

**université
PARIS-SACLAY**

Up-to-date methods to study bacterial and/or viral outbreaks

Evolutionary genomics, Genotyping, Bacterial Population structures and Phylodynamics

C. Sola, Feb 7th 2024

1

SCHEDULE


- Diapos 4 to 66 = Course
- Diapos 67 to end = some clues on my research topic

« Does my dataset contains temporal and evolutionary signal ? »

Study : Space, Time, Genomic changes and Genomic Interactions between species...i.e. ecological changes

2


Hunter Gatherer Health (in particular : skeletal health) was better than early Neolithic people



L2 scientific english 2024, Jan. 10th 2024

4

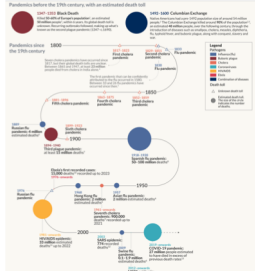
Animal domestication and exchange with livestock together with urban phenomenon triggered infectious diseases expansion



L2 scientific english 2024, Jan. 10th 2024

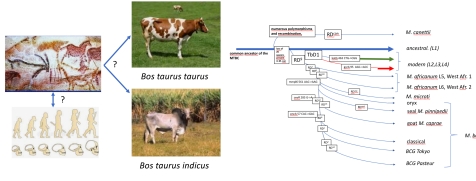
4

Infectious Diseases were and still are the main mass killers before weapons of mass destruction



L2 scientific english 2024, Jan. 10th 2024 <https://ourworldindata.org/historical-pandemic>

Origin of tuberculosis remains uncertain




Sola et al. 2024, Genetics and Evolution of Infectious Diseases. 3rd Edition - Jun 2024
Chapter 27: Evolution, phylogenetics and phylogeography of *Mycobacterium tuberculosis*.

L2 scientific english 2024, Jan. 10th 2024


6

Definitions




AN EPIDEMIC

It is a disease that affects a large number of people within a community, population, or region.




A PANDEMIC

It is an epidemic that's spread over multiple countries or continents, population, or region.



AN ENDEMIC

It is something that belongs to a particular people or country.



AN OUTBREAK

It is a greater-than-anticipated increase in the number of epidemic cases. It can also be a single case in a new area.

Note: If it's not quickly controlled, an outbreak can become an epidemic.

L2 scientific english 2024, Jan. 10th 2024

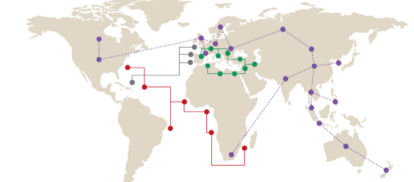
Today's Pandemia have nothing to do with historical pandemia

Post-Columbus contact and European colonization: smallpox, measles and other diseases (fifteenth to eighteenth century)

Classical antiquity trade and war: Antonine plague (second century), Plague of Cyprus (third century), Justinian plague (sixth century)

Transatlantic slave trade and European colonization: P. falciparum malaria (sixteenth to nineteenth century)

International air travel: SARS epidemic (2002-2004)



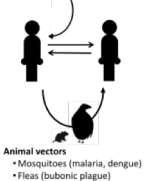
L2 scientific english 2024, Jan. 10th 2024

Mode of Infection

A. General Transmission

Abiotic environmental factors

- Wind
- Water
- Inhalation of spores
- Entry into skin



Animal vectors

- Mosquitoes (malaria, dengue)
- Fleas (bubonic plague)

B. Human to Human Transmission

Direct Contact

- Pathogen survives best inside the body
- Eg: HIV, Herpesviruses, Ebola

Indirect Contact

- Pathogen survives harsh environment
- Pick up pathogen from surface or air
- Eg: influenza, norovirus

Droplets

- Pathogens are in droplets, but do not survive long this way
- Eg: Ebola, Bordetella pertussis

Airborne

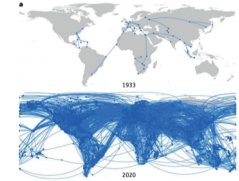
- Pathogens aerosolized and stay infective
- Eg: Influenza, Tuberculosis

Fecal - Oral

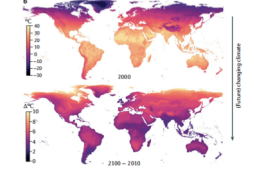
- Through contaminated water or food
- Eg: Cholera, Norovirus, Shigella

L2 scientific english 2024, Jan. 10th 2024

Increased Networks and Trade, Temperature, Population, Migration, Ecological changes... are promoting factors of new emerging diseases



The global international air travel network



Average monthly maximum temperature in 1970-2000 and difference between 2070-2100 and 1970-2000 averages

L2 scientific english 2024, Jan. 10th 2024

Outbreak Definition How can we detect outbreaks ?

Experts

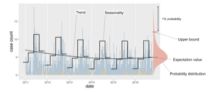
CDC 1981 : AIDS

Acquired Immune Deficiency Syndrome (AIDS) was first recognized as a new disease in 1981 when increasing numbers of young homosexual men succumbed to unusual opportunistic infections and rare malignancies

Automated Outbreak Detection

"Are there too many cases, here and now, compared with expectations?"


One standard approach: Univariate time series + Regression + Confidence Interval



For example: `famrgoutofusable` (from R package `surveillance`), used here for benchmarking

Nishiura et al. (2020) *Statistics in Medicine* 39(7): 1788-1800. doi:10.1002/sim.8488
Sullivan et al. (2005) *Journal of Statistical Software* 19(1): 1-16. doi:10.18637/jss.v019.i01


How infectious Diseases spread : Modeling Epidemiology



<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology


Case contact
(patient 0)




<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

Cluster apparition







<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology


Cluster Development





<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology




Mathematical Modeling
allowed to
prevent Emergency wards
collapse
during Covid19 epidemics

<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

R_0
(the Reproductive Rate)
is a characteristic
of the pathogen

Measles +++
Tuberculosis +



Mathematical Modeling
allows to predict
percentage of people to
be vaccinated to obtain


Herd immunity

<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

SIR Model

Susceptible
Infected
Recovered



<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

How infectious Diseases spread : Modeling Epidemiology

<https://media.hhmi.org/biointeractive/click/modeling-disease-spread/introduction.html>

Today's methods of outbreak study

Patient Cluster Analysis
Social network Analysis
Digital Epidemiology

'Close contact detector' can be accessed via three of the most popular mobile social and payment apps in China, namely Alipay, WeChat and QQ

#Tous AntiCovid

Pathogen Genome Cluster Analysis
Genome Analysis
International Databases and GIS

Developed by Lauren Gardner (an epidemiologist) and her team from the John Hopkins U CSSE (Center for Systems Science and Engineering)

Outbreak Prevention Methods

Break the Chain | 5 steps of infection prevention

Outbreak Preparedness and Response

1 Stop at the source

Infection Control Measures

3 Interrupt mode of transportation

4 Protect portal of entry

VACCINATION

5 Increase host's defenses

Chain of Transmission

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

EXPERT COMMENTARY

The Next Pandemic: Prepare for "Disease X"

Kenneth V. Iserson, MD, MBA The University of Arizona, Department of Emergency Medicine, Tucson, Arizona

Section Editor: Mark I. Langford, MD, MPH/PE
 Submission history: Submitted May 14, 2020; Accepted May 14, 2020
 Electronically published June 8, 2020
 Full text available through open access at http://escholarship.uo.uciem_westjem
 DOI: 10.5811/westjem.2020.5.48215
 [West J Emerg Med. 2020;21(4):756-758.]

We are living the last hours of pDST* *phenotypic drug susceptibility testing

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 OCTOBER 11, 2018 VOL. 379 NO. 15

Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing

The CRyPTIC Consortium and the 100,000 Genomes Project

25

OMICS yes ! everywhere ? no !

MICROBIAL GENOMICS

RESEARCH ARTICLE
Bainomugisa et al., Microbial Genomics 2018,4
DOI: 10.1099/mgen.0.000188

MICROBIOLOGY
SOCIETY

DATA MICROBIOLOGY

A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions

Arnold Bainomugisa,^{1,2*} Tania Duarte,¹ Evelyn Lavu,³ Sushil Pandey,⁴ Chris Coulter,⁴ Ben J. Marais⁵ and Lachlan M. Coin¹

-WGS cost : **200 Euros** (going down though... **40 Euros** in 2022)

-A 50-Plex hybridization bead-based surveillance assay : **10-20 Euros**

WGS still difficult to implement outside EU, however implementation in progress in Africa, South-Africa, Ghana Madagascar

26

What is an outbreak ?

« **A sudden increase of the incidence of a disease** »

Definition: Endemia, epidemia/outbreak, pandemia

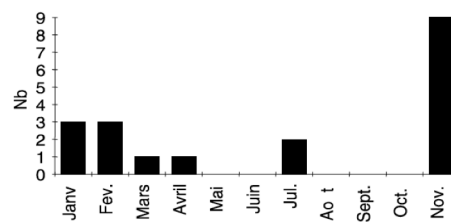
In 2008 "the plane is the key factor in the spread (of epidemics) worldwide. (...) The lines on which there are large passenger flows create preferential paths for the disease. SARS (2003) arrived in France and Canada by flights from Hong Kong ».

Goal of epidemiological surveillance: **anticipate, forecast, limit the spread**
Goal of Government : **take actions (test, isolate, trace, vaccinate, cure)**

27

An example

Exemple 1: Hemoculture positive to *Klebsiella pneumoniae* :



Question: Outbreak ?

Tiré du cours de Dominique Blanc, MER PD Service de médecine préventive hospitalière
CHUV pdf hb_epidem-moi2007.pdf disponible sur le web.

28

What is epidemiology ?

- « *Epidemiology is a reasoning and a method specific to perform objective work in medicine and in other health sciences, applied to the description of health phenomena, to the explanation of their etiology, and to the research of methods of most effective intervention* » (Milos Jenicek)
- Study of links between the diseases and some factors that may have an influence on their frequency, their distribution, their evolution
- **Descriptive, Analytical, Evaluative, Theoretical, Modeling**

29

What is molecular epidemiology ?

- *It is the indirect study of epidemics by the phenotypic, genotypic, genomic characterization of the diversity of populations of pathogens, and through the study of the dynamics of spatio-temporal distribution of their genotypes or phenotypes or genomes, and their variability, as characterized by the most effective methods, **in association** with the characteristics of the patients (demographic, medical, genetic, anthropological, geographic, social, etc.)*
(C. Sola)

30

Other definition (broader)

- “The application of molecular biology to the answering of epidemiological questions. The examination of patterns of changes in DNA to implicate particular carcinogens and the use of molecular markers to predict which individuals are at highest risk for a disease are common examples”

<http://www.termsscience.fr/>

31

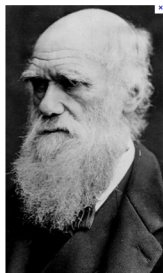
Taxon (OTU), Speciation, molecular evolution species definition in microbiology...

- **Speciation** = the evolutionary process by which new living species emerge
- **Taxon** = conceptual entity which is supposed to group together all living organisms having in common certain well-defined taxonomic or diagnostic characters
- **Molecular Evolution** : evolution of genetic material via its own mechanisms which are at the origin of the creation of genetic novelties

32

Selection : Darwinian ? Neutralist ?

« the overwhelming majority of genetic mutations that we accumulate in our genome are neutral »



C. Darwin



M. Kimura

<http://du-cote-de-chez-elysia-chlorotica.blogspot.fr>

33

Population Structure Panmixia or clonal ?

- **Panmixic** : an individual belonging to a panmictic population has as much chance of reproducing with any other individual of this population. (*Neisseria meningitidis*)
- **Clone** : group of cells, or organism that is descended from and genetically identical to a single common ancestor, such as a bacterial colony whose members arose from a single original cell. (*Mycobacterium tuberculosis*)

34

Lateral Gene transfer



Michael Syvanen



Papers By
Dr. Michael Syvanen on
Horizontal Gene Transfer
<http://www.yvm.net/vme/ht/>

« horizontal gene transfer exists, it has real biological importance, and it is a process that has shaped the history of evolution from the start of life on Earth »

K. Ochiai, T. Yamanaka, K. Kimura et O. Sawada, « Inheritance of drug resistance (and its transfer) between *Shigella* strains and Between *Shigella* and *E. coli* strains », Hihon Iji Shimpou, vol. 1861, 1959
Michael Syvanen, « Cross-species Gene Transfer: Implications for a New Theory of Evolution », *J. Theor. Biol.*, vol. 112, p. 333-343, 1985.

35

Different Phases of an epidemiological study

1. **Selection of populations**
 - Convenience sample,
 - statistical study design
 - Duration and places of sampling,
 - inclusion criteria
2. **Study Methods**
 - Preparation of material,
 - distribution,
 - choice of genetic methods,
 - production of results
3. **Statistical Analysis of Data**
 - Type of analysis (global, sequential)
 - detailed study of the clusters (by location, by year)
4. **Interpretation**
 - Calculation of recent transmission rate,
 - Data-mining of big datasets and GWAS studies (pangenome)
 - intervention epidemiology

36

Phenetic/Genetic/Genomic Methods

• Enzymatic/Protein Polymorphism

[Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics](#), Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS, 1986

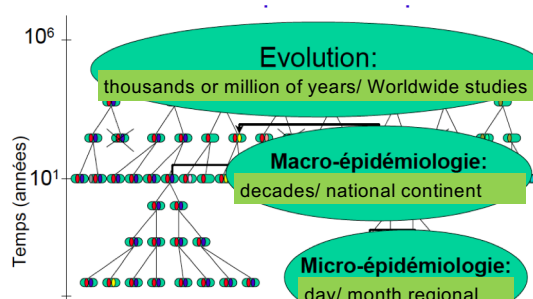
- Biochemistry (biotyping)
- Resistance to antibiotics (antibiogram/DST)
- Antibody-based (serotyping)
- Resistance to phages (lysotyping)

• Genetic Polymorphism

- Repeated Sequences (IS, VNTR-MLVA, CRISPR)
- MLST scheme (multi-locus sequence typing)
- WGS/NGS, SNPs analysis, and Metagenomics

37

How to choose one method -1 ?



38

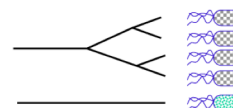
How to choose one method -2 ?

- Typability, Reproducibility, Discriminatory power (>0,95)
Index of Hunter & Gaston- (J Clin. Microbiol. 1988)
- Economical and practical criteria
Exemple = MLVA or Spoligotyping ?
- Others
 - Versatility, Rapidity, Accessibility, Facility of use, level of expertises, downstream expertise needed, etc...

39

Interpretation of Typing results: the paradox

- two isolates of the same type are not always identical (only a fraction of genomes is analyzed) we say that they are « **undistinguishable** »
- Within an epidemiological investigation context, two isolates with identical profiles come from a common ancestor by a same **transmission chain**. We are supposing that they are epidemiologically linked. The strength of the conclusion depends on the **discriminatory power** and on the **stability of markers**



40

Performance Criteria

Discriminatory Power= ability to differentiate between isolates

Comparison between different methods.

Number of types/ number of tested isolates

Discriminatory Index= probability that the method will assign to different types to two independent isolates

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^N a_j$$

a_j = nombre d'isolats qui sont du même type que l'isolat j

Indice de diversité de Simpson (écologie des populations) repris par Hunter et Gaston pour la caractérisation des méthodes de typage en épidémiologie. 41

Numerical or analogical Typing ?

- **Analogic**
 - Bands analysis, variations of migration, physical variation (temp, buffer, agarose, etc...)
- **Numerical**
 - Generation of figures
 - Necessity of coding : information can be raw or interpreted
 - Standardization
 - SNPs

42

Some *historical* ancient methods


- **PFGE= pulse-field gel electrophoresis**
Staphylococcus
- **REA= restriction Enzyme Analysis**
- **RFLP= Restriction Fragment Length Polymorphism**
IS901-RFLP = *Mycobacterium avium* and *paratuberculosis*
IS6110-RFLP = *Mycobacterium tuberculosis*

MLST= Multi-Locus sequence Typing
(detect alleles in 7 house keeping genes)

43

Exemple: spoligotyping in *M. tuberculosis* (CRISPR typing)

- binary coding
- octal coding
- Attribution of « shared-types » or « spoligo-international types »
- creation of databases



44

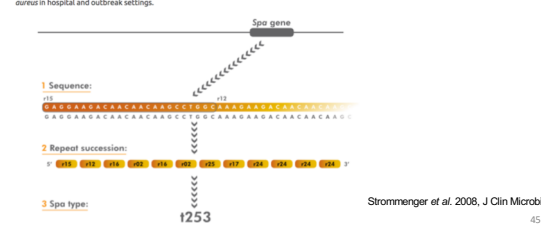
Exemple: spa typing *Staphylococcus*

spa Typing of *Staphylococcus aureus* as a Frontline Tool in Epidemiological Typing

Staphylococcus aureus spa typing

The spa typing technique uses the sequence of a polymorphic VNTR in the 3' coding region of the *S. aureus* specific staphylococcal protein A (*spa*). Each new base composition of the polymorphic repeat found in a strain is assigned a unique repeat code. The repeat succession for a given strain determines its spa type.

The individual repeat length for the *spa* VNTR is usually 24 bp, but exceptions of 21 to 30 exist. Although spa typing is a single-locus typing technique, it offers a subtyping resolution comparable to more expensive and/or laborious techniques such as MLST and PFGE. The technique is widely used for sub-typing of *S. aureus* in hospital and outbreak settings.



Strommenger et al. 2008, J Clin Microbiol
45

Trees or Evolutionary networks ? why we should do bioinformatics...

- Phylogenetic trees
 - PHYLIP®
 - PAUP®
 - PAML®
- Networks
 - Networks®
 - MST : minimum spanning tree
 - e-BURST®
- Populations Structure -Structure®
- Bayesian inference methods and other probability approaches
Maximum Likelihood
 - BEAST (Drummond et al. 2007)
 - <https://neast.community/>
 - <https://evolution.genetics.washington.edu/phylip/software.html>
 - RAxML (Stamatakis et al. 2014)

BMC Evolutionary Biology
Software
BEAST: Bayesian evolutionary analysis by sampling trees
Alexis J Drummond*1,2 and Andrew Rambaut*
*Corresponding author

The Exelixis Lab
Enabling Research in Evolutionary Biology
46

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 9 2014, pages 1312-1313
doi:10.1093/bioinformatics/btu033

Phylogenetics Advance Access publication January 21, 2014

RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies

Alexandros Stamatakis^{1,2}

¹Scientific Computing Group, Heidelberg Institute for Theoretical Studies, 69118 Heidelberg and ²Department of Informatics, Institute of Theoretical Informatics, Karlsruhe Institute of Technology, 76129 Karlsruhe, Germany

Associate Editor: Jonathan Wein

Bioinformatics, 30(9), 2014, 4453-4455
doi:10.1093/bioinformatics/btu033
Advance Access Publication Date: 9 May 2014
Applications Note

OXFORD

Phylogenetics

RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference

Alexey M. Kozlov^{1,*}, Diego Darriba², Tomás Flouri¹, Benoit Morel¹ and Alexandros Stamatakis^{1,2}

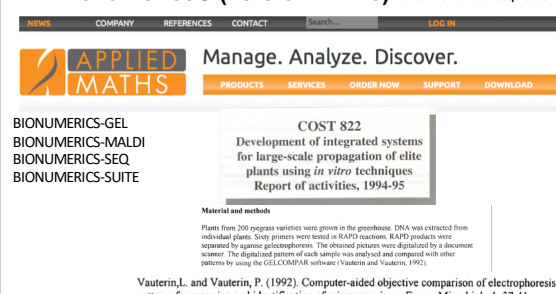
¹Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany and ²Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

47

Databases and data management

« BIONUMERICs: one universal platform to store and analyze biological data »

- **Bionumerics® (version 7-> 10) 2016: Biomérieux acquisition**




Vauterin, L. and Vauterin, P. (1992). Computer-aided objective comparison of electrophoresis patterns for grouping and identification of microorganisms. *Europ. Microbiol.*, 1, 37-41.

GIS Geographic information systems

- <http://philcarto.free.fr/> Philippe Waniez, IRD (under Windows only)
- **QGIS** www.qgis.org


the best choice that allows to go deep into GIS (landscape genetics) (works under all kind of systems)



49

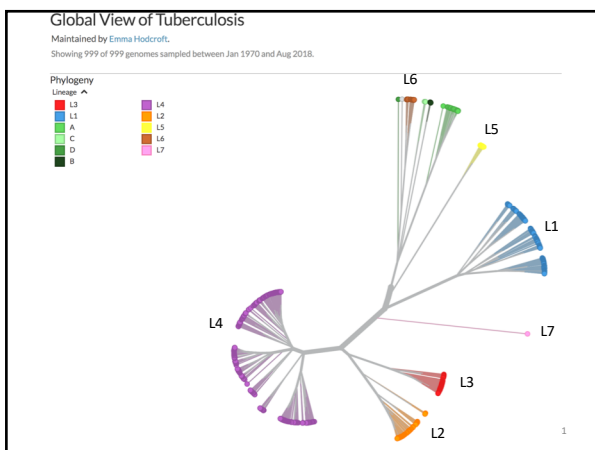
Use of molecular typing in real-time infectious diseases epidemiology

1. Investigation and control of outbreaks
2. Surveillance of infectious agents (endemic)
3. Study of infectious agents epidemiology

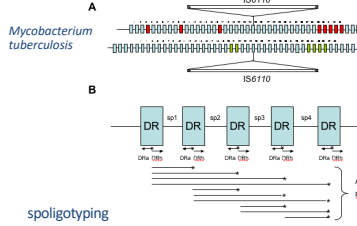


DIGITAL EPIDEMIOLOGY AND OTHER TECHNOLOGIES / INNOVATIONS RELEVANT FOR PUBLIC HEALTH FUNCTIONS

50



How to study the MTB CRISPR locus ? the « spoligotyping » principle



1744 references on Feb 7th 2024

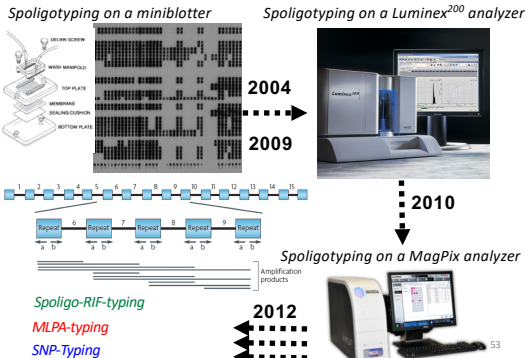
Journal of Clinical Microbiology, Apr. 1997, p. 907-914
0095-1137/97/34100-0
Copyright © 1997, American Society for Microbiology

Vol. 35, No. 4

Simultaneous Detection and Strain Differentiation of *Mycobacterium tuberculosis* for Diagnosis and Epidemiology
JUDITH KAMERBEK,¹ AREND KOLK,¹ MIRANDA VAN ACTERVELD,¹ DICK VAN SOOLINGEN,¹ SIOUKE RUIJTER,¹ ANNELES BUNSCHOTEN,¹ HENRI MOLHUIZEN,¹ RORY SHAW,² MADHU GOYAL,² AND JAN VAN EMBDEN^{1*}

52

Four generations of CRISPR and SNP typing Typing for *Mycobacterium tuberculosis*



Spoligotyping on a miniblatter (2004)

Spoligotyping on a Luminex²⁰⁰ analyzer (2009)

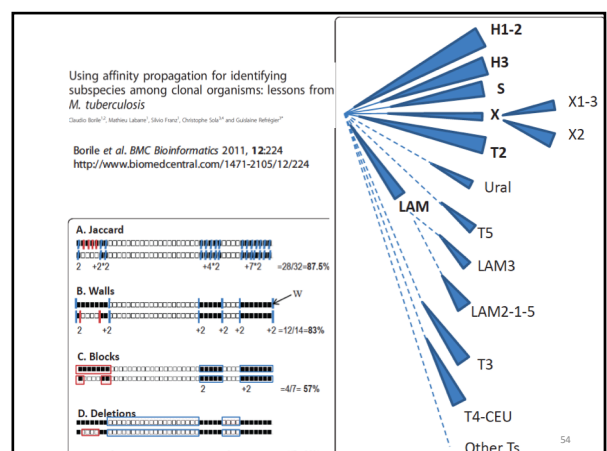
Spoligotyping on a MagPix analyzer (2010)

Spoligo-RIF-typing (2012)

MLPA-typing (2012)

SNP-Typing (2012)

53



http://crispr.u-psud.fr/Server/

http://www.pasteur.fr/recherche/genopole/Pf8/crispr/CRISPRDB.html

INSTITUT PASTEUR

PF8
Genotyping of Pathogens and Public Health Platform
250 rue de la Salette
75231 Paris Cedex 18
FRANCE
Tel: (33) 1 40 69 30 24
Fax: (33) 1 40 69 30 43

Institut Pasteur CRISPR Query Page

Please select the register you want to work on:
[Substrates entries] [CRISPRs]

And enter in the corresponding CRISPR Database:
[CRISPRs]

The application was developed at the Genotyping of Pathogens and Public Health Platform (Platform 8) of Institut Pasteur, under the leadership of G. Denon.

http://www.miru-vnrplus.org/MIRU/index.faces

MIRU-VNTRplus
Navigate Database Edit Options Tools Help

Home
Browse Database
Identification by Similarity Search
Nomenclature
Background Policy
Help
About us
Contact us
Imprint

Welcome to the MIRU-VNTRplus web application!

New to MIRU-VNTRplus? [Click here to open an example data set!](#)
Data from Homolka et al., BMC Microbiol 2008 ([PubMed](#))

Use the Home-button to return to this page at any time, or select "Home" from the Navigate-me Select one of the commands below to work with the database:

You have uploaded 0 user strains.

Enter a single user strain

Import multiple strains from file or clipboard

Change Settings
Change VNTR loci set: 24 loci

Reset default parameters

http://tbinsight.cs.rpi.edu/run_tb_lineage.html

TB insight
Tuberculosis Tracking and Control

TB-insight TB-Lineage TB-Via SPOTCLUST
Home About TB-Lineage About TB-Via About SPOTCLUST
Help Run TB-Lineage Run TB-Via Run SPOTCLUST
Documentation Help TB-Lineage Help TB-Via Complex Function

Run TB-Lineage [Please click!](#)
A tool to determine the major genetic lineage of a *Mycobacterium tuberculosis* isolate given its spoligotypes and optionally MIRU locus 24.

Tools: [Octal to Binary](#) / [Binary to Octal](#)

Select Model
Rules: CBN:

Select data format
[Spoligo only]

Please enter single strain information: Spoligotype (octal or binary format) and MIRU data (optional).
Click [here](#) for information on the MIRU Loci Set.

Spoligotype:

(Optional) MIRU:

Or you can upload a file with multiple strains. Click [here](#) for input format information.

Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database.
Couvin D, David A, Zozio T, Rastogi N.
Infect Genet Evol. 2019 Aug;72:31-43. doi: 10.1016/j.meegid.2018.12.030. Epub 2018 Dec 26.

http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE.

SITVIT2
HOME DESCRIPTION SEARCH ANALYSE ONLINE TOOLS OTHER STATISTICS SUBMIT YOUR DATA TB UNIT LINKS

Welcome to SITVIT2 (an update of SITVITWEB)

This online access to a WorldWide database of genotyping markers for *Mycobacterium tuberculosis* is provided by the Institut Pasteur de la Guadeloupe. You will find more details in the description section. The aim of this website is to provide a platform using several Bioinformatics tools for scientists to better monitor and describe the worldwide spread of *Mycobacterium tuberculosis* strains. SITVIT2 website allows the scientific community who is studying members of the *Mycobacterium tuberculosis* complex, to get information on epidemiology, demography, drug resistance, and molecular genotyping. This resource also provides information about statistical correlations that may exist between phylogenetic, clinical, demographic, and epidemiological data.

For information on how to use this service please read the [USER GUIDE](#)

For full description, see the following publication:
Couvin D, David A, Zozio T, Rastogi N. 2019. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infect Genet Evol.* (Epub ahead of print)

20 years of MTC genome analysis technological changes

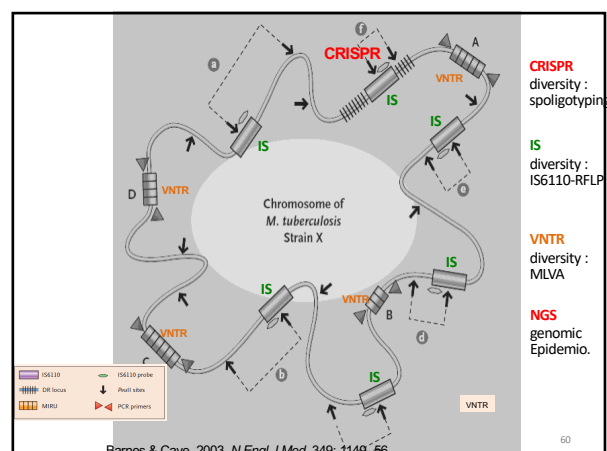
- The « early » age (1990-1997)**
 - IS6110-RFLP, IS1081-RFLP J. van Embden et al. 1993
 - PGRS-RFLP, DR-RFLP B. Ross et al. 1992
 - Pulse-Field Gel Electrophoresis E. Olson et al. 1995

FORMER GOLDEN STANDARD
- The « middle » age (1997-2001)**
 - Mixed-linker-IS6110 PCR, DRE-PCR, RAPD, LM-PCR ...
 - Spoligotyping J. Kamerbeek et al. 1997
 - VNTR (5 loci), QuB, Mtub R. Frothingham et al. 1998
 - extended spoligotyping J. van Embden et al. 2000
 - MIRU-VNTR (12 loci) P. Supply et al. 2001
- The « modern » age or « post genomic era » (2001-2008)**
 - Multi Locus Sequence Typing (MLST) L. Baker et al. 2004
 - Automatized HT-VNTR typing (24 loci) P. Supply et al. 2006
 - Automatized HT spoligotyping L. Couvin et al. 2004; Zhang et al. 2010
 - HT SNP typing D. Alland et al. 2004; Filippi et al. 2006
 - HT Deletion Region typing (RD) A. Tsolaki et al. 2004
 - HT Sequencing+MALDI-TOF MS MV Afanasiev et al. 2007
 - DNA-Chip spoligotyping El Song et al. 2007
 - MALDI-TOF MS spoligotyping C. Hansch et al. 2010
 - WGS whole genome next generation sequencing et al. 2010, 2011, J. Garay et al. 2011, Walker et al. 2013, 2015, Iina et al. 2013, Comas et al. 2013

NEW GOLDEN STANDARD

TODAYS

IMPORTANT SYNTHESIS



Birth of molecular epidemiology of tuberculosis : the IS6110-RFLP technique in 1993

[Cited 3163 times according to Google Scholar](#)

JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 1993, p. 406-409
0095-1137/93/020406-04\$02.000

Vol. 31, No. 2

NOTES

Strain Identification of *Mycobacterium tuberculosis* by DNA Fingerprinting: Recommendations for a Standardized Methodology

JAN D. A. VAN EMBDEN,¹ M. DONALD CAVE,² JACK T. CRAWFORD,³ JEREMY W. DALE,⁴ KATHLEEN D. EISENACH,⁵ BRIGITTE GICQUEL,⁶ PETER HERMANS,⁷ CARLOS MARTIN,⁸ RUTH McADAM,⁹ THOMAS M. SHINNICK,¹⁰ AND PETER M. SMALL¹¹

¹Unit Molecular Microbiology, National Institute of Public Health and Environmental Protection, P.O. Box 1, 3720 BA Bilthoven, The Netherlands; ²University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205; ³Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333; ⁴Department of Microbiology, University of Surrey, Guildford GU2 5XH, France; ⁵Armed Forces Research Institute, Adelaide, Elizabeth, Australia; ⁶Microbiologia, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain; ⁷Albert Einstein College of Medicine, Yeshiva University, Bronx, New York 10461; and ⁸Beckman Center B211, Howard Hughes Medical Institute, Stanford University, Stanford, California 94305-5429

Received 26 August 1992/Accepted 26 October 1992

cours. 2016 Octobre 26

A change of paradigm in 2006: VNTR Takes the lead !

REVUE DE CLINIQUE MICROBIOLOGIE, Dec. 2006, p. 498-510
0950-1270/06/44(6):498-510\$12.00
Copyright © 2006, American Society for Microbiology. All Rights Reserved.


Vol. 44, No. 12

Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis* H

Philip Supply,^{1,2*} Caroline Allin,³ Sarah Lejeune,^{1,2} Mara Cardoso-Oelemann,^{1,2} Sabine Riisch-Gertjes,⁴ Eric Willem,⁵ Tjibbe ten Haaft,⁶ Heek van Disselhorst-Klug,⁷ Sooyoung Rhee,⁸ Pablo Bernal,⁹ Natalia Kuregina,¹⁰ Barry Kreiswirth,¹¹ Christoph Soth,¹² Naim Raouf,¹³ Vincent Valleron,¹⁴ Maria Cristina Gutierrez,¹⁵ Merve Kavakci,¹⁶ Stefan Neumann,¹⁷ Robin Staue,¹⁸ Kristin Kremer,¹⁹ Camille Leach,²⁰ and Dick van Soolingen^{1,2}

¹INSERM U1057 and Institut Pasteur de Lille, Lille, France; ²Institut Pasteur de Lille, Lille, France; ³Department of Microbiology, Ghent University Hospital, Ghent, Belgium; ⁴Forschungszentrum Borstel, National Reference Center for Mycobacteria, Borstel, Germany; ⁵Diagnostic Laboratory for Infectious Diseases and Parasitology, National Institute for Public Health and the Environment, 3720 Bilthoven; ⁶Department of Tuberculosis Control, Municipal Health Service, Amsterdam; ⁷The Netherlands, Entomology Science Division, Department of Agriculture and Rural Development, Belfast, Northern Ireland; ⁸TB Center, Public Health Research Institute, Norway, Oslo, Norway; ⁹Unit of Infection and Immunity, London School of Hygiene and Tropical Medicine, Keppel Street, London, United Kingdom; ¹⁰Department of Microbiology, National Institute of Health, Hanoi, Vietnam; ¹¹Department of Microbiology, University of California, San Francisco, California, USA; ¹²Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹³Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁴Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁵Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁶Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁷Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁸Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ¹⁹Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania; ²⁰Unit of Tuberculosis, National Institute for Research and Development, Bucharest, Romania

Received 6 July 2006/Returned for modification 14 August 2006/Accepted 12 September 2006



[Cited 1613 times](#)

A worldwide database of spoligotyping : SpoDB4

39,295 clinical isolates
141 countries

[Cited 1276 times](#)

BMC Microbiology

Research article
***Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpoDB4) for classification, population genetics and epidemiology**

Lucia Garcia,¹ Lucija B. D'Almeida,² Leon Karam,³ Magdalena M. Prodingier,⁴ Andrea Cori,⁵ Sarah A. G. Hoogstraal,⁶ Caroline Allin,⁷ Lindiane de Almeida,⁸ Jody Anon,⁹ Yoonah Baek,¹⁰ Lohar Bhandari,¹¹ Patricia Calzavara,¹² Angel Calzavara,¹³ Soledad Lavigne,¹⁴ Richard Dier,¹⁵ Christyphel Elbertse,¹⁶ Jason F. Doran,¹⁷ Marjan Faustina,¹⁸ Doriane Neveline Erdmann,¹⁹ Daria Garcia de Velasco,²⁰ Carl Campbell,²¹ Eddie Cassa,²² Harrison M. Gomez,²³ M. Cristina Gutierrez,²⁴ Peter M. Hawkey,²⁵ Paul D van Helden,²⁶ Cornelia V. Kraljic,²⁷ Barry N. Krieger,²⁸ Kristina Kravari,²⁹ Milan Kubisa,³⁰ Sandra P. Kulkarni,³¹ Benjamin Leno,³² Troels Lillebaek,³³ Hyuk-Jin Lim,³⁴ Carlos Martin,³⁵ Christian Marnett,³⁶ Igor Medvedev,³⁷ Olga Narkalak,³⁸ Van-Fong Nguyen,³⁹ Ludmila Nuzumova,⁴⁰ Stefan Neumann,⁴¹ Kyo-Pyeong Park,⁴² Zoran Ralovic,⁴³ Vuokko Rintala,⁴⁴ Ratanapattanasri,⁴⁵ Taina Raaijmakers,⁴⁶ M. Lucia Rosenthal,⁴⁷ Sabine Riisch-Gertjes,⁴⁸ Anna Sakhakidze,⁴⁹ Sofia Sanchez,⁵⁰ Igor C. Schenkel,⁵¹ Urvashi B. Singh,⁵² Alex Sotomayor,⁵³ Robin A. Staue,⁵⁴ Dick van Soolingen,⁵⁵ Elizabeth B. Steinbock,⁵⁶ Philip N. Supply,⁵⁷ Frances Tenover,⁵⁸ Tatjana Trankovic,⁵⁹ Vincenzo Viscenzi,⁶⁰ Tommie C. Victor,⁶¹ Robin M. Warren,⁶² Suk-Fan Yip,⁶³ Rhonda Zaman,⁶⁴ Francine Zornick,⁶⁵ Nadia Zornick,⁶⁶ and Christophe Soth⁶⁷

Figure 1
Bioinformatic (32 binomially spaced percentage patterns) and statistical (50 most frequent classification) analysis of SpoDB4. First column: 32 "binomially spaced" (BS) number of percentage patterns for the binomially spaced. Second column: binomially spaced patterns. Third column: binary strings (BS) with black circles (BS) for respective binomially spaced patterns. Fourth column: color code (BS) for respective color code. Fifth column: total absolute number of isolates of the isolates.

Genomic Epidemiology arrives !!! 2010 change of era

- Schürch et al. 2010;** Netherlands, 104 clinical isolates, identical IS6110-RFLP, eight SNPs (specific for the outbreak), five SNPs clusters, each with an index case.
- Gardy, et al. 2011** Vancouver, 41 cases, MIRU-VNTR, 1 genotype, WGS revealed two outbreaks
- Walker TM et al. 2013** Midlands, UK, 390 isolates, 254 Patients

SCHEDULE

- Diapos 4 to 66 = Course
- Diapos 67 to end = some clues on my research topic

« Does my dataset contains temporal and evolutionary signal ? »

Study : Space, Time, Genomic changes and Genomic Interactions between species...i.e. ecological changes

Systems Epidemiology and Evolution of Tuberculosis

Global history

co-evolution

Arms race (darwinian selection) or **Pure co-divergence (neutral)**

Purifying selection
Positive selection
Balancing selection

TIME

SPACE

Genetic geography

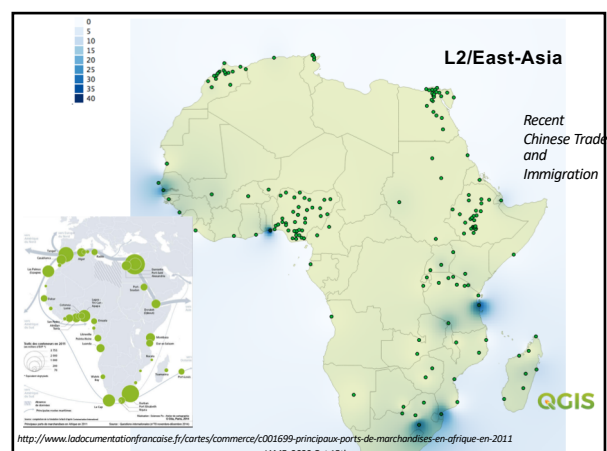
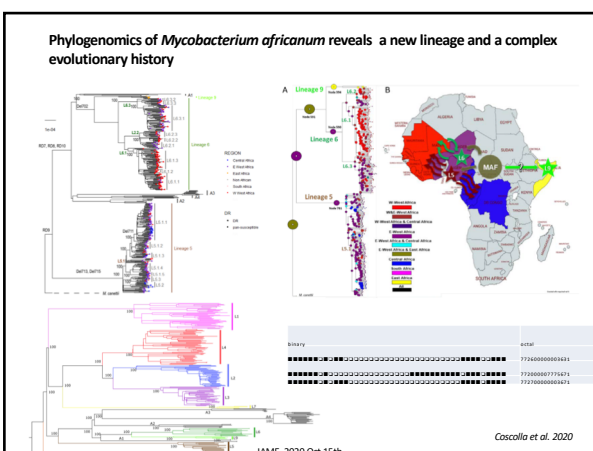
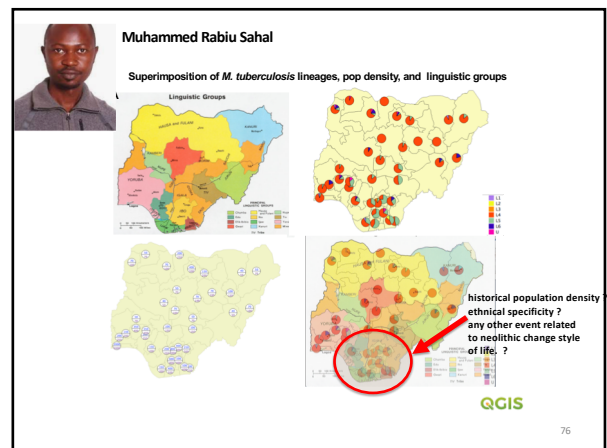
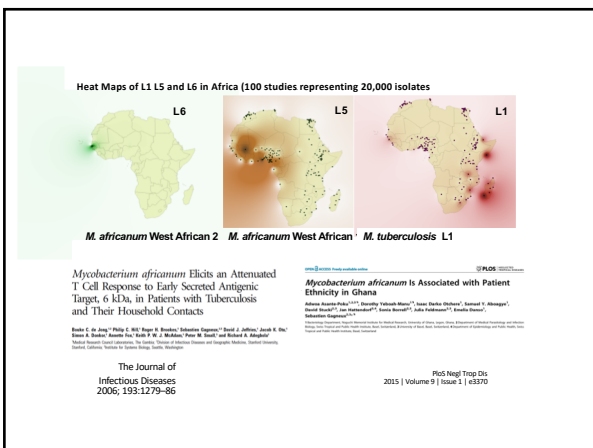
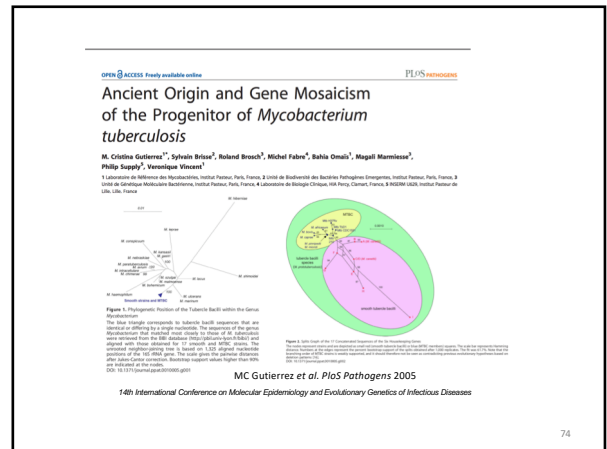
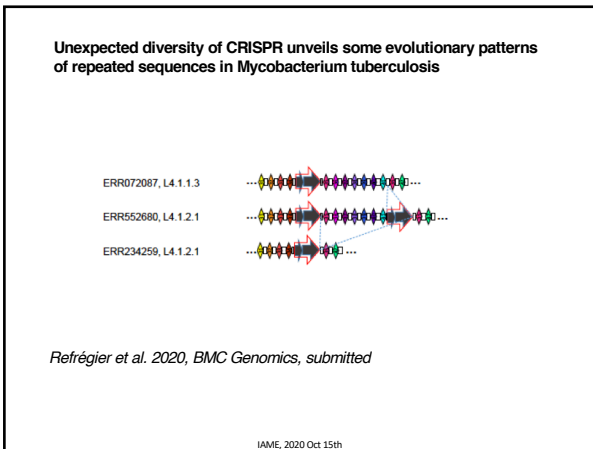
Pathocenosis

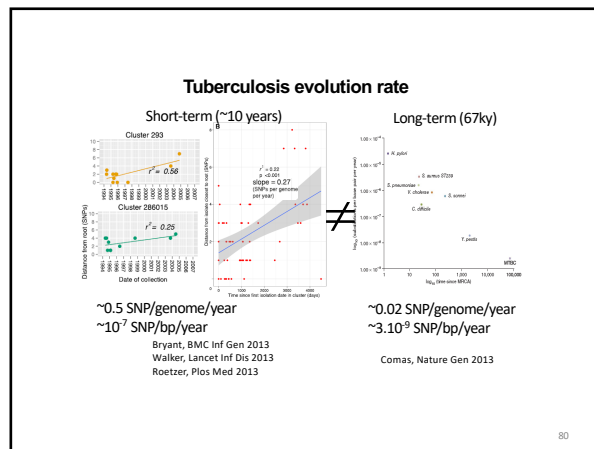
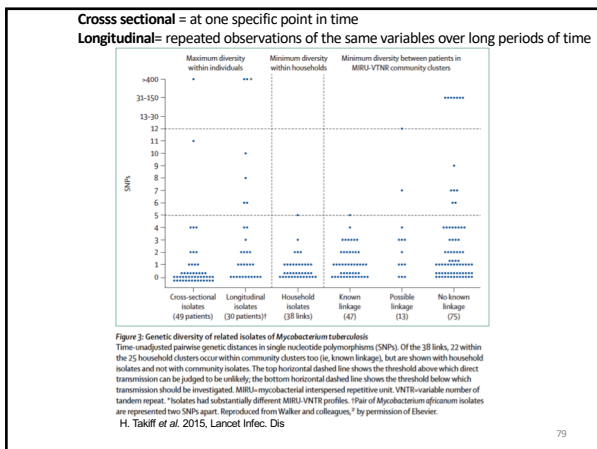
ECOLOGY

Fernand Braudel 1902-1985

Luigi Luca Cavalli-Sforza 1922-2018

Mirko Grmek 1924-2000





Tuberculosis 143 (2023) 102578

Contents lists available at ScienceDirect

Tuberculosis

ELSEVIER

journal homepage: www.elsevier.com/locate/tuber

Review

The paradoxes of Mycobacterium tuberculosis molecular evolution and consequences for the inference of tuberculosis emergence date

R. Zein-Eddine^a, F. Hak^b, A. Le Meur^c, C. Genestet^{d,e}, O. Dumitrescu^{f,g}, C. Gueyex^h, G. Senelleⁱ, C. Sola^{j,k,l,m}, G. Refrégierⁿ

^a Laboratoire d'Épidémiologie et Biostatistiques, École Polytechnique, Institut National de la Santé et de la Recherche Médicale: U1182, Centre National de la Recherche Scientifique: UMRS164, France
^b Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Évolution, 91190, Gif sur Yvette, France
^c Centre International de Recherche en Infectiologie, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, Inrae U1111, CNRS UMS308, Lyon, France
^d Hospices Civils de Lyon, Institut des Agents Infectieux, Laboratoire de bactériologie, Lyon, France
^e CDC-Computer Science Department, EMR/ST Institute, EMR 4274 CNRS, Univ. Bourgogne Franche-Comté (UBFC), 16 Route de Gray, 25000, Besançon, France
^f Université de Paris, IAME, UMR1137, INSERM, Paris, France
^g AP-HP, GHU Pitié-Salpêtrière, Service de microbiologie spécialisée et de référence, Paris, France
^h Université Paris-Saclay, Saint-Aubin, France

OPEN ACCESS Freely available online

Evolution and Diversity of Clonal Bacteria: The Paradigm of Mycobacterium tuberculosis

Tago Dos Vultos¹, Olga Mestre¹, Jean Raulier², Marcin Golec³, Nalin Rastogi⁴, Voahangy Rasolofo⁵, Tone Tonjum^{6,7}, Christophe Sola^{8,9}, Ivan Matic¹⁰, Brigitte Gicquel¹¹

Research paper
 Turkish and Japanese Mycobacterium tuberculosis sublineages share a remote common ancestor

Guislaine Refrégier^{1,2}, Edgar Abadia³, Tomohige Matsumoto⁴, Hiromi Ano⁵, Tetsuya Takashima⁶, Izu Tsuyuguchi⁷, Elif Aktas⁸, Fuisun Cömert⁹, Michel Kiroport Gomgnimhou¹⁰, Stefan Panaiotov¹¹, Jody Phelan¹², Frances Coll¹³, Ruth McNeerney^{14,15}, Taane G. Clark¹⁶, Christophe Sola¹⁷

IAME, 2020 Oct 15th
 Dos Vultos et al. 2018, Plos ONE
 Refrégier et al. 2016, Infect Genet Evol

Molecular Anthropology application, tuberculosis history in Japan-1

Infection, Genetics and Evolution 45 (2016) 401–473

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

ELSEVIER

journal homepage: www.elsevier.com/locate/iegid

Research paper

Turkish and Japanese Mycobacterium tuberculosis sublineages share a remote common ancestor

Guislaine Refrégier^{1,2}, Edgar Abadia³, Tomohige Matsumoto⁴, Hiromi Ano⁵, Tetsuya Takashima⁶, Izu Tsuyuguchi⁷, Elif Aktas⁸, Fuisun Cömert⁹, Michel Kiroport Gomgnimhou¹⁰, Stefan Panaiotov¹¹, Jody Phelan¹², Frances Coll¹³, Ruth McNeerney^{14,15}, Taane G. Clark¹⁶, Christophe Sola¹⁷

¹ Institut für Integrative Biologie der Carl-Neuberg-Universität, CNRS, Univ. Paris-Saclay, Université Paris-Saclay, 91190, Gif sur Yvette cedex, France
² Institut Universitaire de Recherche Clinique (IURC), CNRS, France
³ Department of Clinical Immunology and Development, Osaka Prefecture Hospital Organization, Osaka Prefecture Medical Center for Respiratory and Allergic Diseases, Habikino-city, Japan
⁴ Iqbal Khan Research and Training Hospital, Istanbul, Turkey
⁵ Faculty of Medicine, Bilkent University, Ankara, Turkey
⁶ National Center of Parasitic and Infection Diseases, Subtropical Parasitology, Taipei, Taiwan
⁷ Faculty of Infection and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK
⁸ Infection and Immunity Unit, UCL East Institute, University of College London, Old Man Building, Gower Street Hospital, Gower Street, London, UK
⁹ Pathogen Genomics Group, Biological, Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Molecular Anthropology application, tuberculosis history in Japan-2

Connection between two historical tuberculosis outbreak sites in Japan, Honshu, by a new ancestral Mycobacterium tuberculosis L2 sublineage

Published online by Cambridge University Press: 19 January 2022

Christophe Gueyex, Gaetan Senelle, Guislaine Refrégier, Florence Bretelle-Establie, Emmanuelle Cambau and Christophe Sola

Tochi province and of Kobe where the « G3 » group was described.

SCIENCE ADVANCES | RESEARCH ARTICLE

HEALTH AND MEDICINE

Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation

Ola B. Brynildsrud¹, Caitlin S. Pepperell^{2,3}, Philip Suffys⁴, Louis Grandjean⁵, Johana Monteserin^{6,7}, Nadia Debech⁸, Jon Bohlin⁹, Kristian Aifsnies¹⁰, John O.-H. Pettersson^{14,15,19}, Ingerid Kirkeleite¹, Fatima Fandinho¹¹, Marcia Aparecida da Silva¹¹, Joao Pervigiao¹², Isabel Portugal¹², Miguel Viveiros¹³, Taane Clark^{14,15}, Maxine Caw^{16,17}, Sarah Dunstan¹⁸, Phan Vuong Khac Thai¹⁹, Beatriz Lopez²⁰, Viviana Ritacco^{6,7}, Andrew Kitchen²⁰, Tyler S. Brown²¹, Dick van Soolingen²², Mary B. O'Neill^{2,23*}, Kathryn E. Holt^{14,24}, Edward J. Feil²⁵, Barun Mathema²⁶, Francois Balloux²⁷, Vegard Eldholm^{1*}

On the basis of population genomic and phylogeographic analyses of 1669 *Mycobacterium tuberculosis* lineage 4 (L4) genomes, we find that dispersal of L4 has been completely dominated by historical migrations out of Europe. We demonstrate an intimate temporal relationship between European colonial expansion into Africa and the Americas and the spread of L4 tuberculosis (TB). Markedly, in the age of antibiotics, mutations conferring antimicrobial resistance overwhelmingly emerged locally (at the level of nations), with minimal cross-border transmission of resistance. The latter finding was found to reflect the relatively recent emergence of these mutations, as a similar degree of local restriction was observed for susceptible variants emerging on comparable time scales. The restricted international transmission of drug-resistant TB suggests that containment efforts at the level of individual countries could be successful.

Copyright © 2018 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

85

The TB-Annotator Project, February 2022, 15901 Genomes

Guillaume Refregier, Emmanuelle Cambau, Flore Breteille-Establat, Gaetan Senelle, Christophe Guyeux

ASSISTANCE PUBLIQUE HÔPITAL DE PARIS

Hierarchy of aspects to consider for sustainable Data analysis

but... all of this requires programming expertise, trained bioinformatics staff... and is difficult starting from scratch...

Mölder F, et al. Sustainable data analysis with Snakemake [version 2]. F1000Research 2021, 10:23 [doi: 10.12688/f1000research.20622.2]

F1000Research

Senelle et al. 2023 Tuberculosis

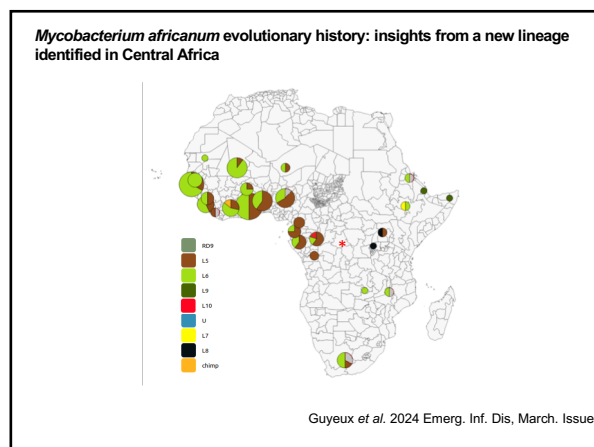
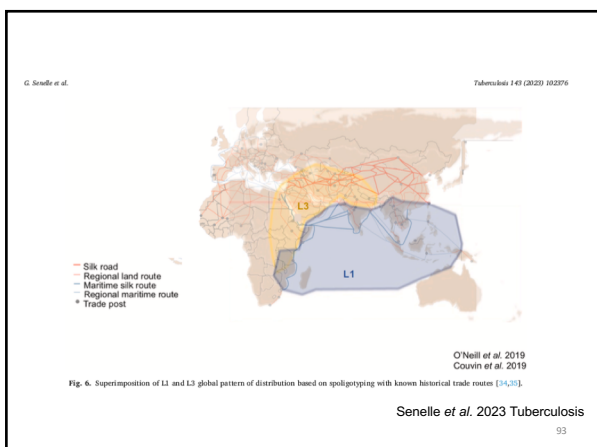
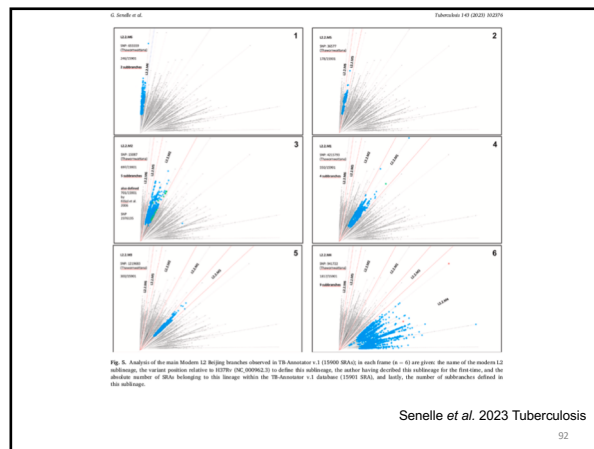
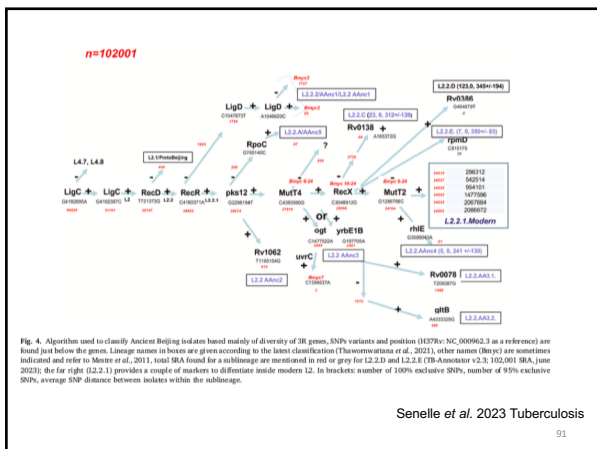
88



Senelle et al. 2023 Tuberculosis

89

Senelle et al. 2023 Tuberculosis

90



FEMS Microbiology Reviews 27 (2003) 145–163

www.fems-microbiology.org

The MerR family of transcriptional regulators

Nigel L. Brown ^{*}, Jivko V. Stoyanov ¹, Stephen P. Kidd ², Jon L. Hobman

School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 26 November 2002; received in revised form 23 January 2003; accepted 27 January 2003

First published online 6 May 2003

ESM 2023 Tirana, Albania, Jun 25–28th 2023

Another way: asking ChatGPT ?

How do I classify Mycobacterium tuberculosis genomes, the agent of human tuberculosis ?

As an AI language model, I can provide you with guidance. Here's a step-by-step method:

Data Preprocessing: Obtain the genomic data of *Mycobacterium tuberculosis* strains that you want to classify with a suitable format, such as FASTA or GenBank. Genome Assembly: With raw sequencing reads, you'll need to perform genome assembly to reconstruct the complete genomes. Tools like **SPAdes** or **Velvet** can be used for this purpose.

Genome Annotation: Annotate the assembled genomes to identify various genomic features such as coding sequences, regulatory elements, and other relevant genetic elements. Tools like **Prokka** or **RAST** can assist in genome annotation.

Core Genome Alignment: Perform a core genome alignment to identify conserved regions across different *Mycobacterium tuberculosis* strains. This step helps in identifying common genes and variations within them. Tools like **Roary** or **Parasp** can be used for core genome alignment.

Phylogenetic Analysis: Construct a phylogenetic tree based on the aligned core genome sequences. Phylogenetic analysis helps in determining the evolutionary relationships and relatedness among different strains. Tools like **RAxML** or **FastTree** can be employed.

SNP Analysis: Identify single nucleotide polymorphisms (SNPs) within the aligned genome sequences. SNPs can provide insights into genetic variations between strains. Tools such as **Snippy** or **GATK** can be used for SNP analysis.

Drug Resistance Prediction: Determine the presence of drug resistance-associated mutations within the genomes. Tools like **TB-Profit** or **Mykrobe Predictor** can aid in predicting drug resistance profiles based on genomic variations.

Classification Model: Utilize **machine learning** or **deep learning algorithms** to develop a classification model.

Train the model using annotated genomes with known classifications (e.g., drug-resistant or drug-sensitive). Features such as SNPs, gene presence/absence, or drug resistance-associated mutations can be used for classification.

Model Evaluation: Assess the performance of your classification model using appropriate metrics such as accuracy, precision, recall, or F1-score. Cross-validation or holdout testing can be used to evaluate the model's generalizability.

Classification of Unknown Genomes: Apply the trained classification model to classify unknown *Mycobacterium tuberculosis* genomes based on their genomic features. The model will provide predictions regarding drug resistance profiles or other classification categories.

Remember: the accuracy of your classification will depend on the quality and diversity of the training data, as well as the robustness of your classification model. It's crucial to ensure the accuracy of the annotations, perform appropriate quality control steps, and stay updated with the latest research in Mycobacterium tuberculosis genomics.