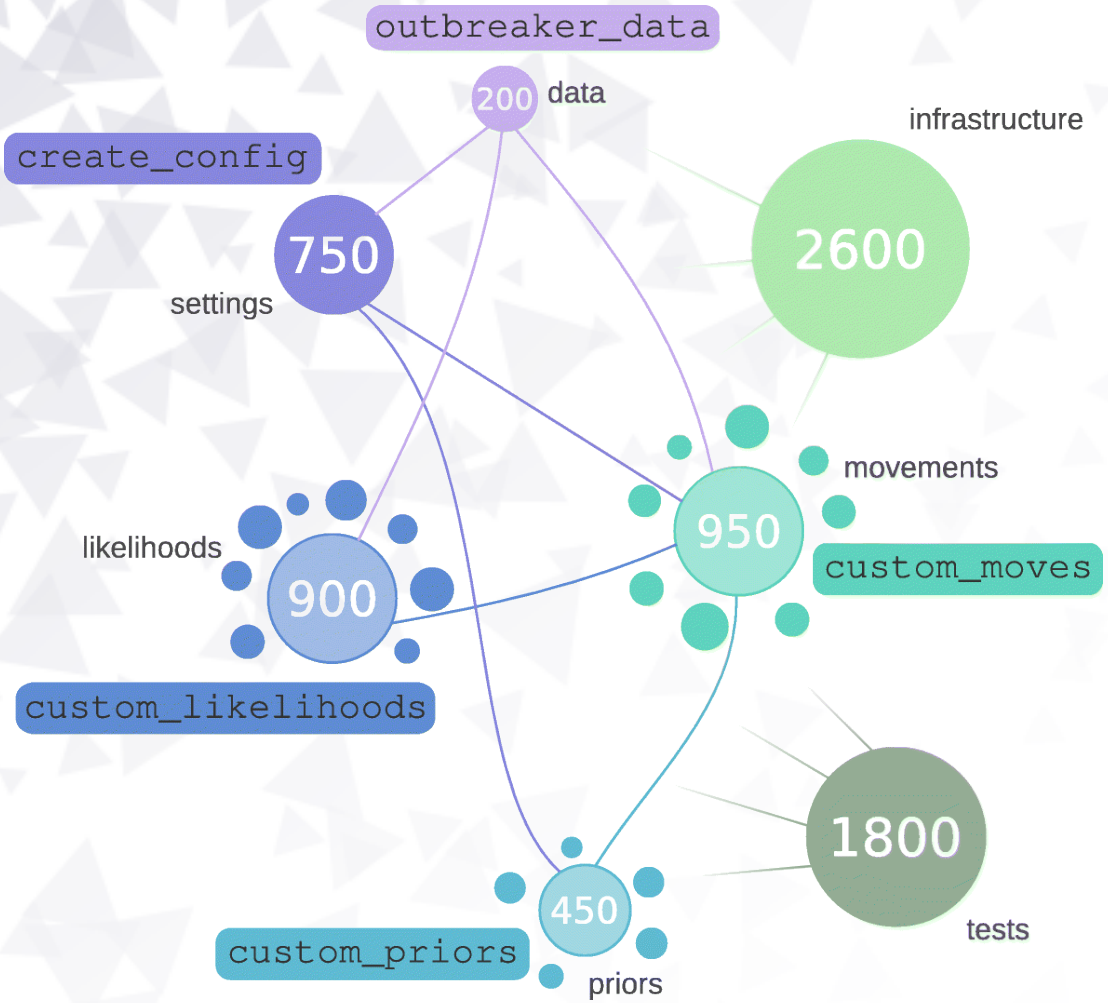
Modularity: customise data, prior, likelihood, MCMC.



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

What is inside the pot?

Module sizes in lines of code, and entry points:



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

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* x *incubation* x *genetic (simple)* x *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* x *incubation* x *genetic (simple)* x *missing cases*

TransPhylo likelihood

- $p(G|\beta, \gamma, N_e g, \alpha) = p(G|N_e g, \alpha) \times p(\alpha|\beta, \gamma)$
- i.e. *phylogeny (coalescent)* x *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* x *incubation* x *genetic (simple)* x *missing cases*

TransPhylo likelihood

- $p(G|\beta, \gamma, N_e g, \alpha) = p(G|N_e g, \alpha) \times p(\alpha|\beta, \gamma)$
- i.e. *phylogeny (coalescent)* x *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 **probPTreeGivenTTree**. 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$nam)
  ttree$ttree[which(is.na(ttree$ttree[,3])),3] <- 0
  txt <- capture.output(ctree <- combine(ttree,data$ptree))
  if (length(txt)==0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
    } else {
      prob <- -Inf
      }
  return(prob)
}
```

```
args(api$cpp_ll_all)

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

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:data$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(-100:-1,1:100), 1)
    param$t_inf[i] <- param$t_inf[i] + modif
    new_ll <- api$cpp_ll_all(data, param, i = NULL, list_custom_ll)
    if (log(runif(1)) > (new_ll - current_ll)) {
      param$t_inf[i] <- param$t_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 //
## $mu
```

[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 **probPTreeGivenTTree**. 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$nam)
  ttree$ttree[which(is.na(ttree$ttree[,3])),3] <- 0
  txt <- capture.output(ctree <- combine(ttree,data$ptree))
  if (length(txt)==0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
    } else {
      prob <- -Inf
      }
  return(prob)
}
```

likelihood

```
args(api$cpp_ll_all)

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

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:data$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(-100:-1,1:100), 1)
    param$t_inf[i] <- param$t_inf[i] + modif
    new_ll <- api$cpp_ll_all(data, param, i = NULL, list_custom_ll)
    if (log(runif(1)) > (new_ll - current_ll)) {
      param$t_inf[i] <- param$t_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 //
## $mu
```

[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 `probPTreeGivenTTree`. 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$nam)
  ttree$ttree[which(is.na(ttree$ttree[,3])),3] <- 0
  txt <- capture.output(ctree <- combine(ttree,data$ptree))
  if (length(txt)==0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
    } else {
      prob <- -Inf
      }
  return(prob)
}
```

likelihood

movement function

```
args(api$cpp_ll_all)

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

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:data$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(-100:-1,1:100), 1)
    param$t_inf[i] <- param$t_inf[i] + modif
    new_ll <- api$cpp_ll_all(data, param, i = NULL, list_custom_ll)
    if (log(runif(1)) > (new_ll - current_ll)) {
      param$t_inf[i] <- param$t_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 //
## $mu
```

[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 `probPTreeGivenTTree`. 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$nam)
  ttree$ttree[which(is.na(ttree$ttree[,3])),3] <- 0
  txt <- capture.output(ctree <- combine(ttree,data$ptree))
  if (length(txt)==0) {
    prob <- probPTreeGivenTTree(ctree, neg = 365 * 0.25)
    } else {
      prob <- -Inf
      }
  return(prob)
}
```

likelihood

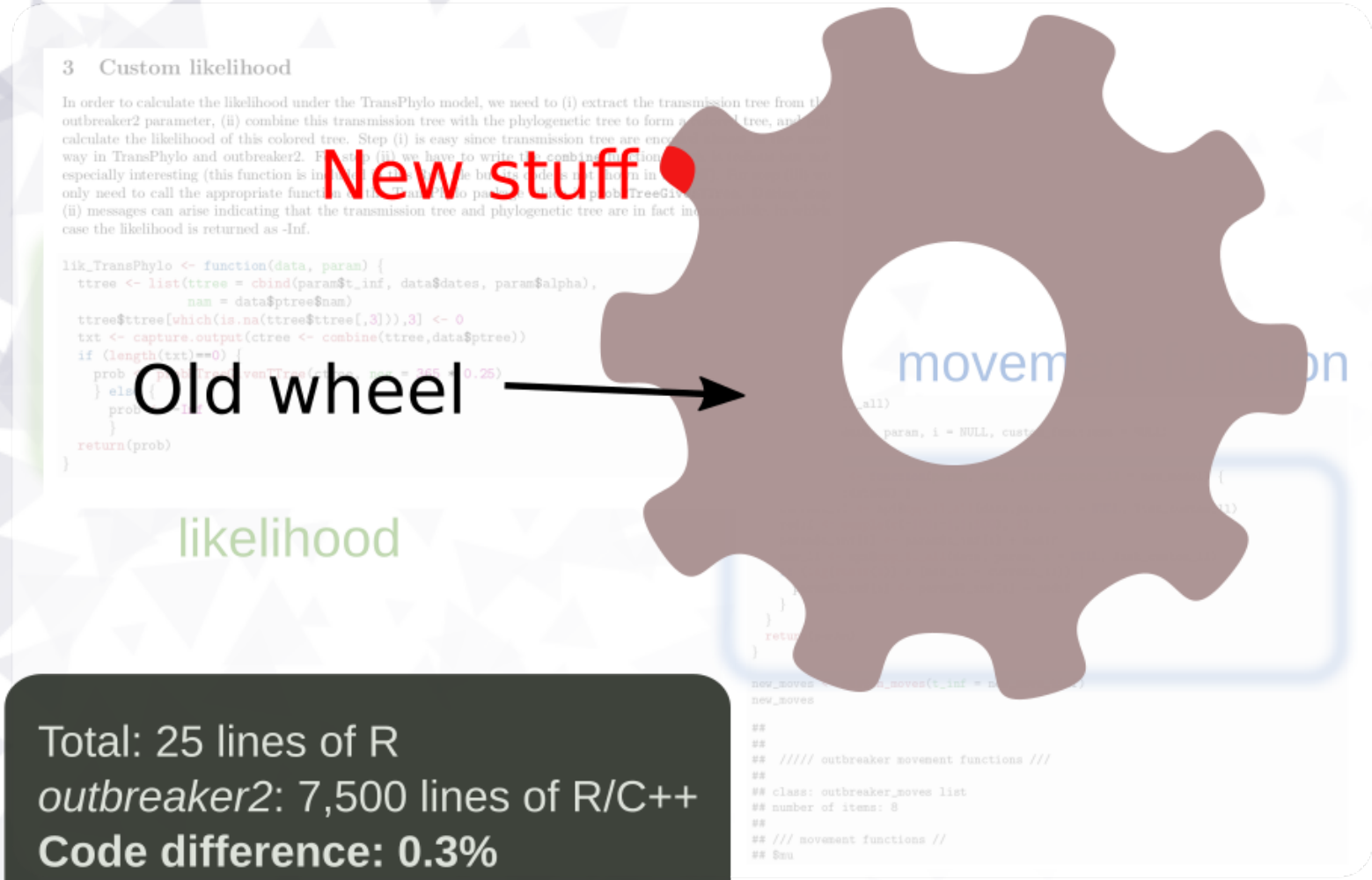
movement function

```
args(api$cpp_ll_all)

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

new_move_tinf <- function(param, data, list_custom_ll = new_model) {
  for (i in 1:data$N) {
    current_ll <- api$cpp_ll_all(data,param, i = NULL, list_custom_ll)
    modif <- sample(c(-100:-1,1:100), 1)
    param$t_inf[i] <- param$t_inf[i] + modif
    new_ll <- api$cpp_ll_all(data, param, i = NULL, list_custom_ll)
    if (log(runif(1)) > (new_ll - current_ll)) {
      param$t_inf[i] <- param$t_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 //
## $mu
```

Total: 25 lines of R
outbreaker2: 7,500 lines of R/C++
Code difference: 0.3%

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

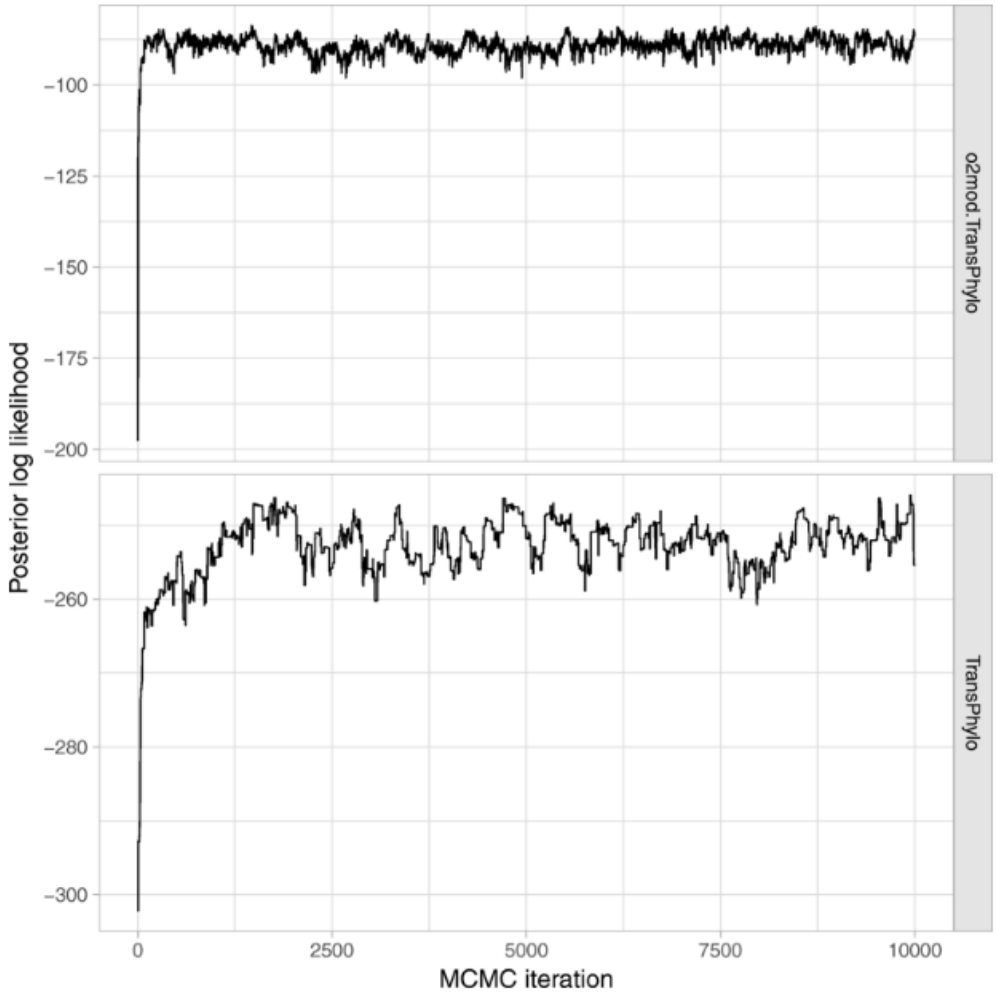
TransPhylo module for *outbreaker2*



Total: 25 lines of R
outbreaker2: 7,500 lines of R/C++
Code difference: 0.3%

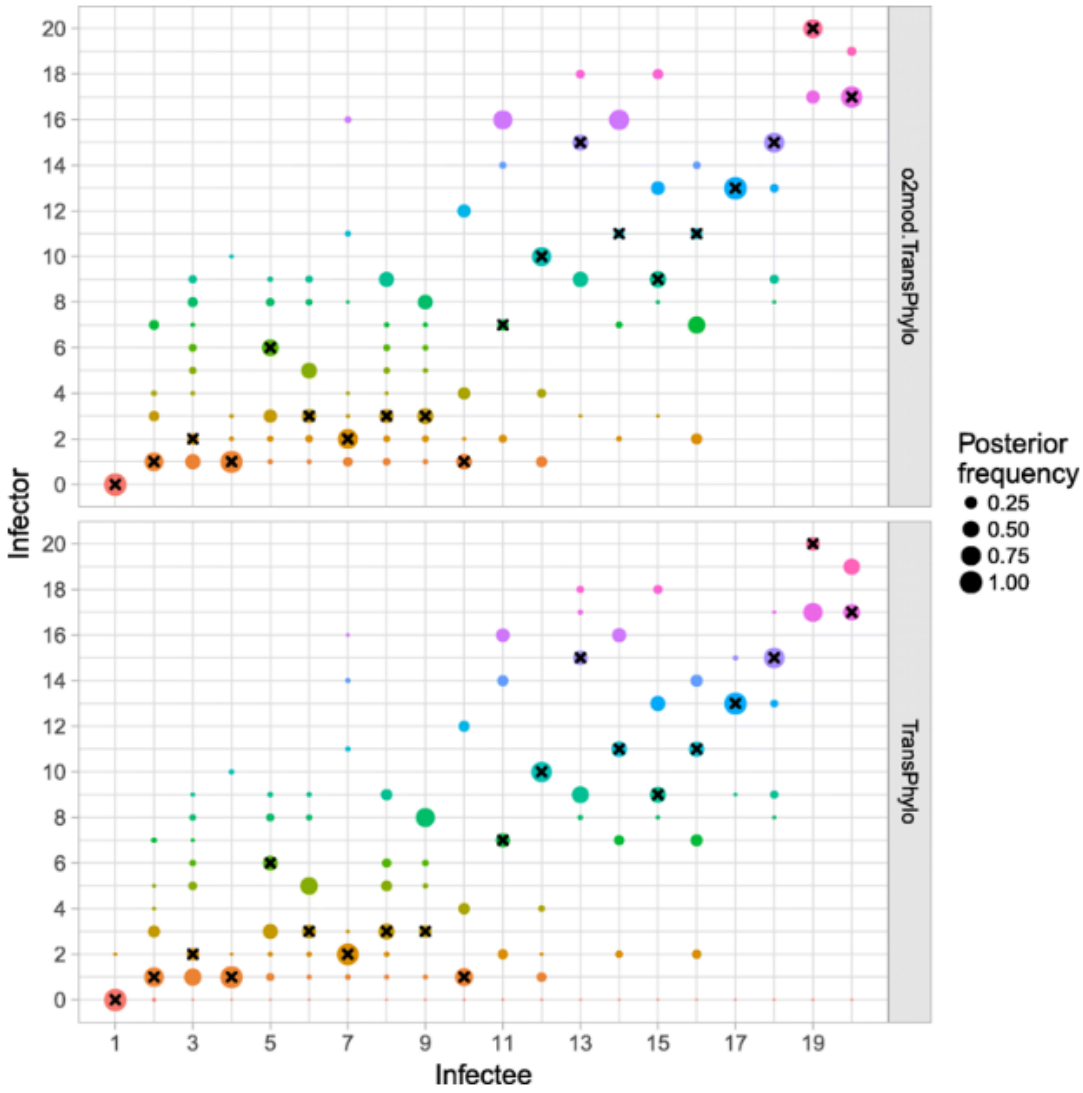
[Campbell *et al.* (2018) BMC Bioinformatics]

Transphylo module results (1/2): convergence



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

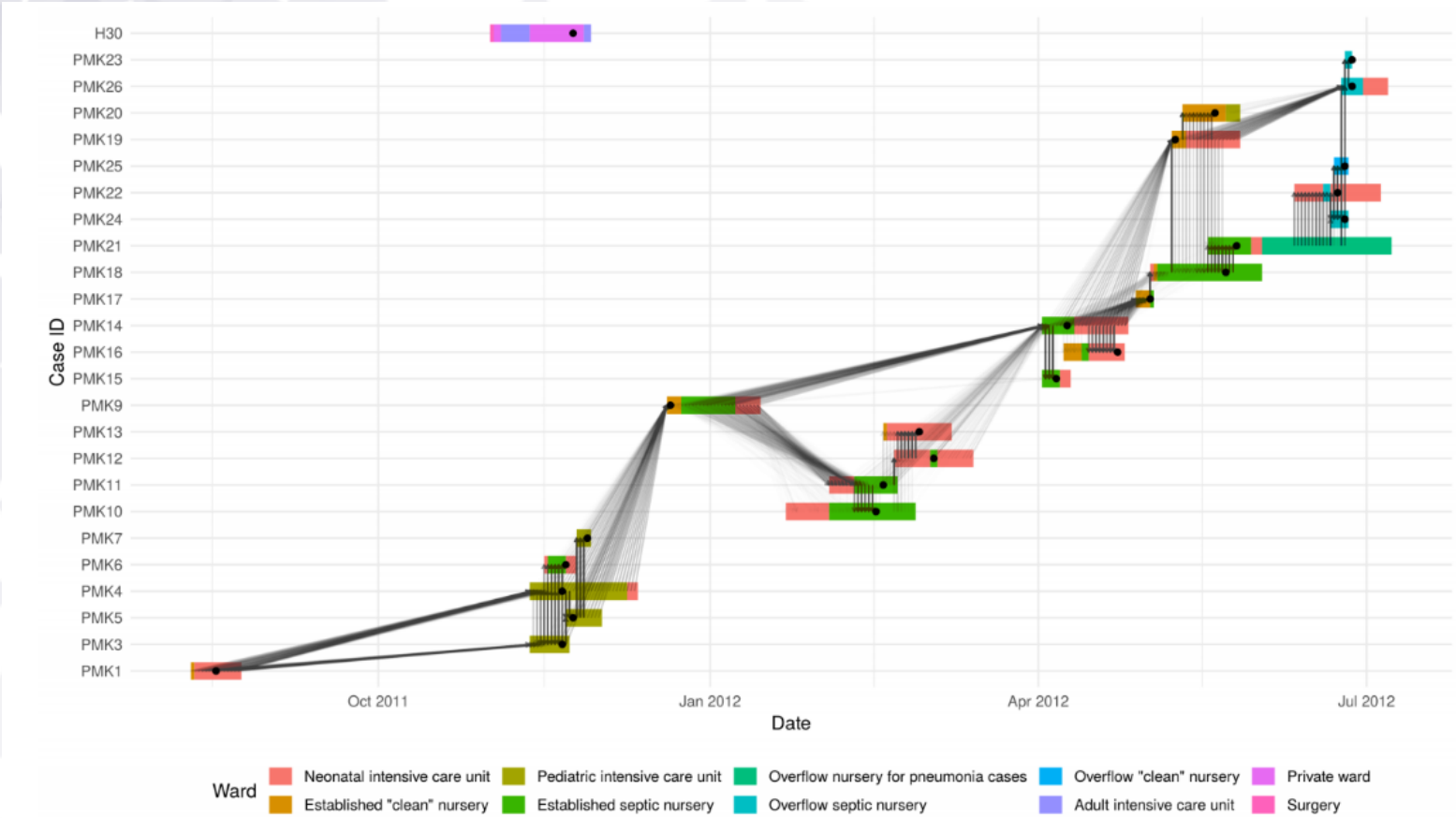
Transphylo module results (2/2): trees



[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

Who infects whom: when do we care?

- complex methods, WGS data costly: is it worth it?

Who infects whom: when do we care?

- complex methods, WGS data costly: is it worth it?
- in general, not useful for forecasting

Who infects whom: when do we care?

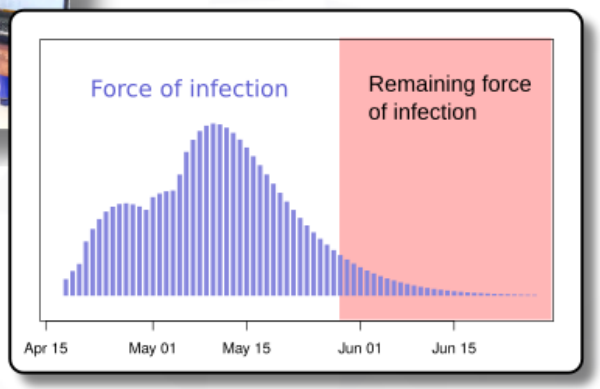
- complex methods, WGS data costly: is it worth it?
- in general, not useful for forecasting
- useful to detect multiple introductions or superspreading

Who infects whom: when do we care?

- complex methods, WGS data costly: is it worth it?
- in general, not useful for forecasting
- useful to detect multiple introductions or superspreading
- complement exposure / contact tracing data

Scaling a response in real time

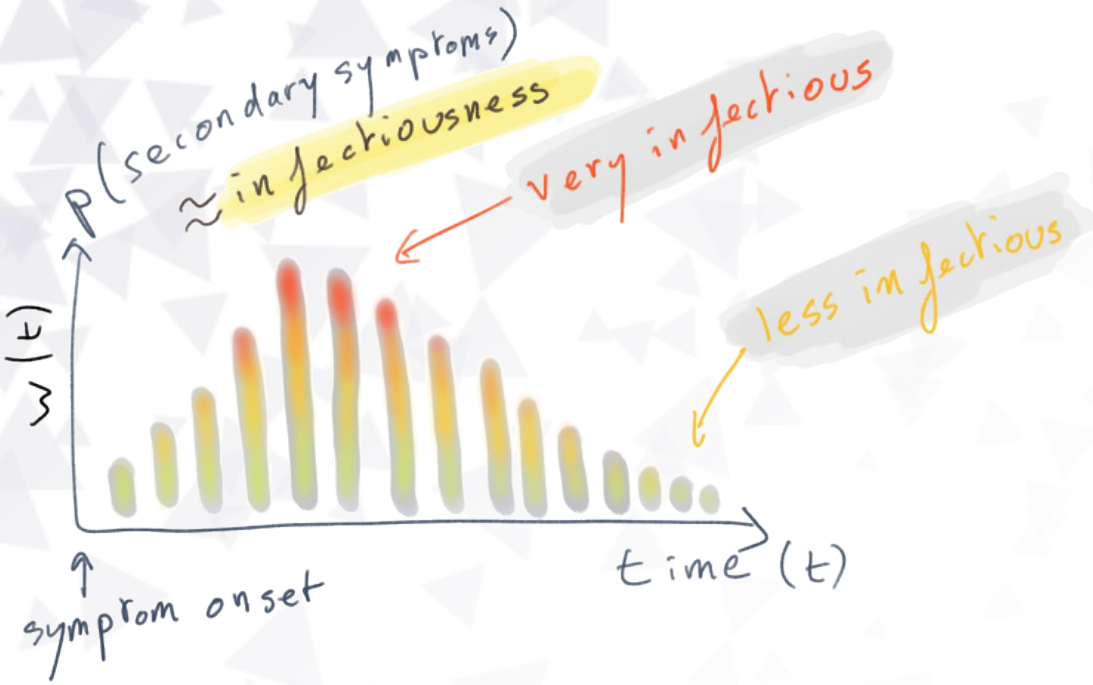
Ebola outbreak, Likati (DRC) 2017



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

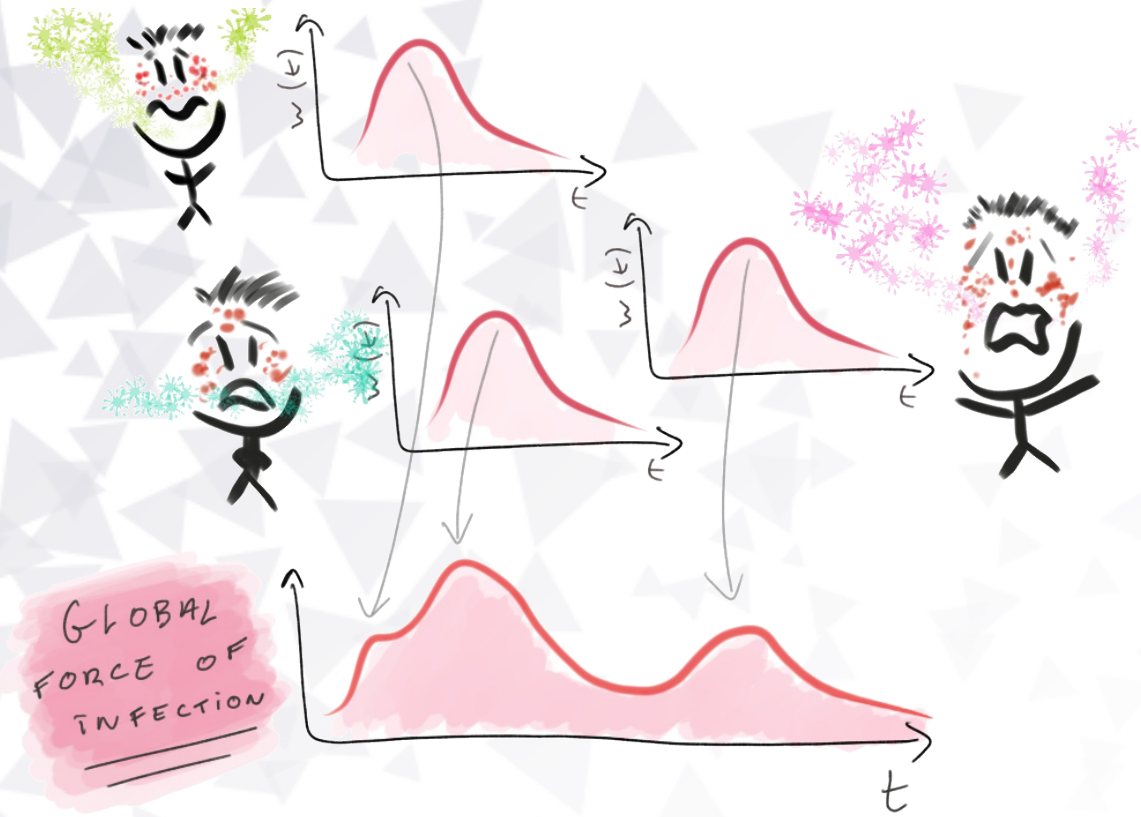
Individual infectiousness over time

Serial interval: delay between symptom onset in infector and infectees



Indicates when we expect new cases, if there are any.

A "simple" branching process model



$$y_t \sim \mathcal{P}(\lambda_t) \quad ; \quad \lambda_t = R_0 \times \sum_i w(t - t_i)$$

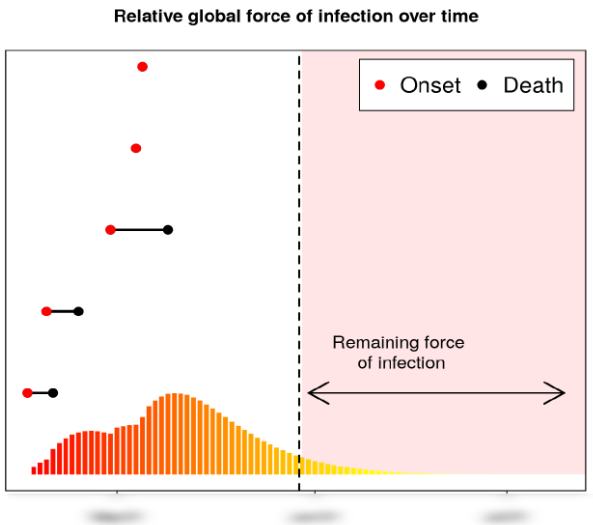
y_t: incidence at time t; $\mathcal{P}()$: Poisson distribution; λ_t: global force of infection; $w()$: serial interval distribution; t_i: date of symptom onset

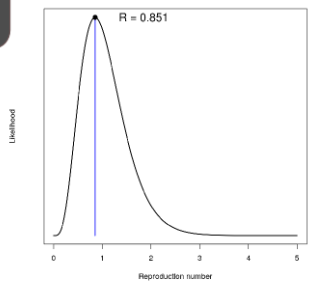
A model for short-term forecasting

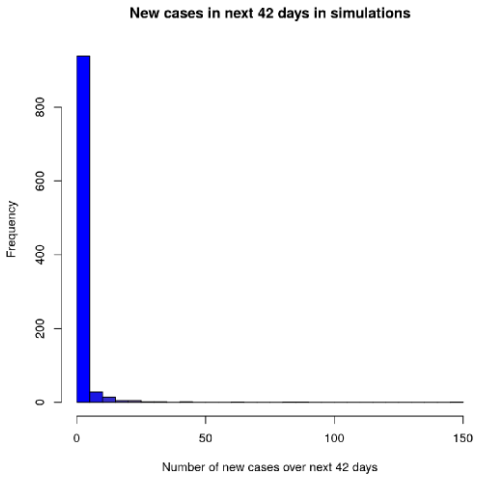
1. estimate R from incidence $y_1, \ldots, y_t$ until time t
2. simulate incidence $_{t+1} \sim \mathcal{P}(\lambda_{t+1})$
3. increate t by one day, repeat

Scaling the response in real-time

Estimating remaining force of infection, transmissibility (R), predicting new cases

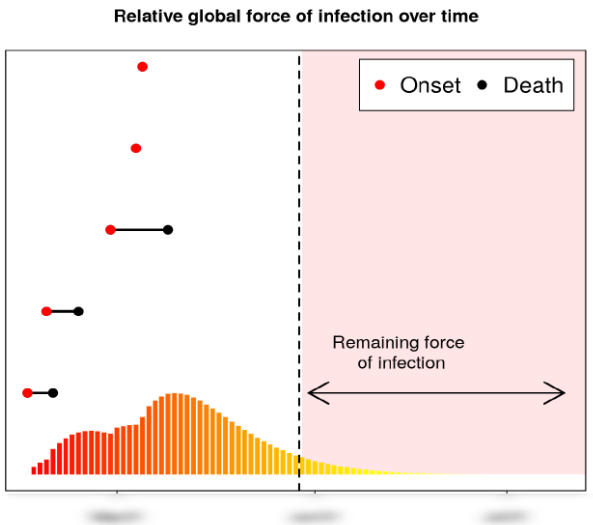


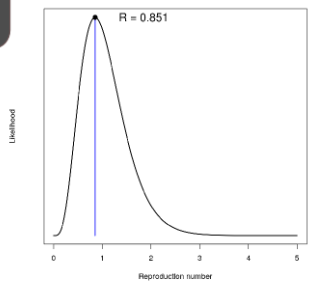


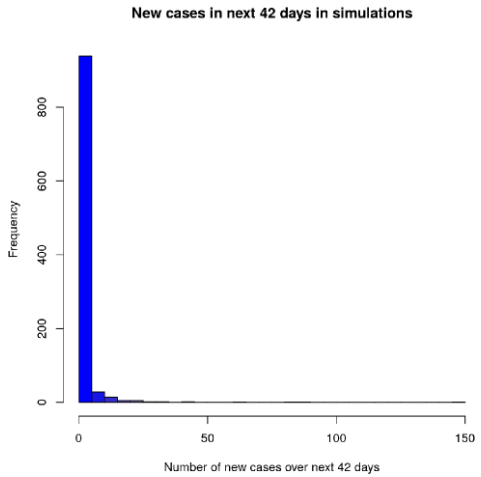


Scaling the response in real-time

Estimating remaining force of infection, transmissibility (R), predicting new cases



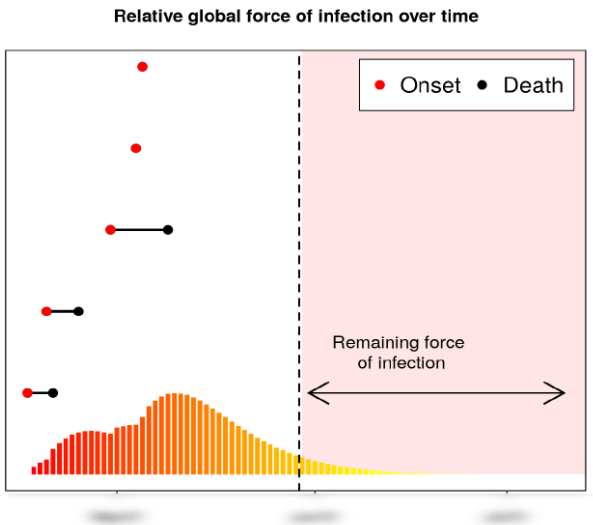


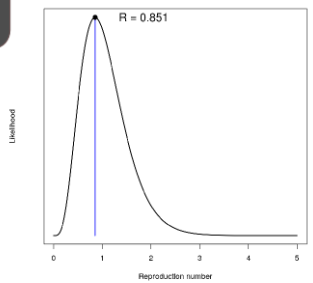


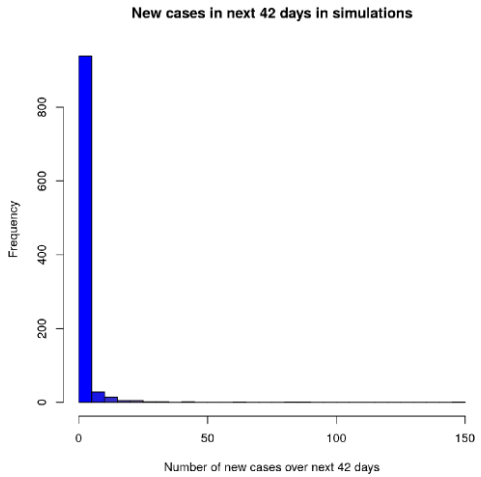
Despite uncertainty in R_0, new cases were unlikely.

Scaling the response in real-time

Estimating remaining force of infection, transmissibility (R), predicting new cases





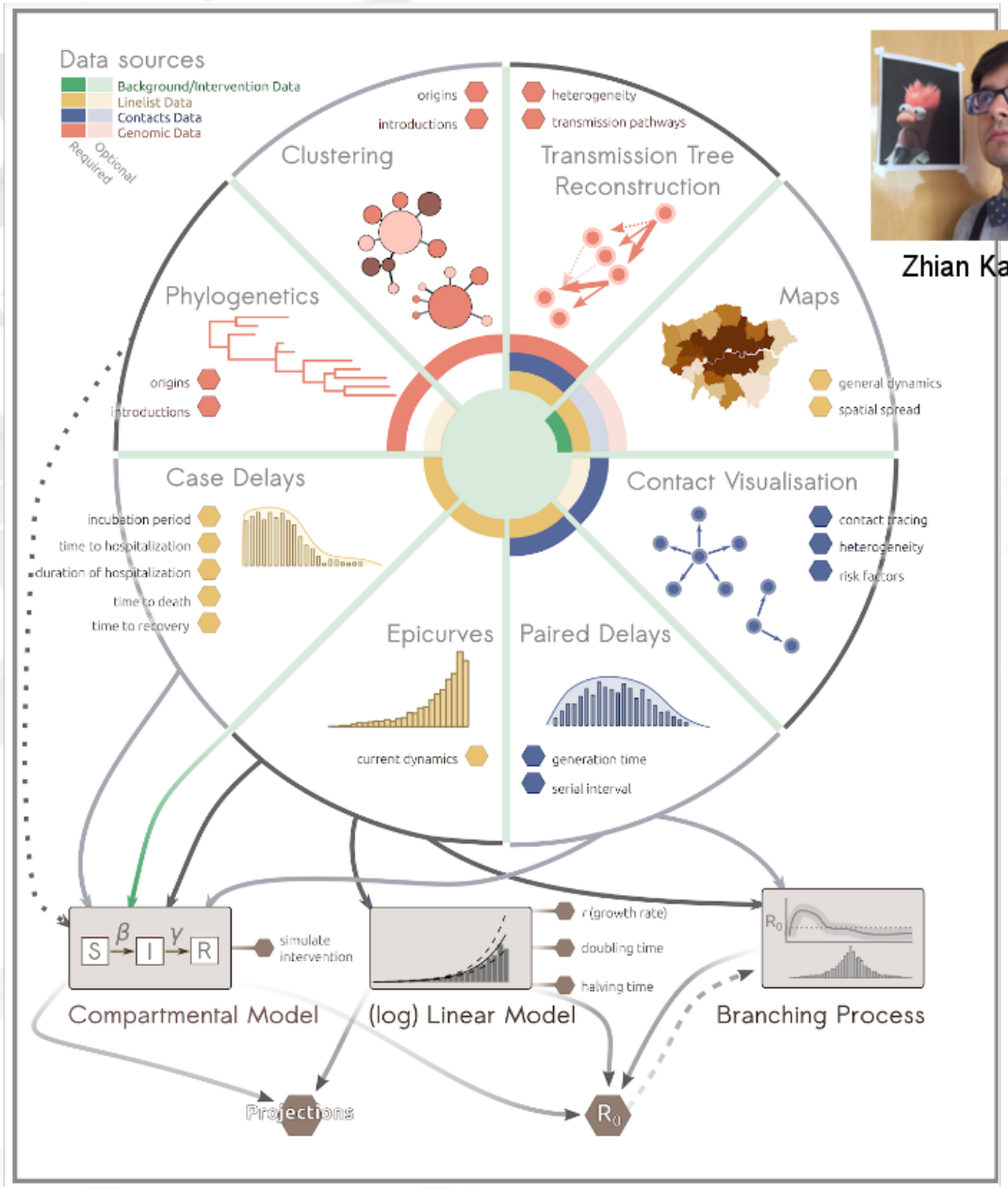


Despite uncertainty in R_0, new cases were unlikely.

Discouraged scaling up in resource-limited context.

Outbreak analytics: some practical considerations

Cost-effective analyses: data needs *vs* actionable intel



Zhian Kamvar

Centralised analyses, distributed delays

Affected countries

data collection

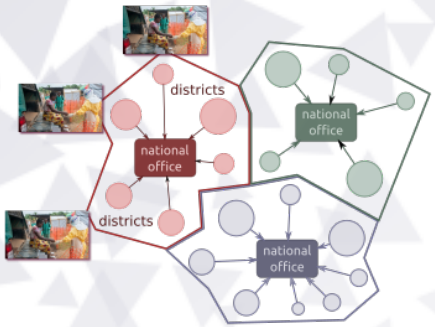
Centralised analyses, distributed delays

Affected countries

time (block = day)

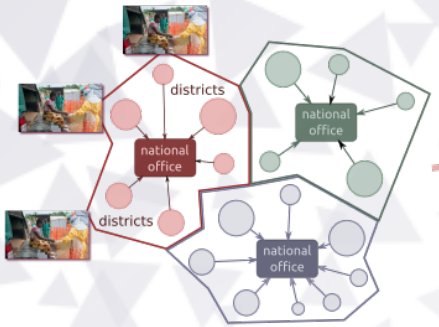
data collection

Centralised analyses, distributed delays



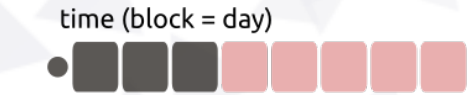
Affected countries

Centralised analyses, distributed delays



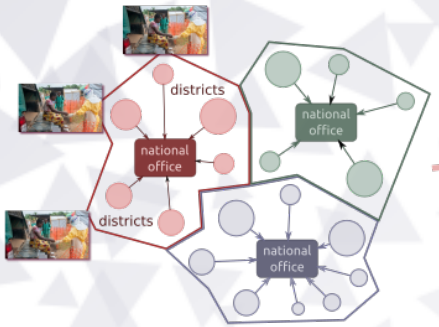
Affected countries

Public Health agencies



data collection

Centralised analyses, distributed delays

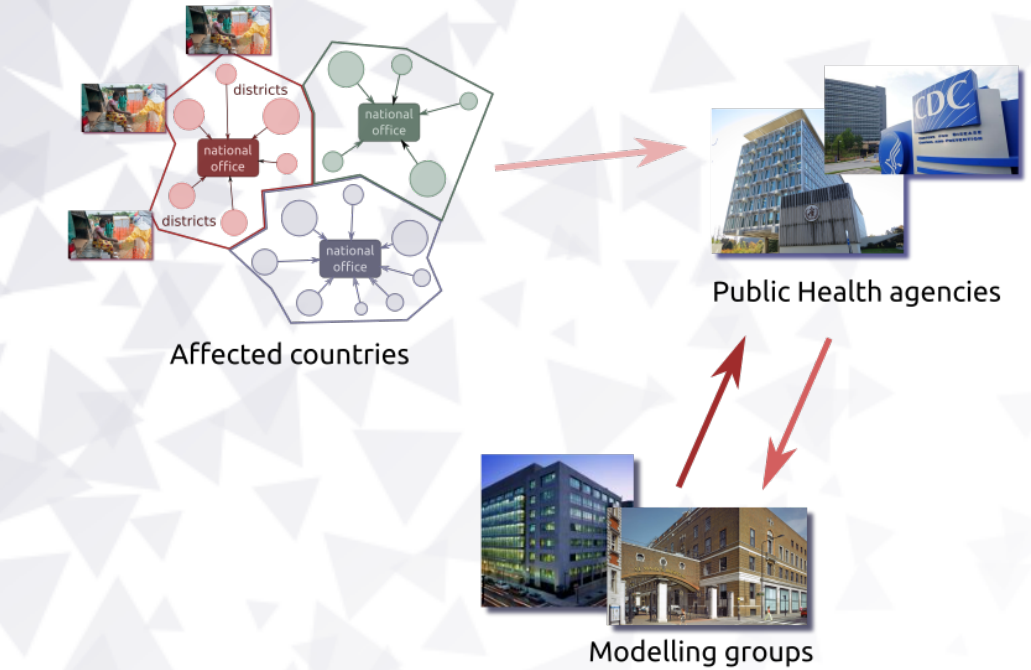


Affected countries

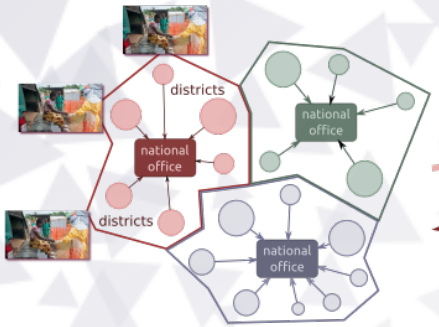
Public Health agencies

Modelling groups

Centralised analyses, distributed delays



Centralised analyses, distributed delays



Affected countries

Public Health agencies

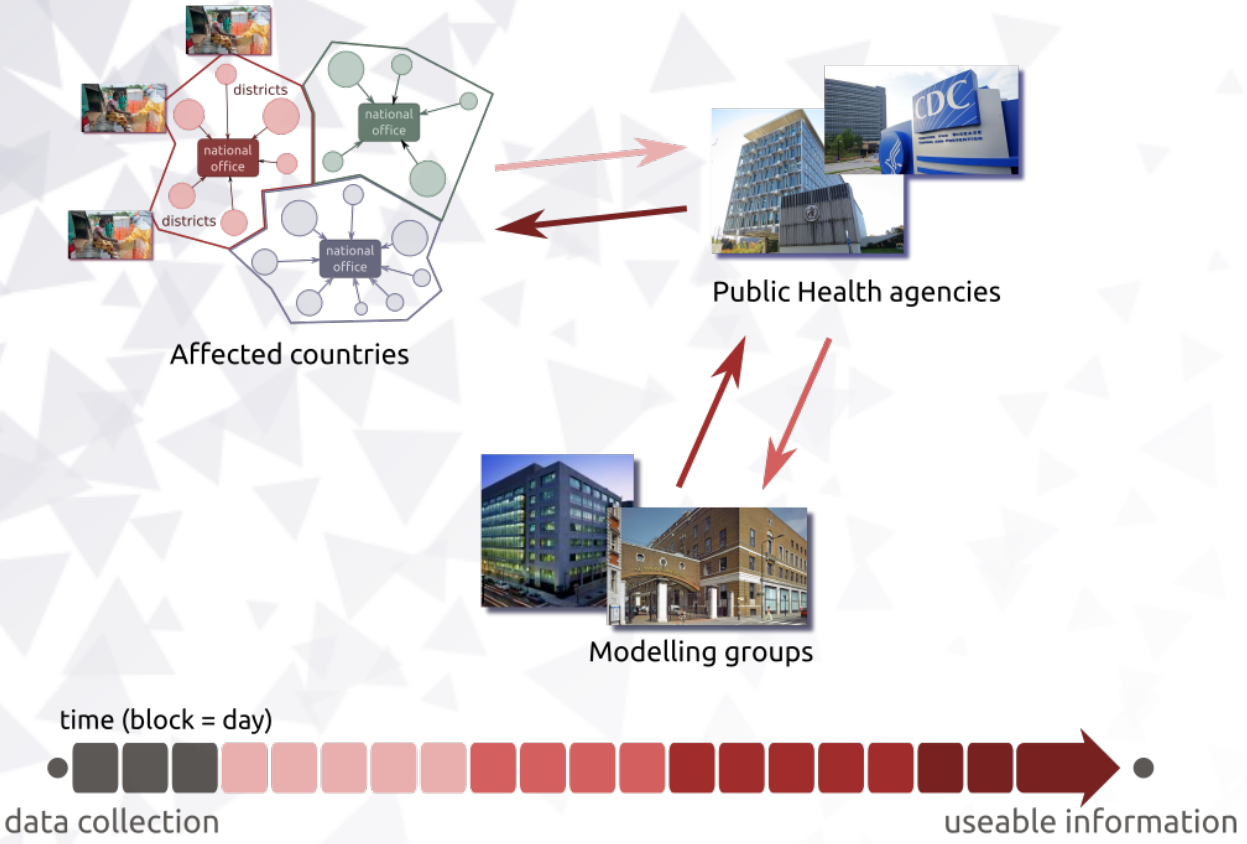
Modelling groups

time (block = day)

data collection

useable information

Centralised analyses, distributed delays



Timeliness is key: need to bring analytics to the field

Bringing analytics resources where they are needed

www.repidemicsconsortium.org

- an NGO for free, open health crisis analytics
- 100-150 subscribers, ∼30 active members

Bringing analytics resources where they are needed

RECON

www.repidemicsconsortium.org

- an NGO for free, open health crisis analytics
- 100-150 subscribers, ~30 active members
- development of free, open-source analysis tools (using R)
- 10 packages released, ~15 in development

Bringing analytics resources where they are needed

RECON

www.repidemicsconsortium.org

- an NGO for free, open health crisis analytics
- 100-150 subscribers, ~30 active members
- development of free, open-source analysis tools (using R)
- 10 packages released, ~15 in development
- short courses with partner institutions (CDC, MSF, WHO, EAN, ...)

Bringing analytics resources where they are needed

RECON

www.repidemicsconsortium.org

- an NGO for free, open health crisis analytics
- 100-150 subscribers, ~30 active members
- development of free, open-source analysis tools (using R)
- 10 packages released, ~15 in development
- short courses with partner institutions (CDC, MSF, WHO, EAN, ...)
- support field deployment

Thanks to:

- **Organisers:** J & J
- **Collaborators:** Finlay Campbell, Anne Cori, Pierre Nouvellet, Zhian Kamvar, Stephen Baker, Amrish Baidjoe, Neil Ferguson, Dan Bausch, Jimmy Whitworth, Bayard Roberts, John Edmunds
- **Groups:** WHO Ebola Likati Response Team
- **Funding:** GCRF project RECAP (ES/P010873/1), UK PH RST, HPRU-NIHR, MRC

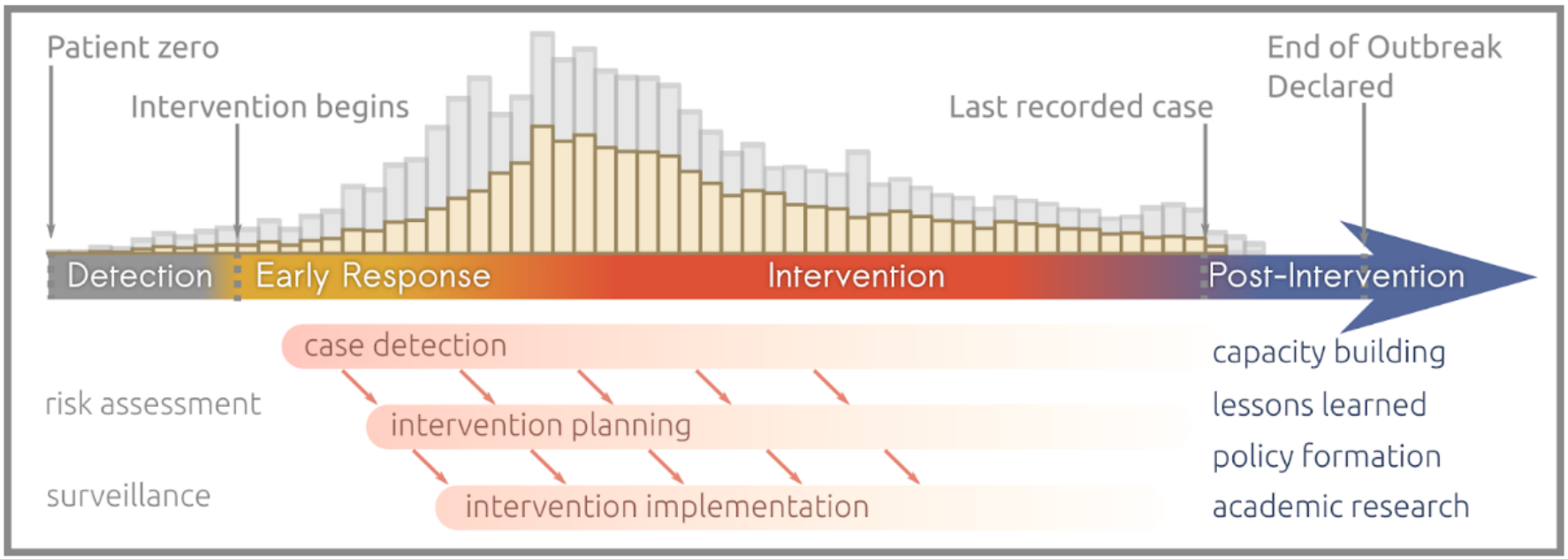
Get these slides at:

@teebzr

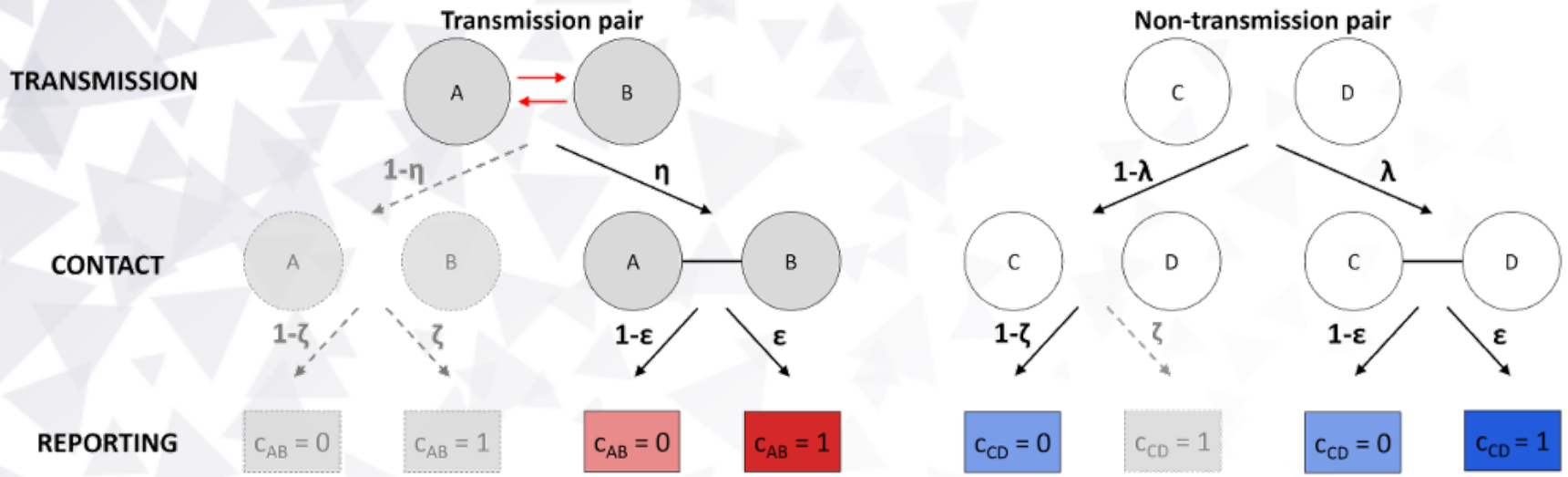
RECON

www.repidemicsconsortium.org

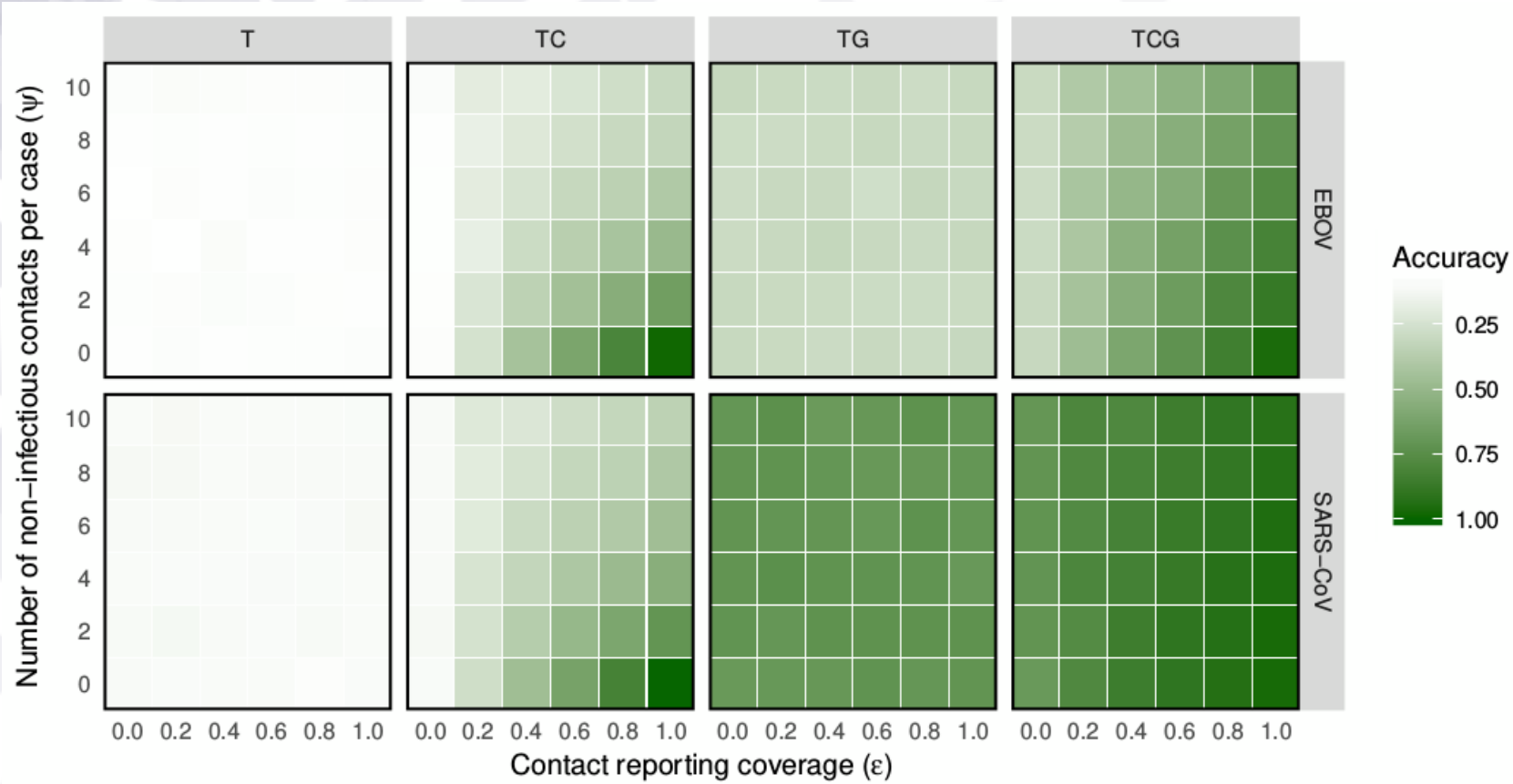
Outbreak analytics timeline



Contact model: how it works



Contact model: results (1/2)



Contact model: results (2/2)