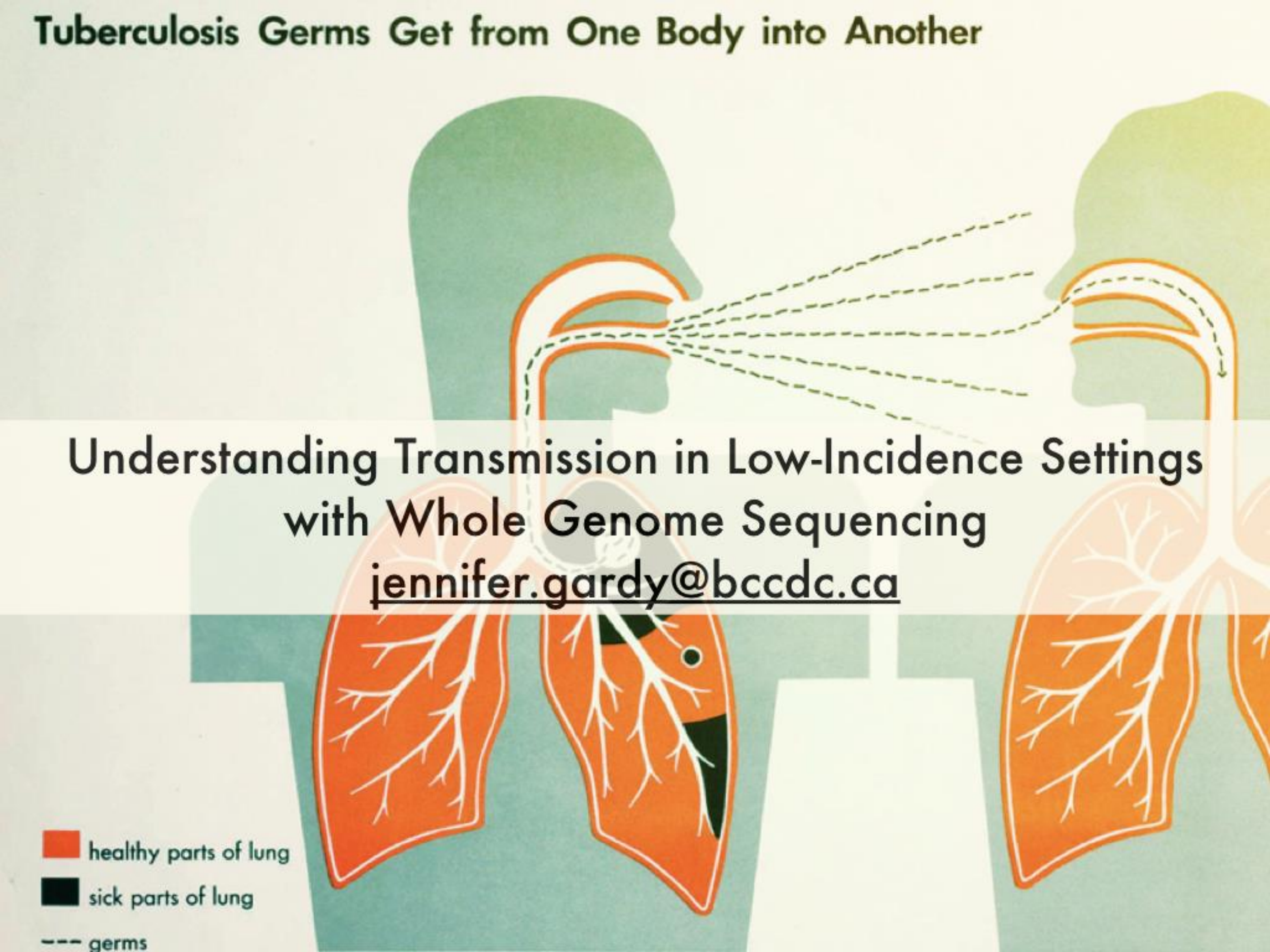
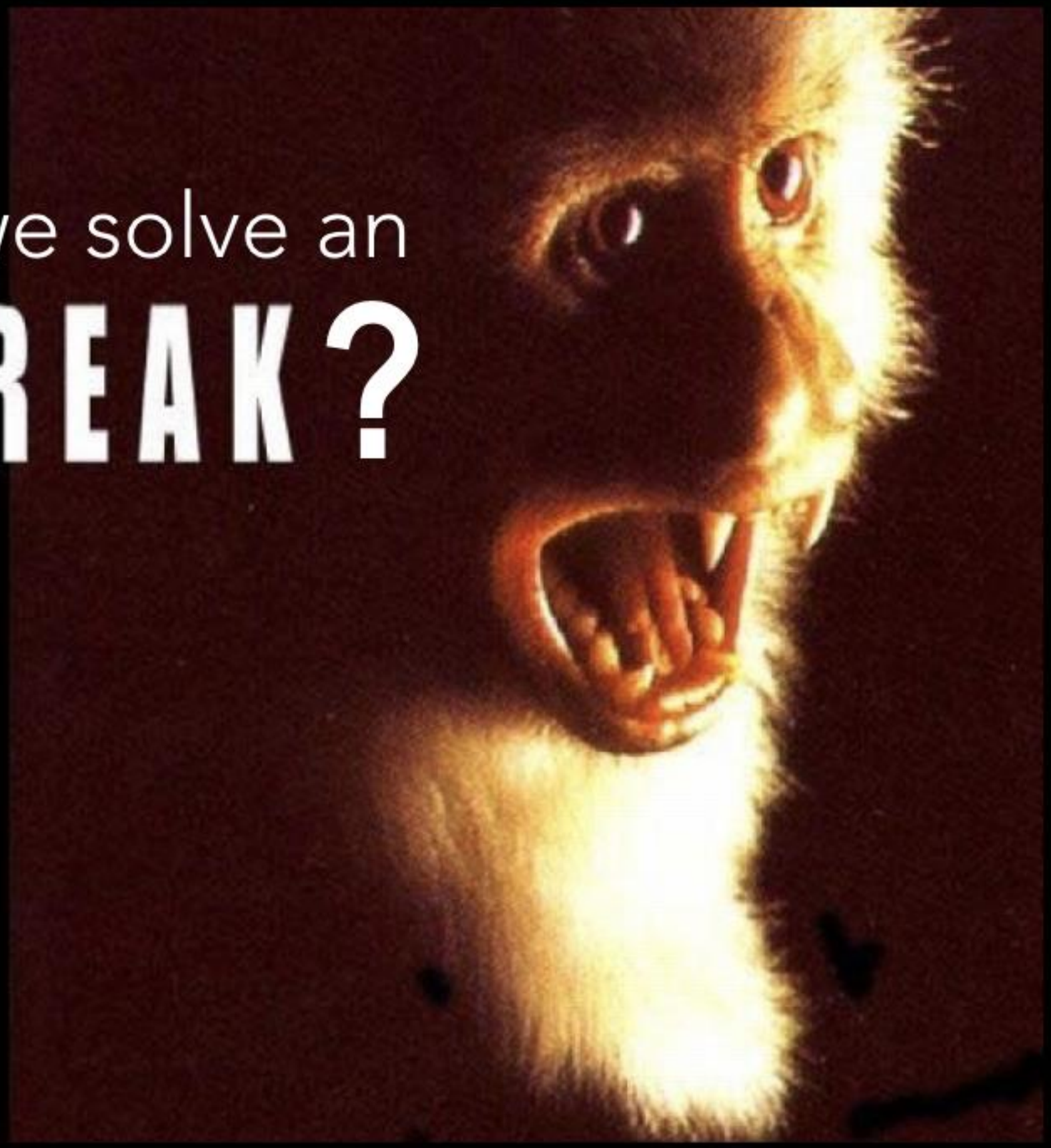


Understanding Transmission in Low-Incidence Settings with Whole Genome Sequencing

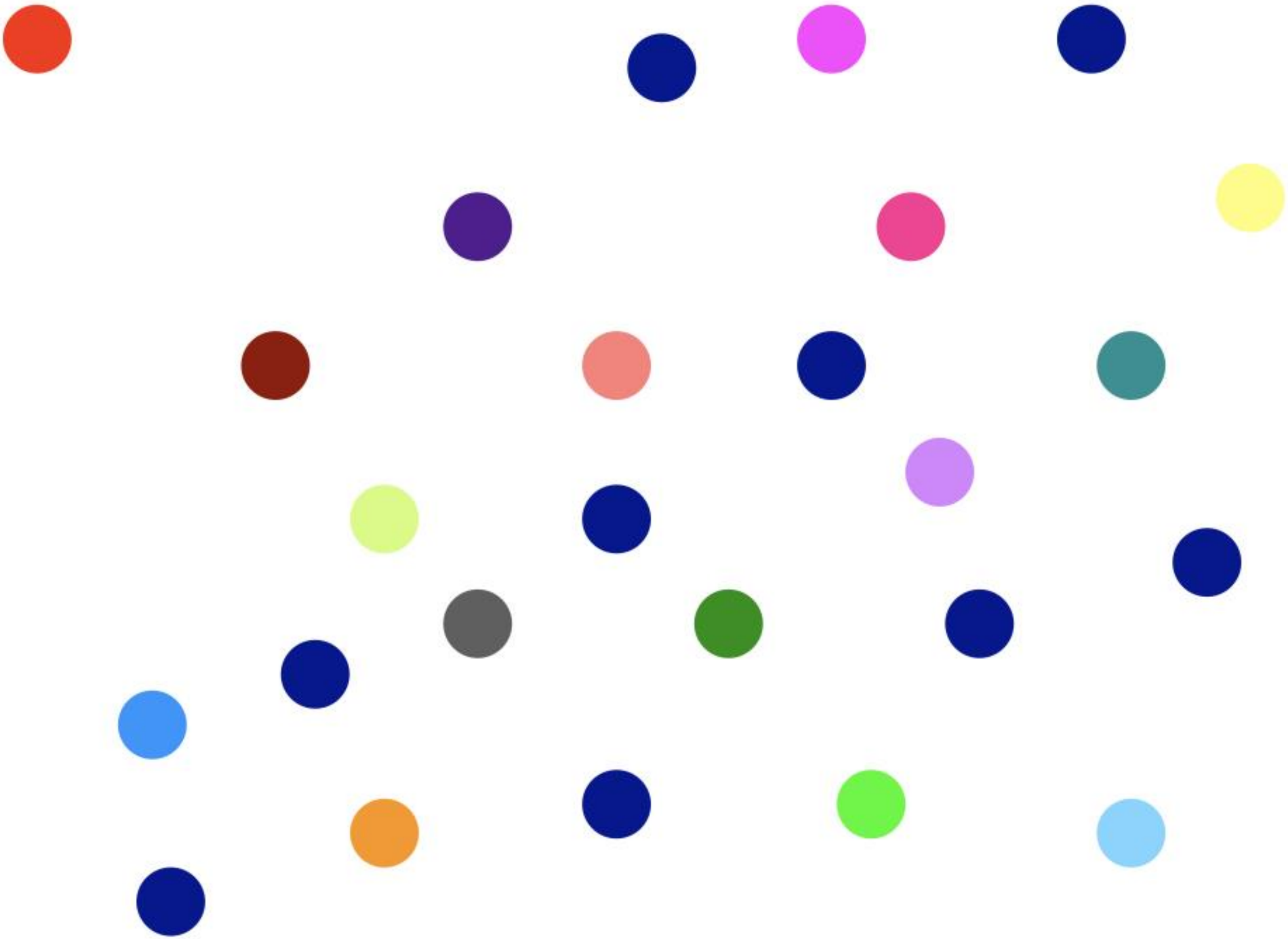
jennifer.gardy@bccdc.ca

How do we solve an

OUTBREAK?

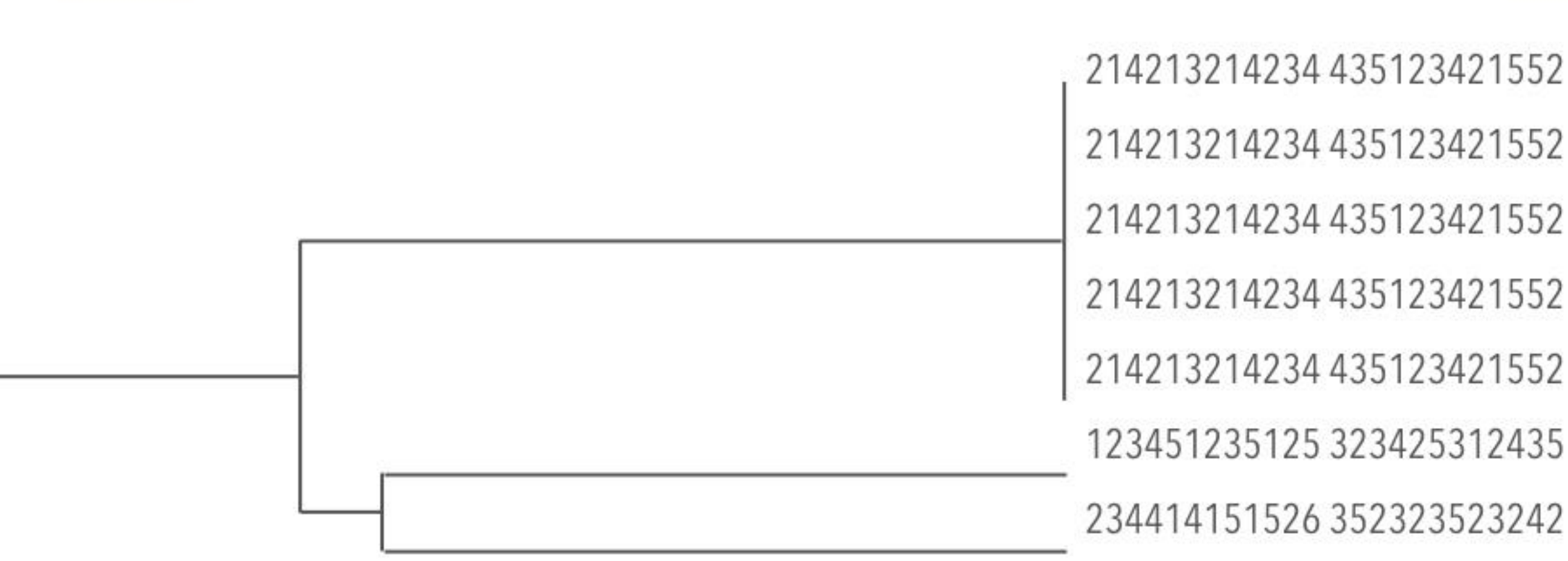


Surveillance identifies a potential outbreak

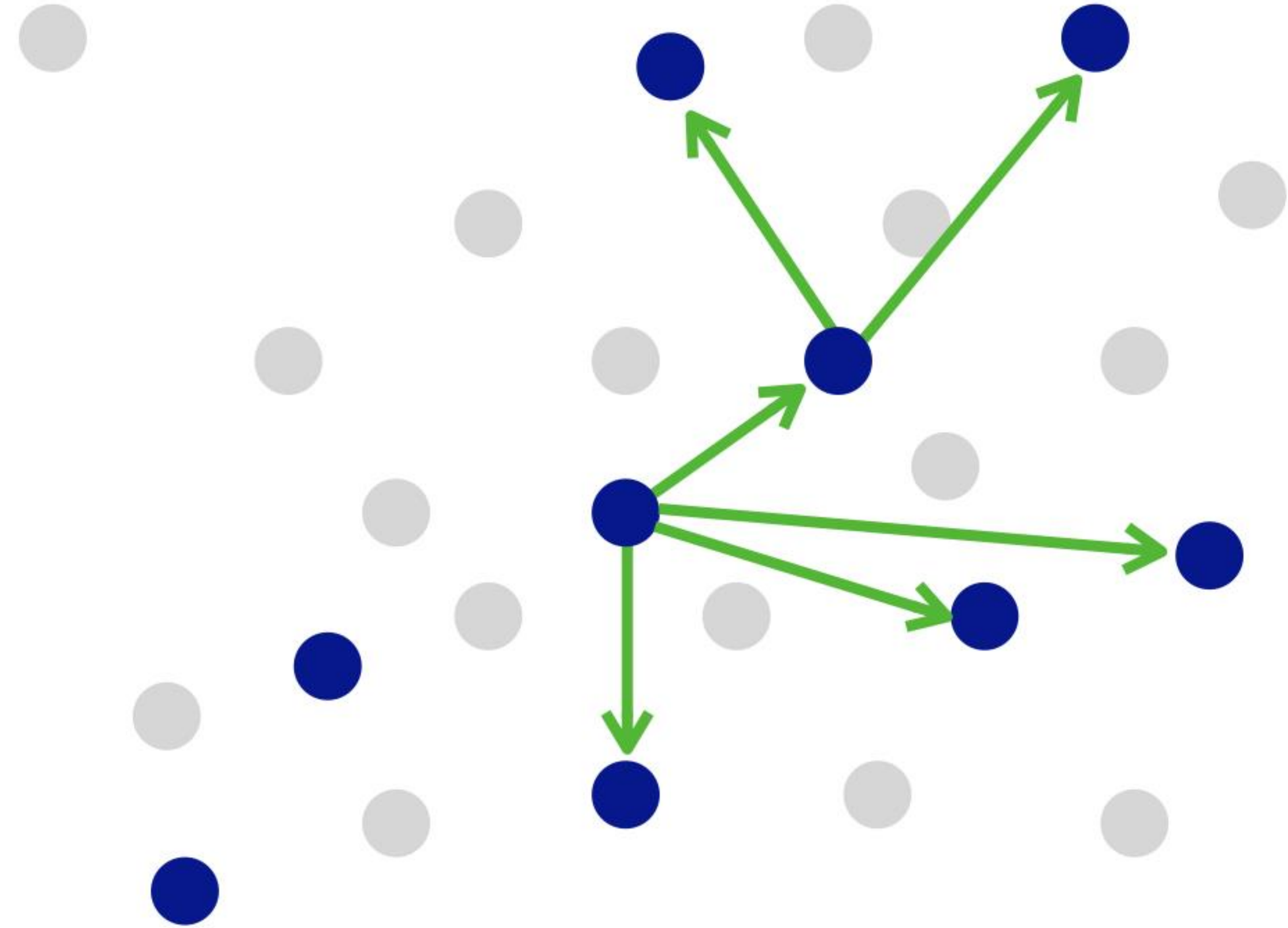


Molecular epidemiology hones in on clusters

MIRU-VNTR GENOTYPING GIVES US DNA FINGERPRINTS FOR A TB ISOLATE



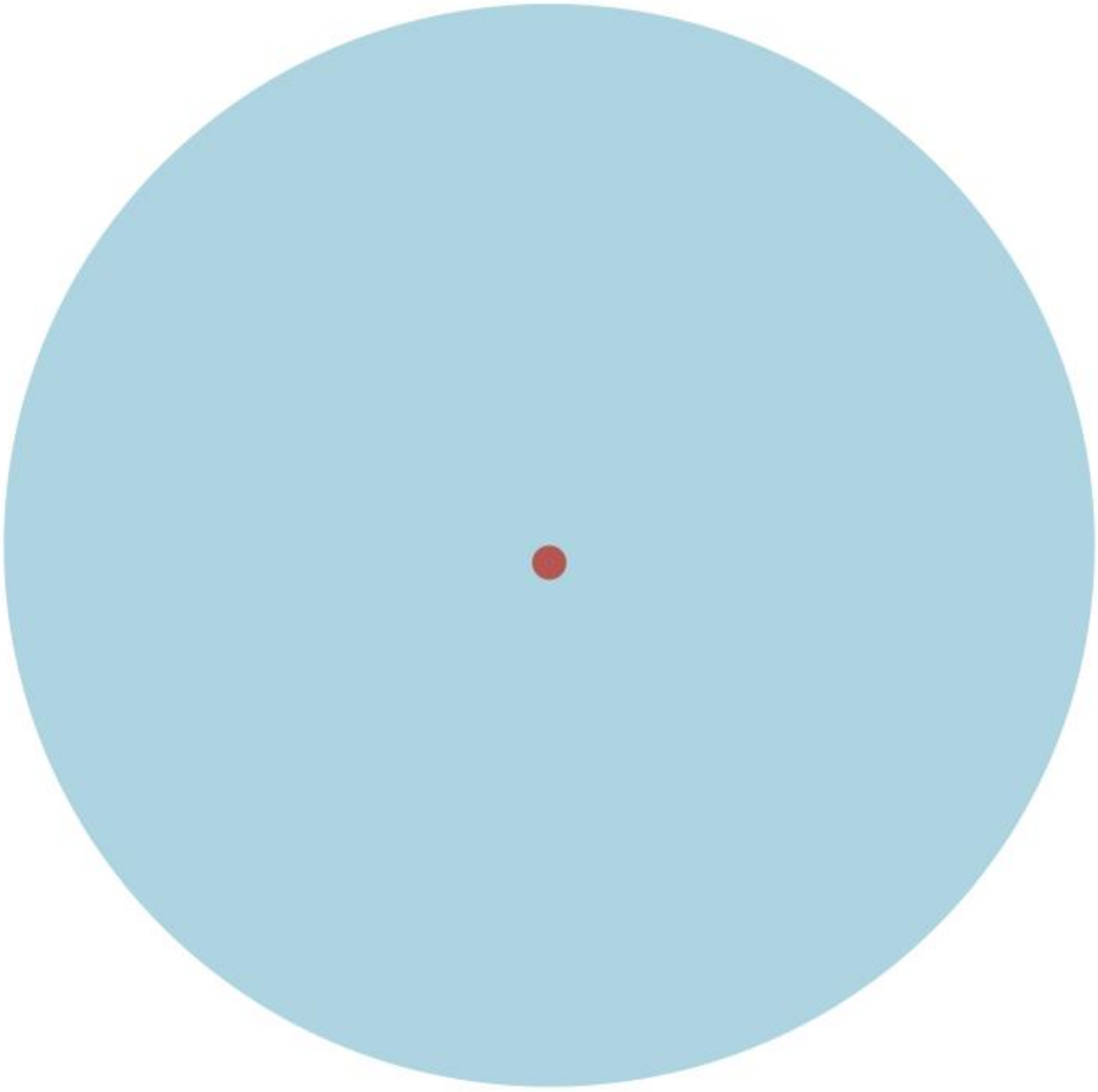
Contact investigation suggests transmissions

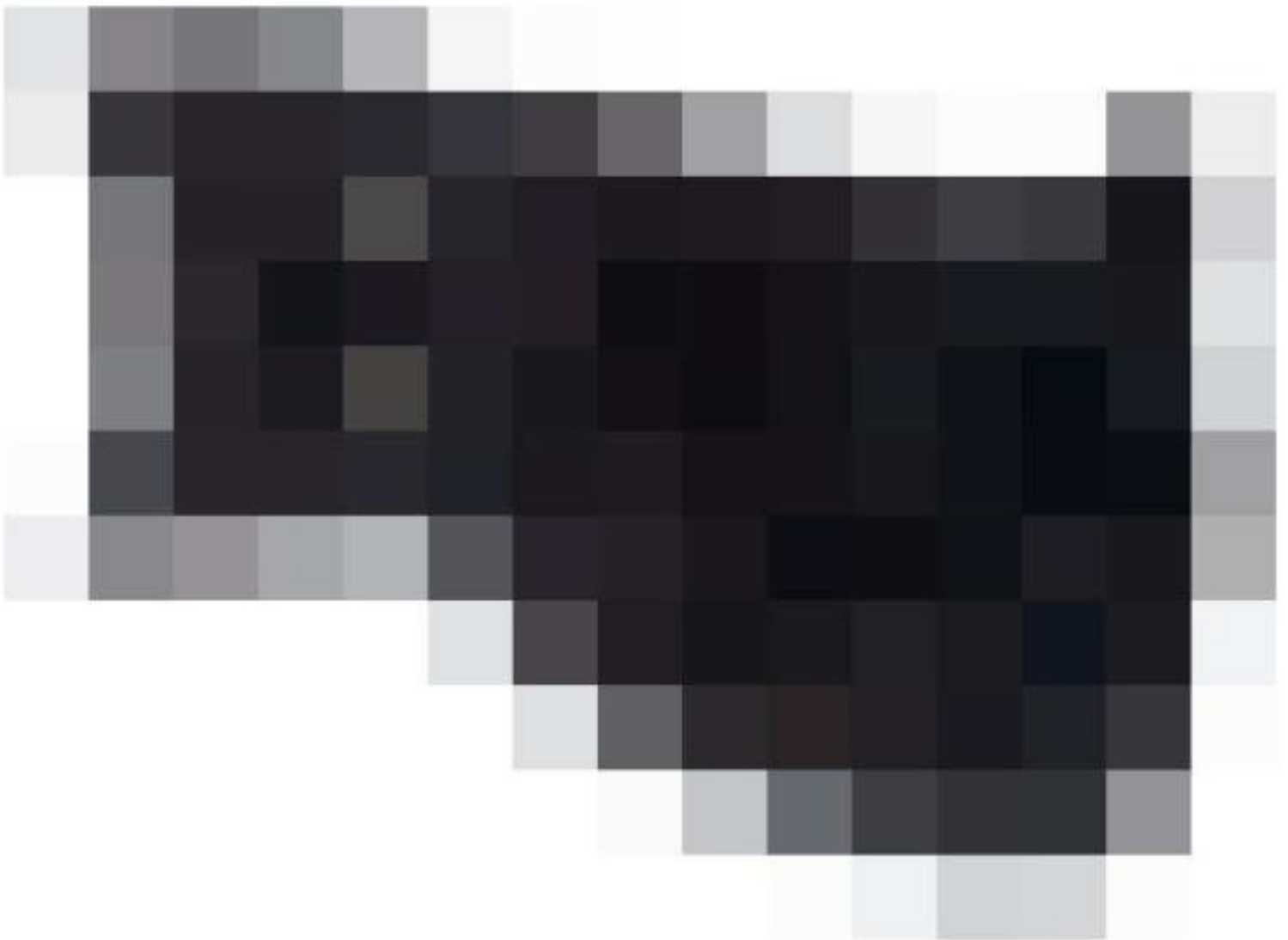


LIMITATIONS OF CURRENT METHODS

- Genotyping methods only tell you a cluster of cases exists, not the order/direction of transmission
- Size/membership of the cluster varies with the molecular typing method(s) used
- Epidemiological investigation is required to derive the links between cases, and may not be available or of sufficient quality

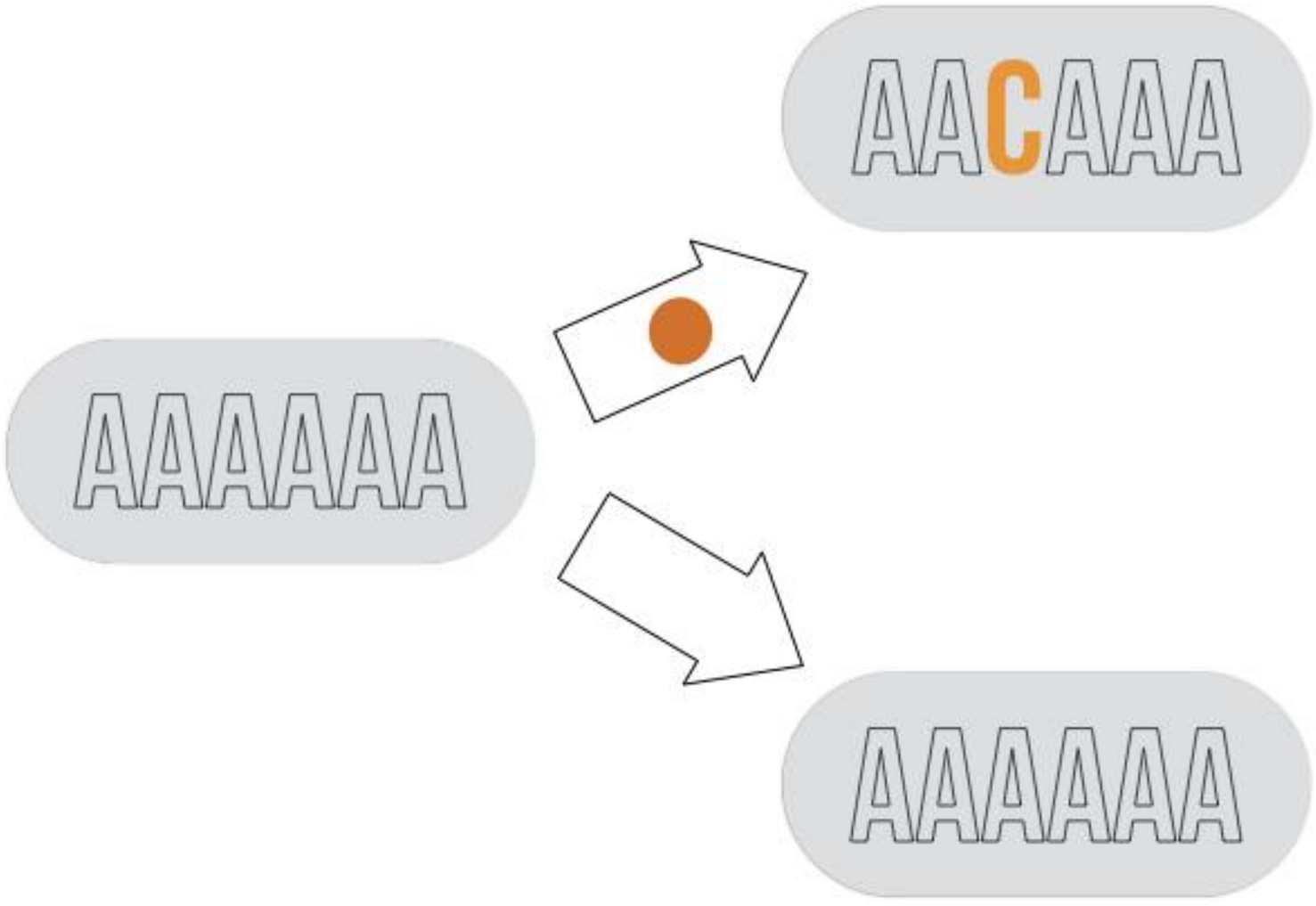
GENOMICS TO THE RESCUE





ge·no·mic ep·i·de·mi·ol·o·gy (jē ˈnōmik ˌepiˌdēmēˈäləjē/)
n. reading whole genome sequences from outbreak isolates to track person-to-person spread of an infectious disease.


AAAAAA


AACAAA
AAAAAA
AAAAAA

AAAAAA

AACAAA

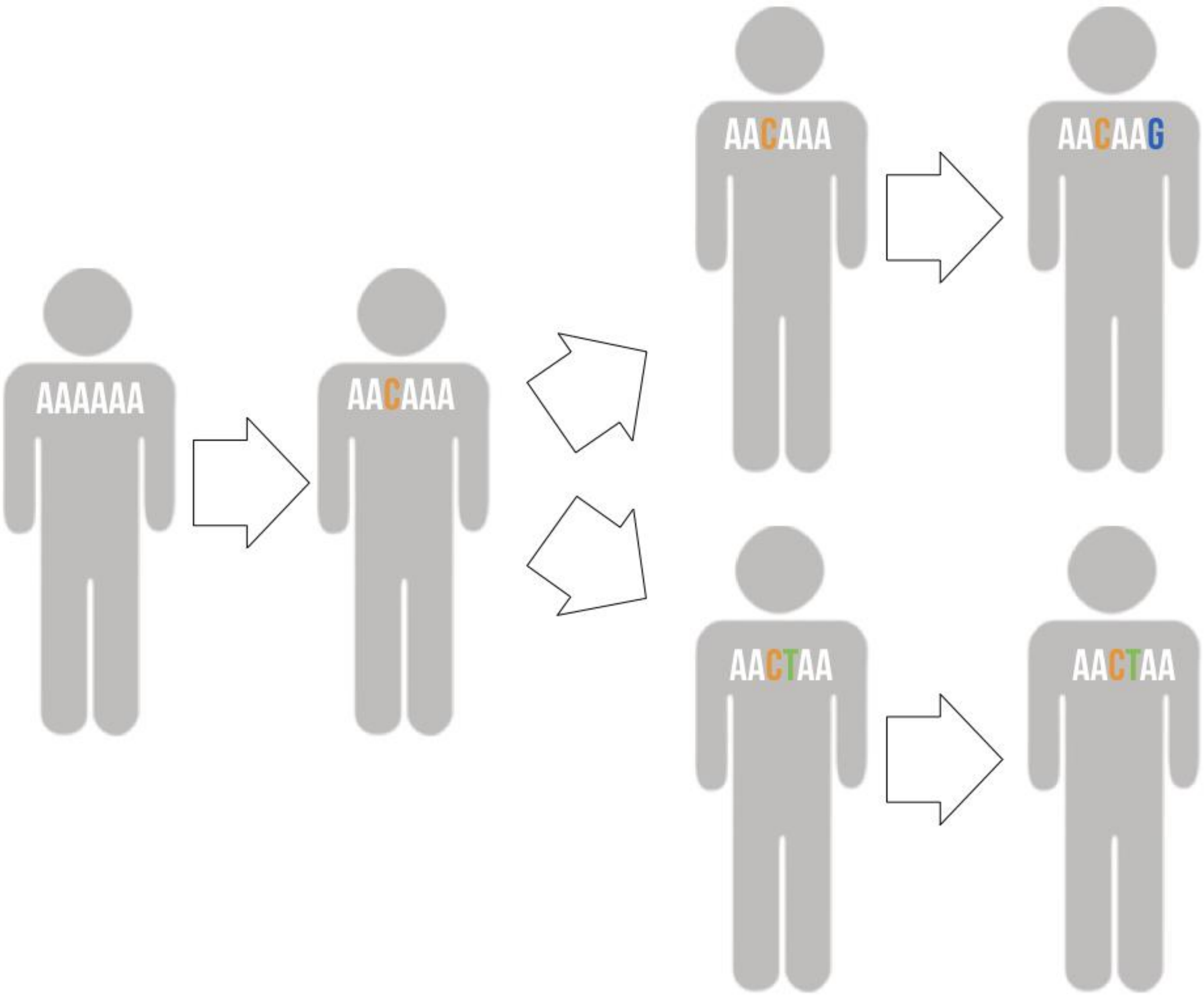
AACAAA

GACAAA

AAAAAA

AAAATA

AAAAAA


AAAAAA
AACAAA
AACAAA
AACAAG
AACTAA
AACTAA

TELEPHONE

ART BY DEVIANTART USER SCUMMY

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak

Jennifer L. Gardy, Ph.D., James C. Johnston, M.D., Shannan J. Ho Sui, Ph.D.,
Victoria J. Cook, M.D., Lena Shah, M.Sc., Elizabeth Brodkin, M.D.,
Shirley Rempel, R.N., Richard Moore, Ph.D., Yongjun Zhao, D.V.M.,
Robert Holt, Ph.D., Richard Varhol, M.Sc., Inanc Birol, Ph.D., Marcus Lem, M.D.,
Meenu K. Sharma, Ph.D., Kevin Elwood, M.D., Steven J.M. Jones, Ph.D.,
Fiona S.L. Brinkman, Ph.D., Robert C. Brunham, M.D.,
and Patrick Tang, M.D., Ph.D.

WGS for TB Outbreaks

- Sequence "all" your isolates (from LJ or 1+ day positive MGIT)
- Assemble against H37Rv (bwa mem)
- Call SNVs (samtools mpileup)
- Filter, filter, filter
- LOOK AT YOUR DATA. LOOK AT IT SOME MORE. DONE LOOKING? WHAT THE HECK, MIGHT AS WELL LOOK AGAIN.
- Incorporate the epi data

1. Get yourself sorted out with a copy of BWA, samtools, bcftools and vcftools

2. Download a TB reference genome (we use H37Rv, NC_000962.3) in fasta format, save as .fa extension

3. Index the reference (~1sec)
bwa index ref.fa

4. Align the reads (in fastq format) against the reference with bwa mem (~1.5min)
bwa mem ref.fa reads1.fastq reads2.fastq > genome.sam

5. Convert sam to bam (~30sec)
samtools view -bS genome.sam > genome.bam
(-b outputs a BAM, -S auto-detects input format)

6. Sort the BAM (~30sec)
samtools sort genome.bam -o genome.sorted
(-o writes the output to the specified file)

7. Index the sorted BAM (~1 sec)
samtools index genome.sorted

8. Call SNVs with mpileup (~5-7min)
samtools mpileup -q 30 -u -f ref.fa genome.sorted > genome.bcf -I
(-q ignores reads that mapped with a quality score below the value you specify and 30 seems like the community standard; -u is uncompressed output; -f points at the reference file you originally assembled against; -I skips indels)

9. Convert the BCF to a VCF (~1sec)
bcftools call -O v -mv genome.bcf > genome.vcf
(-O v writes output to a VCF, -m uses a multiallelic caller, -v keeps variants only, leave v off to get one line for every base pair in the reference)
228 in the current samtools is the PERFECT quality score - if you have a high enough coverage (~100x), you should get perfect, glorious SNVs with 228 scores. Anything else, get suspicious and inspect the DP4 field.
DP= total read depth at the position - if < 3, be wary. if 0, trouble.
DP4= total high quality read depth, forward ref, reverse ref, forward alt, reverse alt. A good SNV will have nothing or only 1 or 2 reads on the ref bits (first two digits), and a nicely balanced distribution across the last two digits (alt bases)

9. Filter the BCF to remove stuff in repetitive regions*** and convert to VCF (~1min)
vcftools --vcf genome.vcf --exclude-bed stufftomask.bed --recode --out genome_masked

***Use a bed format file containing the ranges of repetitive regions you want to mask. These are usually generated via self-self blast searches. If you make your own bed file and wonder why masking isn't working, it's usually because the chromosome name in the bed file and the chromosome name in the first column of the VCF don't match 100%

Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *Antimicrobial Therapeutics Reviews*

A brief primer on genomic epidemiology: lessons learned from *Mycobacterium tuberculosis*

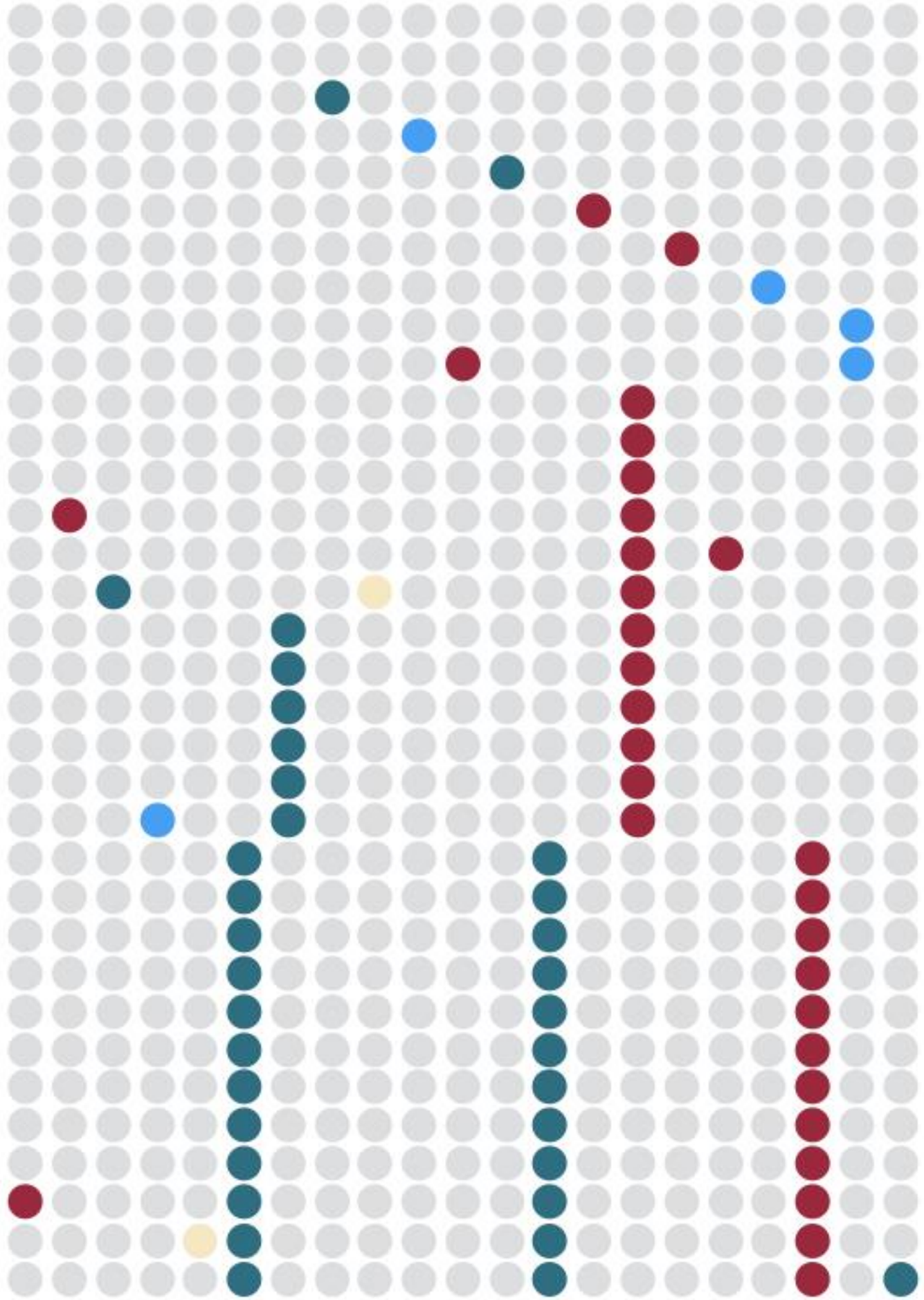
Jennifer L. Guthrie[1] and Jennifer L. Gardy[1,2]

[1]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[2]Communicable Disease Prevention and Control Services, British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada

Address for correspondence: Jennifer L. Gardy, Communicable Disease Prevention and Control Services, British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC, Canada V5Z 4R4. jennifer.gardy@bccdc.ca

Genomics is now firmly established as a technique for the investigation and reconstruction of communicable disease outbreaks, with many genomic epidemiology studies focusing on revealing transmission routes of *Mycobacterium tuberculosis*. In this primer, we introduce the basic techniques underlying transmission inference from genomic data, using illustrative examples from *M. tuberculosis* and other pathogens routinely sequenced by public health agencies. We describe the laboratory and epidemiological scenarios under which genomics may or may not be used, provide an introduction to sequencing technologies and bioinformatics approaches to identifying transmission-informative variation and resistance-associated mutations, and discuss how variation must be considered in the light of available clinical and epidemiological information to infer transmission.

Keywords: tuberculosis; transmission; genomics; resistance

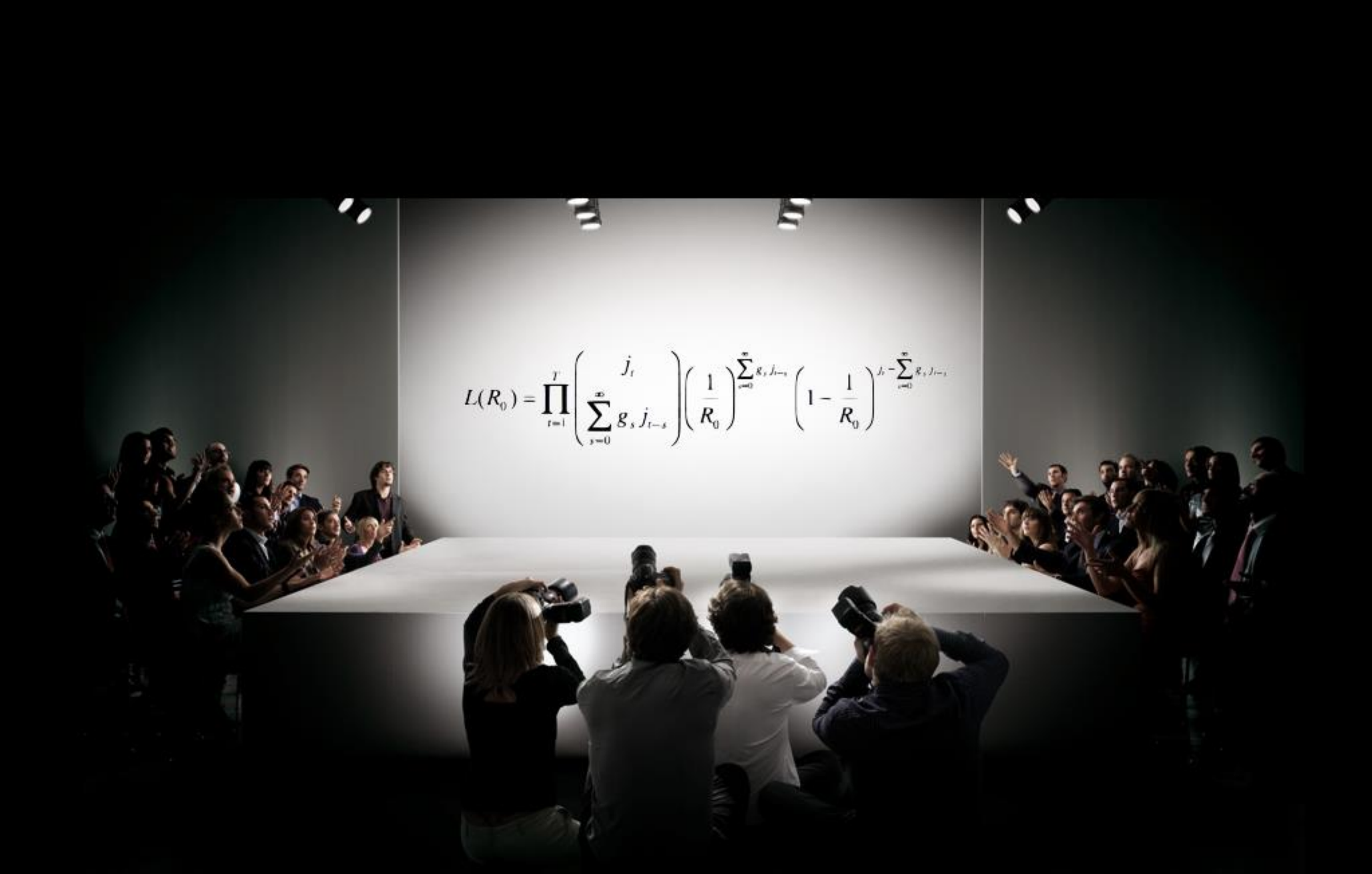


index case

second case

OH $#!T.

WE FORGOT ABOUT WITHIN-HOST GENETIC DIVERSITY



MATH MODELLING

MOLECULAR BIOLOGY AND EVOLUTION

ABOUT THIS JOURNAL · CONTACT THIS JOURNAL · SUBSCRIPTIONS · CURRENT ISSUE · ARCHIVE · SEARCH

Oxford Journals › Medicine & Health & Science & Mathematics › Molecular Biology and Evolution › Advance Access › 10.1093/molbev/msw275

Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks

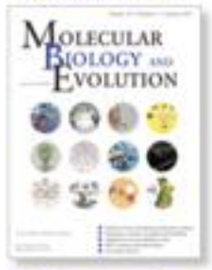
Xavier Didelot[1], Christophe Fraser[1,2], Jennifer Gardy[3,4] and Caroline Colijn[5]

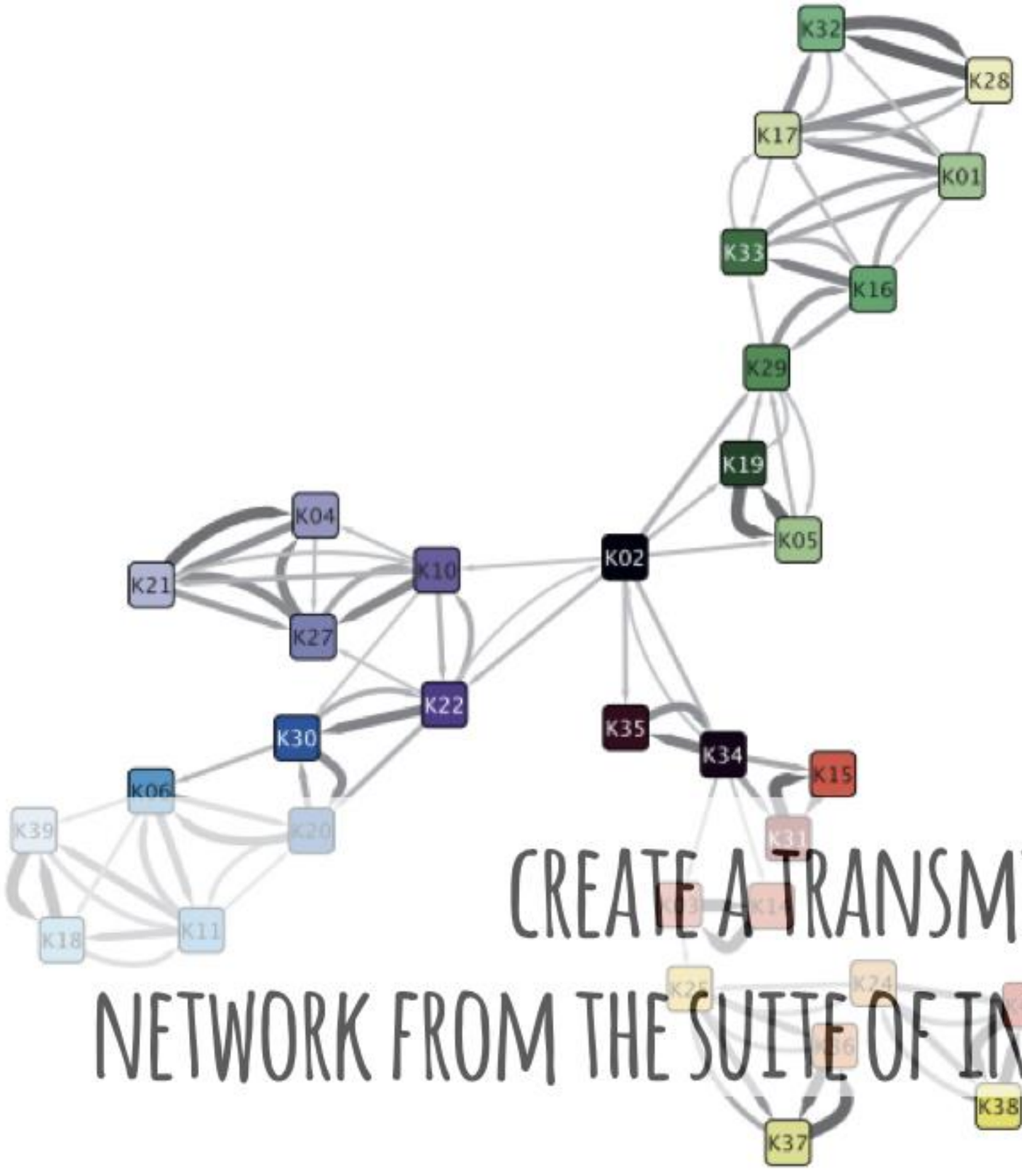
+ Author Affiliations

Received July 22, 2016.
Revision received October 28, 2016.
Accepted November 20, 2016.

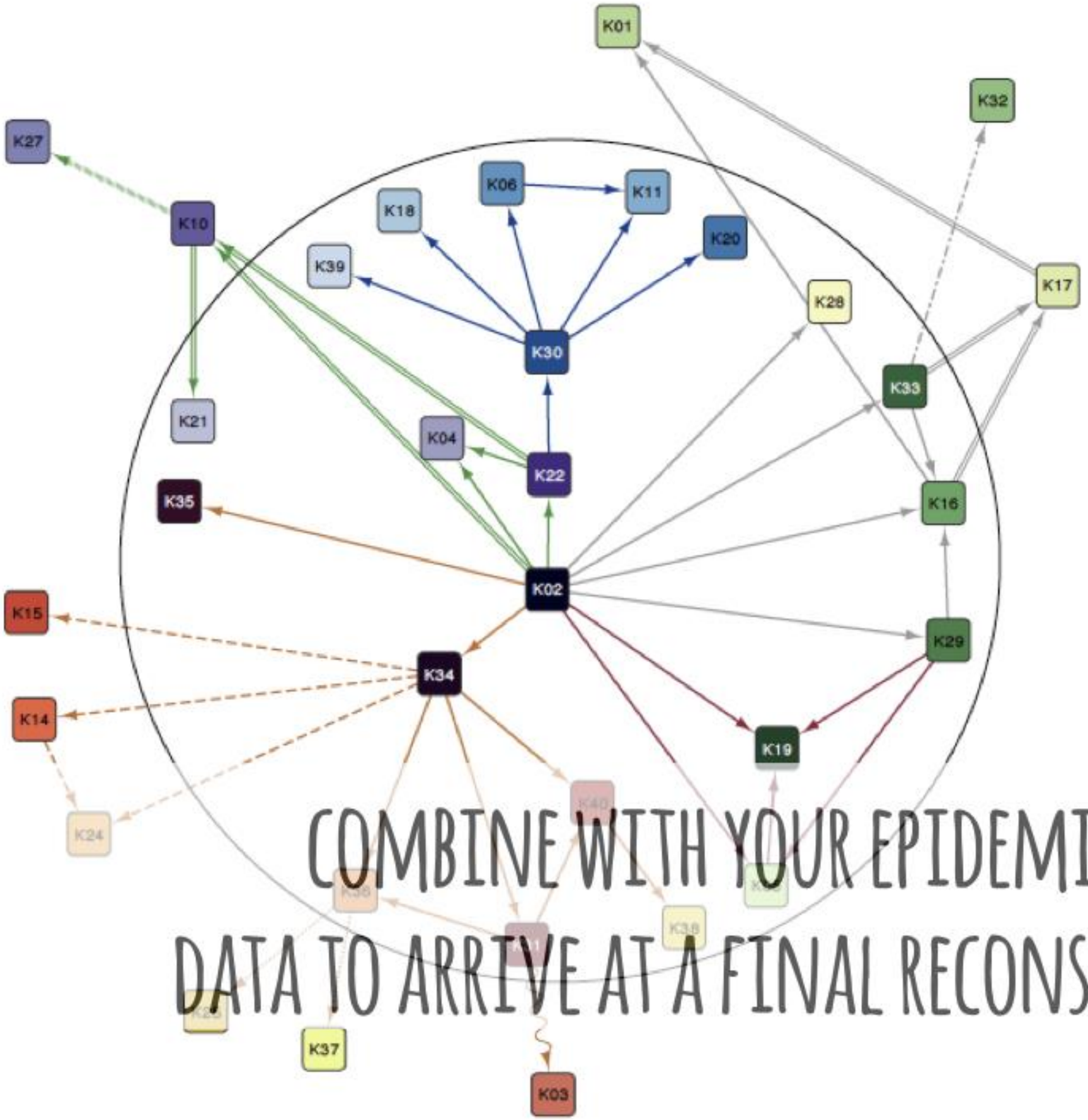
Abstract

Genomic data is increasingly being used to understand infectious disease epidemiology. Isolates from a given outbreak are sequenced, and the patterns of shared variation are used to infer which isolates within the outbreak are most closely related to each other. Unfortunately, the phylogenetic trees typically used to represent this variation are not directly informative about who infected whom – a phylogenetic tree is not a transmission tree. However, a transmission tree can be inferred from a phylogeny while accounting for within-host genetic diversity by colouring the branches of a phylogeny according to which host those branches were in. Here we extend this approach and show that it can be applied to

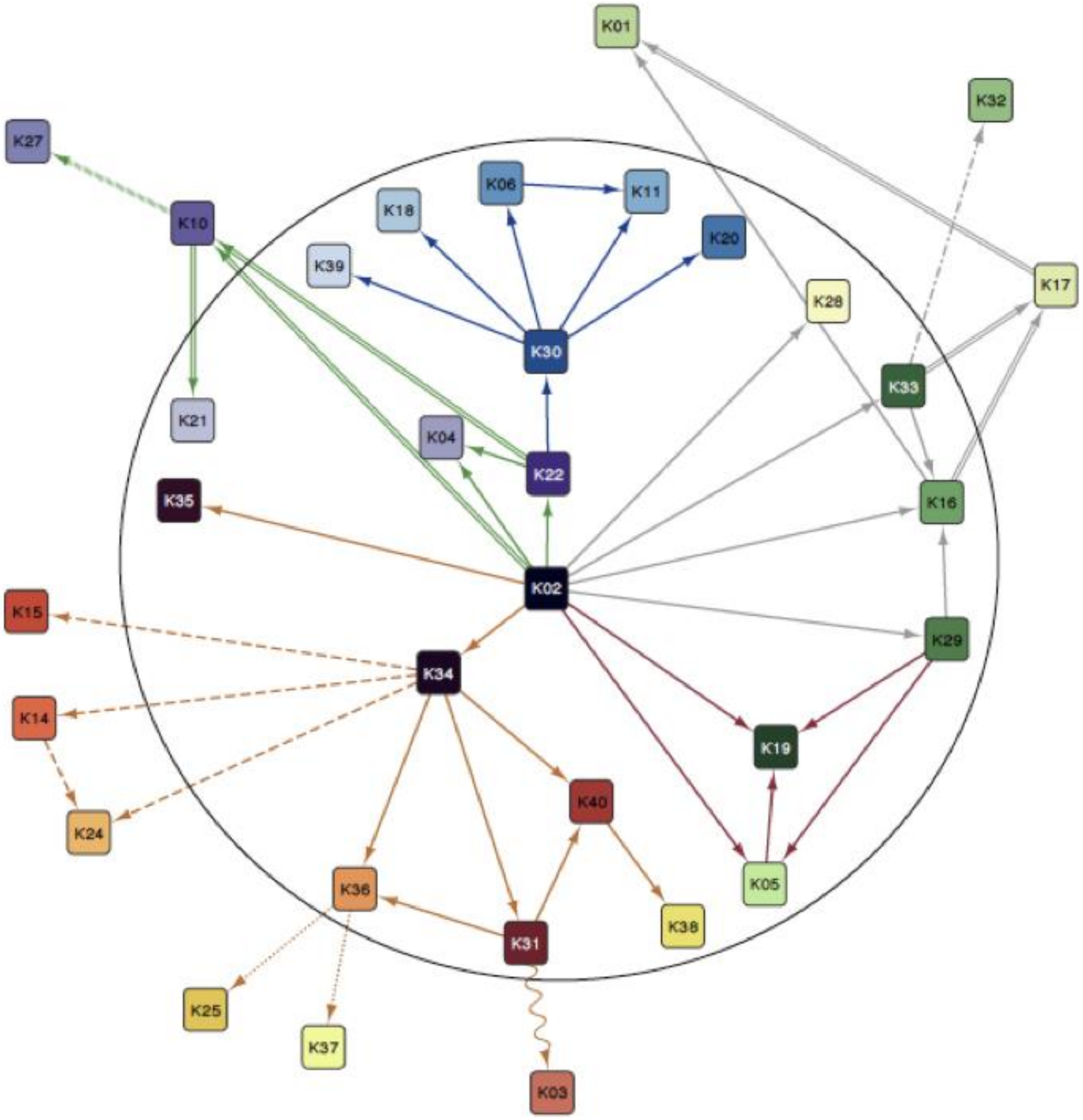




CREATE A TRANSMISSION SOCIAL NETWORK FROM THE SUITE OF INFECTION TREES



COMBINE WITH YOUR EPIDEMIOLOGICAL
DATA TO ARRIVE AT A FINAL RECONSTRUCTION


K01
K32
K27
K10
K06
K11
K18
K20
K39
K17
K28
K30
K33
K21
K04
K22
K35
K16
K02
K15
K29
K34
K14
K19
K40
K24
K05
K36
K38
K31
K25
K37
K03

Caveats

- Automated methods not quite there yet - still need good epi data
- In densely connected networks, inference might be difficult
- Not ideal for high-incidence settings
- Should have an actionable goal in mind, don't just WGS because you can
- Need a team who knows the disease, the epidemiology, and the genomics/bioinformatics

HIV Medicine (2007), 8, 382–387

ORIGINAL RESEARCH

HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission*

EJ Bernard,[1] Y Azad,[2] AM Vandamme,[3] M Weait[4] and AM Geretti[5]

[1]NAM, London, UK, [2]National AIDS Trust, London, UK, [3]Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium, [4]Research Institute for Law, Politics and Justice, Keele University, Staffordshire, UK, and [5]Department of Virology, Royal Free Hospital and Royal Free & University College Medical School, London, UK

Background

Phylogenetic analysis – the study of the genetic relatedness between HIV strains – has recently been used in criminal prosecutions as evidence of responsibility for HIV transmission. In these trials, the expert opinion of virologists has been of critical importance.

Pitfalls

Phylogenetic analysis of HIV gene sequences is complex and its findings do not achieve the levels of certainty obtained with the forensic analysis of human DNA. Although two individuals may carry HIV strains that are closely related, these will not necessarily be unique to the two parties and could extend to other persons within the same transmission network.

Acceptable standards

For forensic purposes, phylogenetic analysis should be conducted under strictly controlled conditions by laboratories with relevant expertise applying rigorous methods. It is vitally important to include the right controls, which should be epidemiologically and temporally relevant to the parties under investigation. Use of inappropriate controls can exaggerate any relatedness between the virus strains of the complainant and defendant as being strikingly unique. It will be often difficult to obtain the relevant controls. If convenient but less appropriate controls are used, interpretation of the findings should be tempered accordingly.

Conclusions

Phylogenetic analysis cannot prove that HIV transmission occurred directly between two individuals. However, it can exonerate individuals by demonstrating that the defendant carries a virus strain unrelated to that of the complainant. Expert witnesses should acknowledge the limitations of the inferences that might be made and choose the correct language in both written and verbal testimony.

Keywords: phylogenetic analysis, criminal investigation, HIV transmission, forensic investigation, molecular epidemiology

THANK YOU

BCCDC: JAMES JOHNSTON, VICTORIA COOK
UBC: JENNIFER GUTHRIE, ANAMARIA CRISAN
BCPHL: PATRICK TANG, CLARE KONG, LINDA HOANG, DANIELLE JORGENSEN, MABEL RODRIGUES
INTERIOR HEALTH: SUE POLLOCK ,LORI HISCOE, ROB PARKER, JOYCE CHEN
IMPERIAL COLLEGE: CAROLINE COLIJN, XAVIER DIDELOT, CHRISTOPHE FRASER
$$$: BCCDC FOUNDATION FOR POPULATION & PUBLIC HEALTH, GENOME BC, BC LUNG ASSOCIATION, CANADA RESEARCH CHAIRS, MICHAEL SMITH FOUNDATION FOR HEALTH RESEARCH