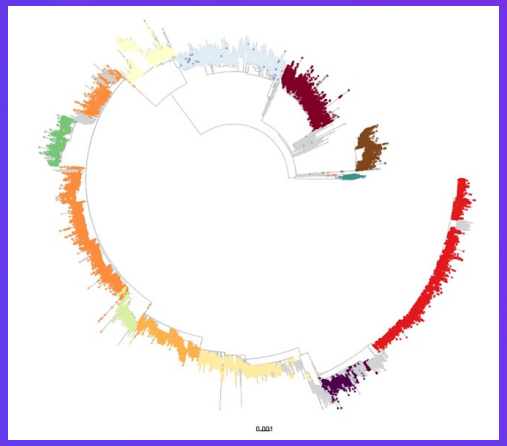
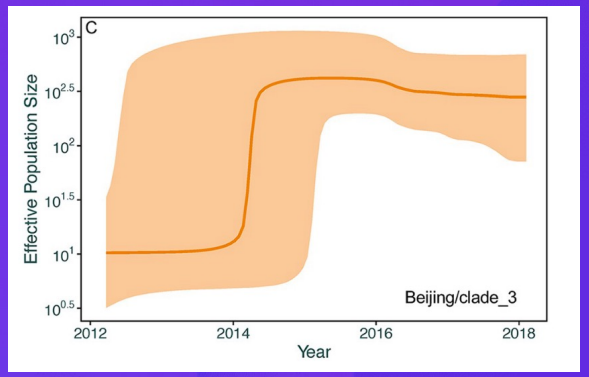
Van ML 2024

Phylogenetic applications to real-world data





Dr. Ben Sobkowiak
Yale University / University College London

Outline of this lecture:

1. The history of phylogenetics in biology and epidemiology
2. Inferring transmission from phylogenies
3. Phylodynamics with BEAST
4. Further applications of phylogenetics – Phylogeography, detecting selection and identifying associated variants

About me:

Consultant Computational Biologist at Yale University
Honorary Research Fellow at UCL
(Former postdoc at SFU & UBC/BCCDC)

Interest in pathogen transmission, bioinformatics, and mathematical modelling of infectious disease.

Harnessing whole genome sequence data to better understand transmission and evolution of pathogens – mainly focus on tuberculosis.

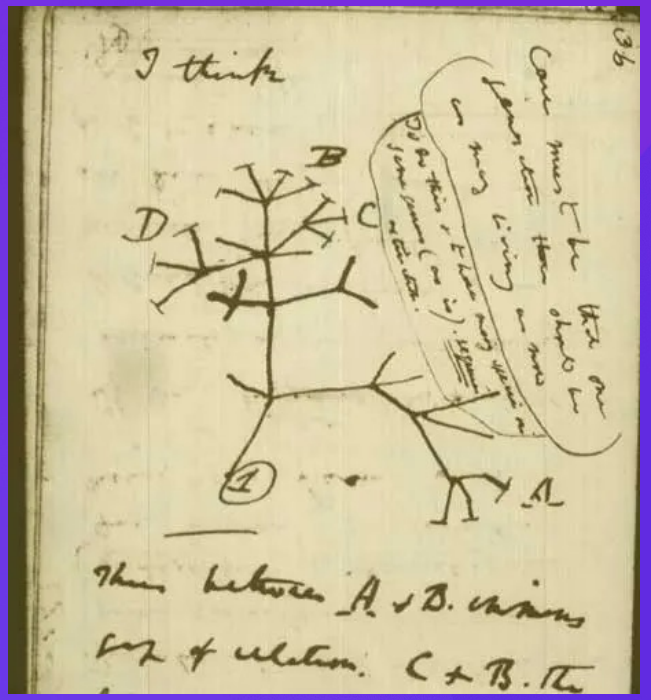
Outline of this lecture:

1. **The history of phylogenetics in biology and epidemiology**
2. Inferring transmission from phylogenies
3. Phylodynamics with BEAST
4. Further applications of phylogenetics – Phylogeography, detecting selection and identifying associated variants

1. The history of phylogenetics

Early concepts

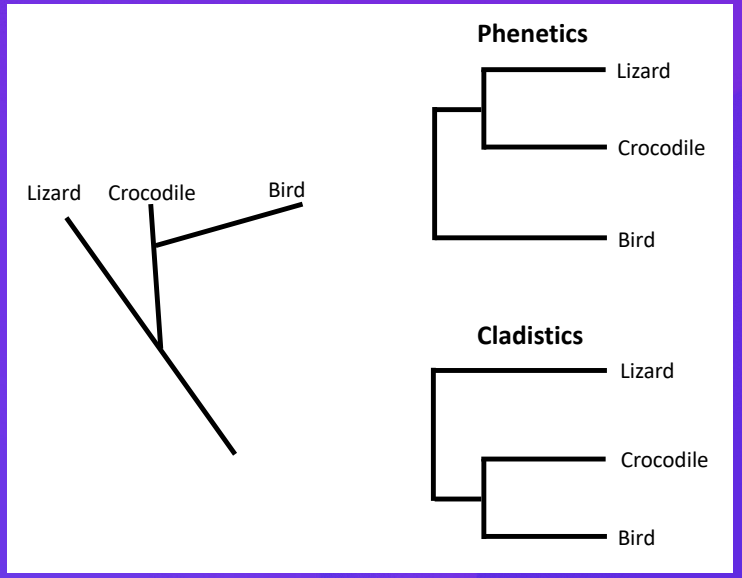
- The concept of a "tree of life" predates Darwin, with early naturalists like Linnaeus (1707-1778) classifying species based on physical characteristics.
- Charles Darwin, in "On the Origin of Species" (1859), proposed the idea of common descent and branching patterns of evolution, laying the groundwork for phylogenetics.



1. The history of phylogenetics

Phenetics vs cladistics

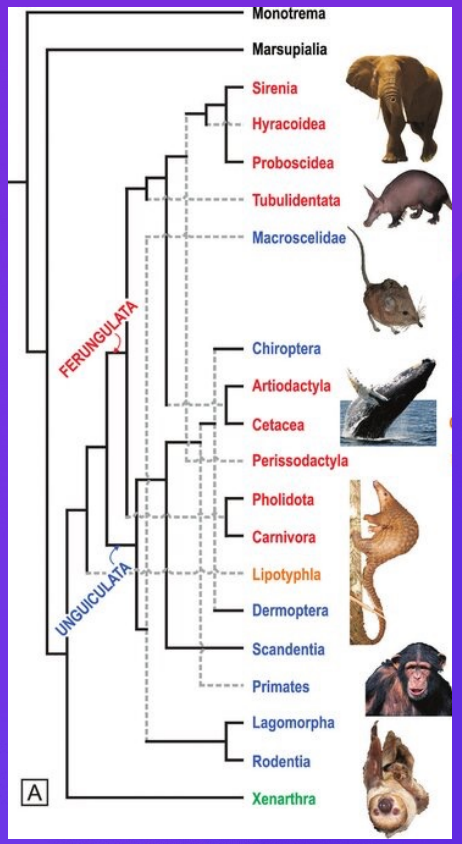
- Cladistics brought a more rigorous and systematic approach to phylogenetics.
- Focused on shared derived characteristics (synapomorphies) to infer evolutionary relationships rather than classifying organisms based on overall similarity (phenetics).
- Cladistics provides a more accurate reflection of the organism's evolutionary history by understanding the evolutionary process.



1. The history of phylogenetics

Morphology-Based Phylogenetics

- Initially, phylogenetic relationships were inferred from morphological (structural) traits.
- With the discovery of DNA structure and the development of molecular biology, scientists began to use genetic information to infer evolutionary relationships.
- In the 1960s brought the "molecular clock" hypothesis, which suggested that rates of molecular change could be used to date evolutionary divergences.

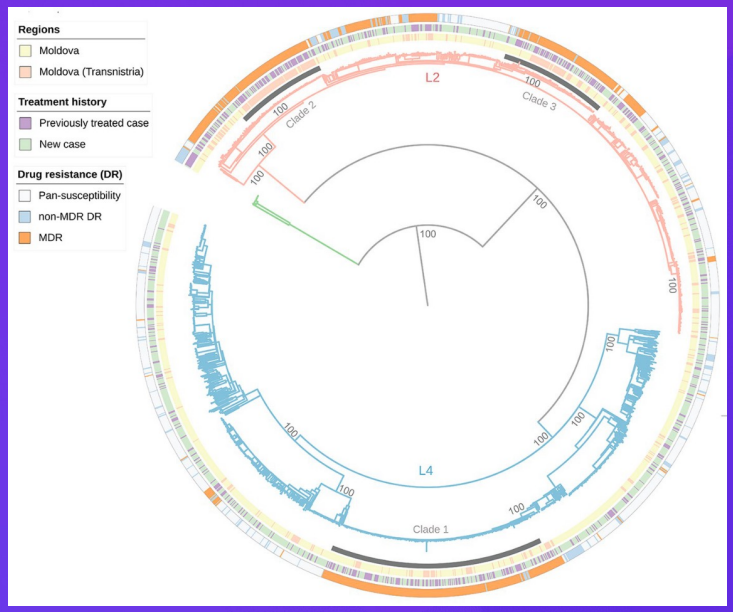


Adapted from Lehmann 2010

1. The history of phylogenetics

‘Phylogenomics’

- The field has entered the era of ’phylogenomics’, where entire genomes are used for phylogenetic analysis, providing unprecedented detail and accuracy.
- Phylogenomics provides more data points so can explore evolutionary relationships at a higher resolution (e.g., transmission pairs or recently diverged individuals)

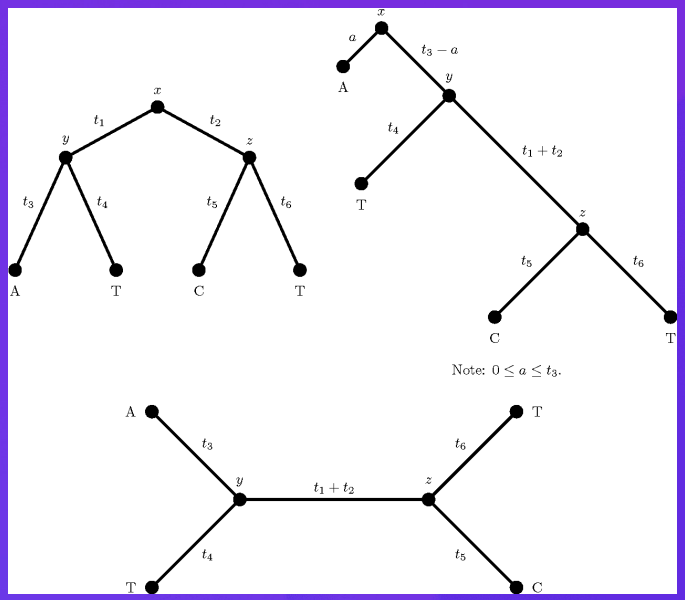


Yang et al. 2022

1. The history of phylogenetics

Computer-Aided Phylogenetics

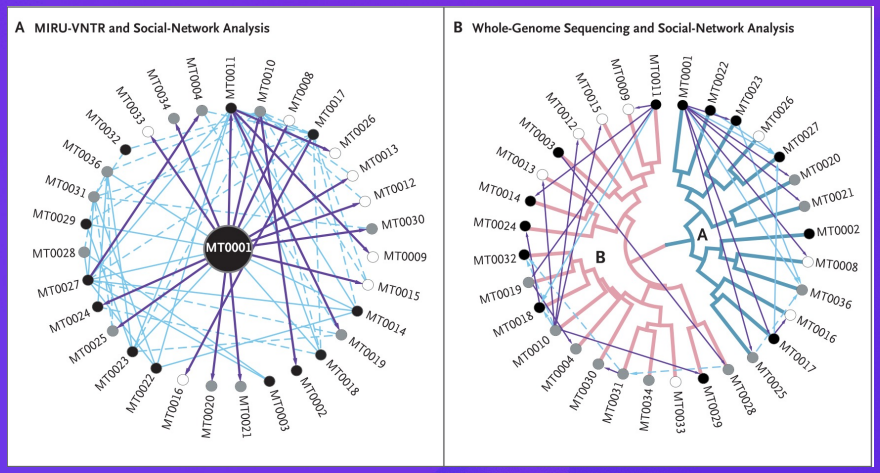
- Methods like Maximum Parsimony, Maximum Likelihood, and Bayesian Inference allow more complex analysis of large datasets.
- From simple distance-based methods such as Neighbour-joining trees to more complex ML and Bayesian approaches.
- Choice of method depends on factors such as the complexity of data, evolutionary history of the taxa, computational power and time.



1. The history of phylogenetics

Phylogenetics in genomic epidemiology

- This integration of phylogenetics allows for detailed tracking of pathogens, insight into their evolutionary dynamics, and aids in public health responses.
    - Transmission/population dynamics
    - Outbreak investigation
    - Emergence of antimicrobial resistance
    - Vaccine development
    - Surveillance
    - Co-evolution between pathogens and/or host



Gardy et al. 2011

Outline of this lecture:

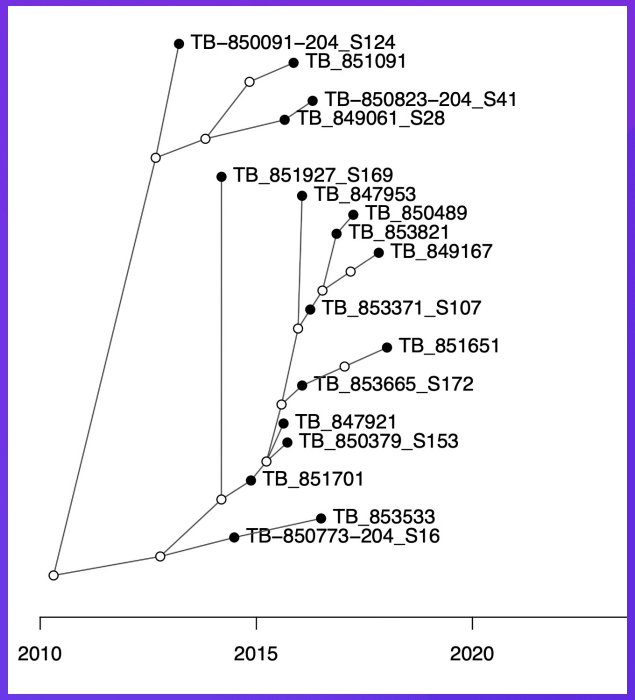
1. The history of phylogenetics in biology and epidemiology
2. **Inferring transmission from phylogenies**
3. Phylodynamics with BEAST
4. Further applications of phylogenetics – Phylogeography, detecting selection and identifying associated variants

2. Inferring transmission from phylogenies

Transmission inference

- We can link phylogenetic trees with transmission to reconstruct transmission trees

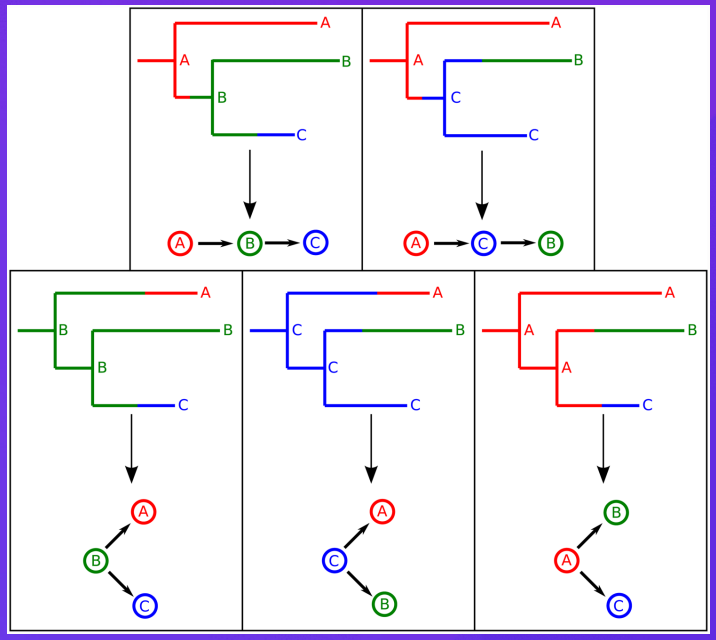
- Transmission trees can tell us:
  - Amongst whom transmission is happening and direction of transmission events
  - The location of transmission (hotspots?)
  - The time between infection and transmission
  - The extent of recent transmission
  - If there are superspreader events



2. Inferring transmission from phylogenies

Transmission inference

- Phylogenetic tree ≠ transmission tree in all instances or it might not be clear who-infected-whom
- Can use computational approaches to incorporate epidemiological models and estimate transmission trees from phylogenies or sequence data directly.



Hall et al, 2015

2. Inferring transmission from phylogenies

Transmission inference

- Example software to estimate transmission networks from phylogenies or genomic data
  - TransPhylo
  - BEASTLIER
  - SCOTTi
  - Outbreaker2
  - Phybreak
- Employ different underlying models and parameters that can result in different inferred transmission networks

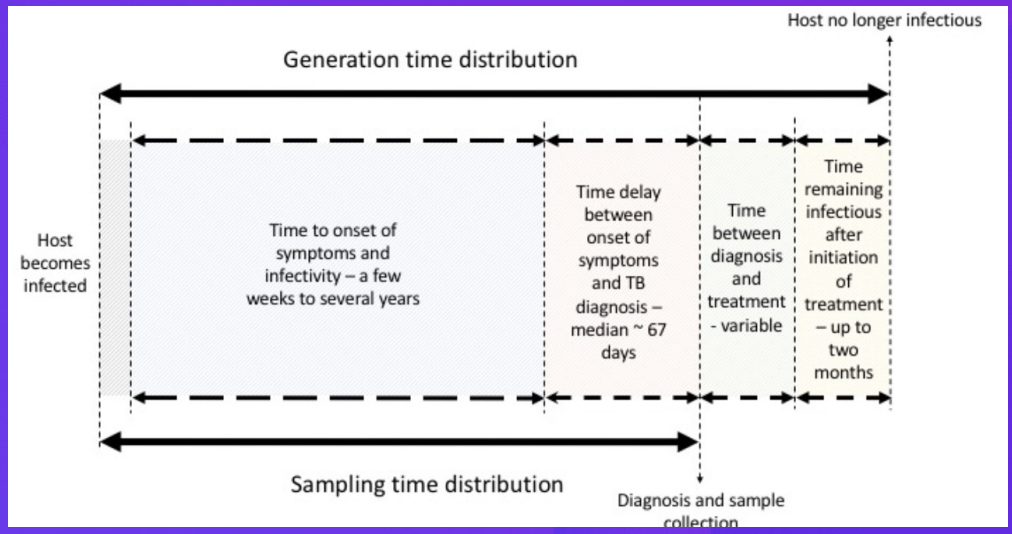
From Sobkowiak et al, 2023

2. Inferring transmission from phylogenies

TransPhylo

- Bayesian method for inferring transmission events from a timed phylogeny (e.g., created using BEAST)

- Epidemiological parameters including:
  - Generation time distribution
  - Sampling time distribution
  - Sampling density
  - Within-host coalescent rate

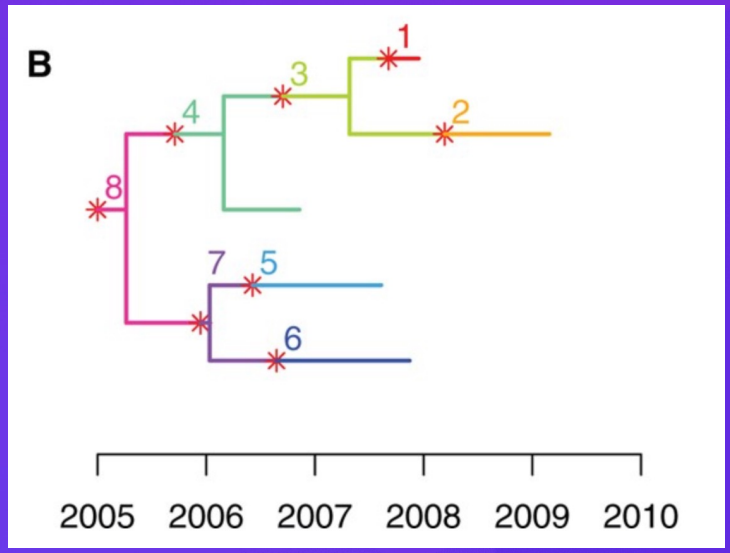


From Sobkowiak et. al. 2020

2. Inferring transmission from phylogenies

TransPhylo

- Colours the phylogeny by transmission events
- Each new sampled or unsampled case is given a unique colour and transmission events illustrated with the star



From Didelot et. al. 2017

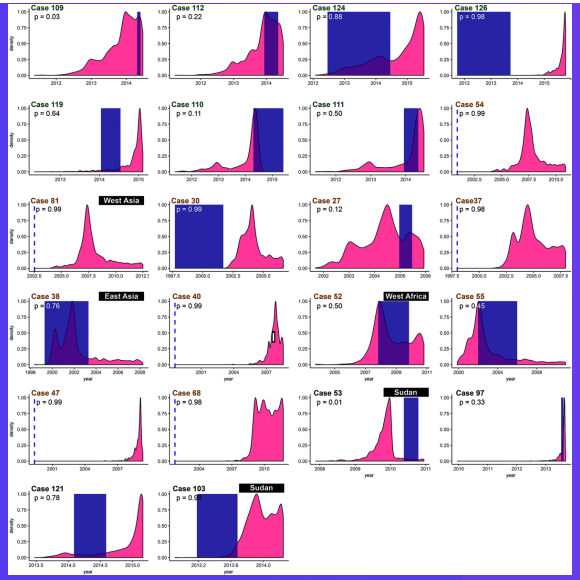
2. Inferring transmission from phylogenies

Estimating transmission networks from past outbreaks

MICROBIAL GENOMICS
Volume 4, Issue 10
Research Article | Open Access

Genome-based transmission modelling separates imported tuberculosis from recent transmission within an immigrant population

Diepreye Ayabina[1], Janne O. Ronning[2], Kristian Alfsnes[2], Nadia Debech[2], Ola B. Brynildsrud[2], Trude Arnesen[2], Gunnstein Norheim[2], Anne-Torunn Mengshoel[2], Rikard Rykkvin[2], Ulf R. Dahle[2], Caroline Colijn[1], Vegard Eldholm[2]



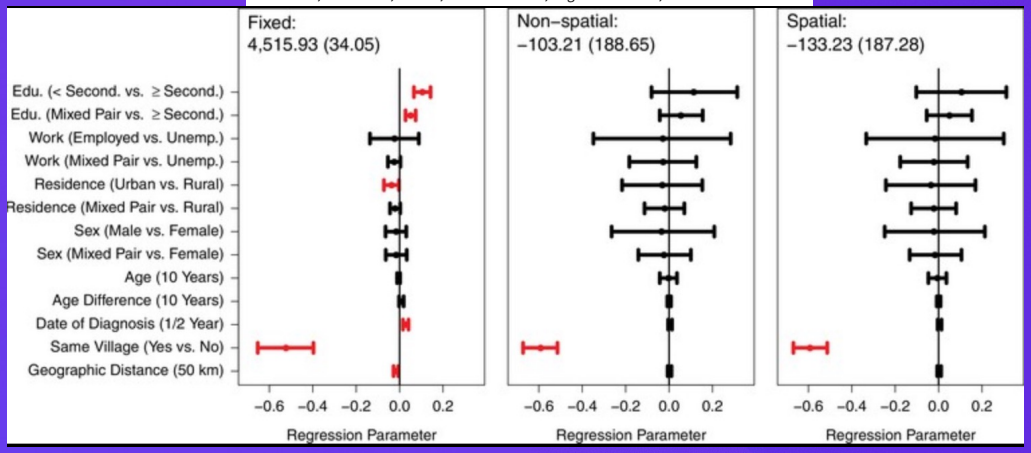
JOURNAL ARTICLE

Spatial Modeling of *Mycobacterium Tuberculosis* Transmission with Dyadic Genetic Relatedness Data

FREE

Joshua L. Warren ✉, Melanie H. Chitwood, Benjamin Sobkowiak, Caroline Colijn, Ted Cohen

Biometrics, Volume 79, Issue 4, December 2023, Pages 3650–3663,



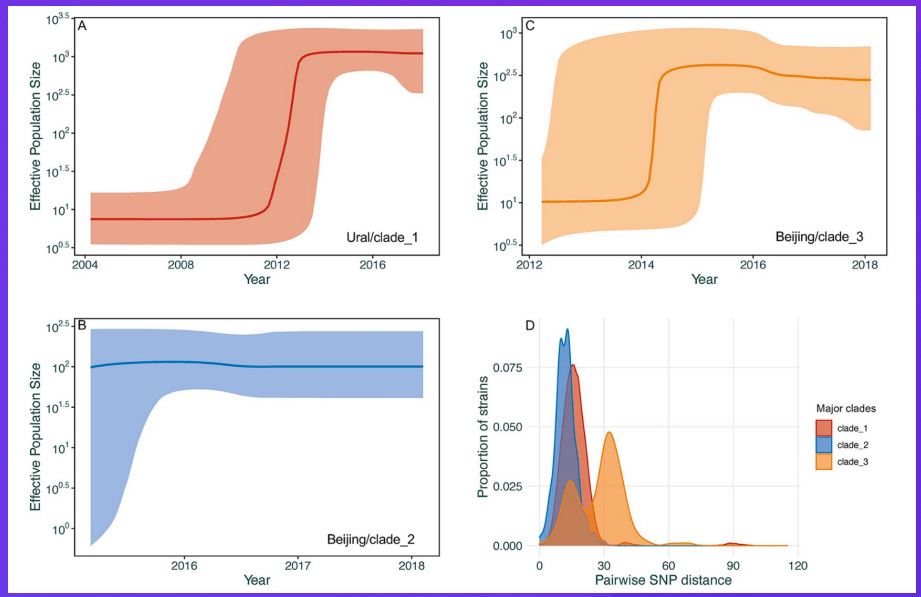
Outline of this lecture:

1. The history of phylogenetics in biology and epidemiology
2. Inferring transmission from phylogenies
3. **Phylodynamics with BEAST**
4. Further applications of phylogenetics – Phylogeography, detecting selection and identifying associated variants

3. Phylodynamics

What is phylodynamics?

- Phylodynamics is the study of how epidemiological and evolutionary processes act to shape phylogenies
- Phylodynamic approaches typically involve the reconstruction of phylogenetic trees, combined with mathematical models that describe evolutionary processes and population dynamics
- Many packages directly in BEAST2 (e.g., MSBD, EpiInf, BASTA)



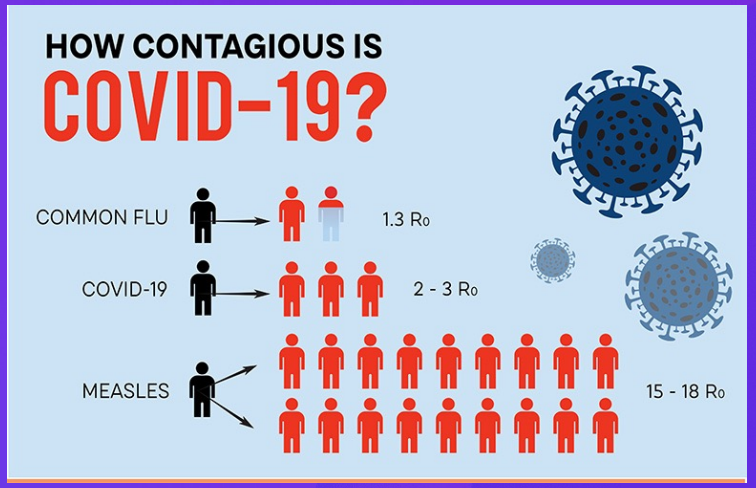
From Yang et. al. 2022

3. Phylodynamics

What can phylodynamics tell us about disease?

- These analyses can help us to estimate key epidemiological parameters:
  - R0
  - Transmission rate
  - Force of infection
  - Serial intervals

- Infer past population demographics to estimate how lineages or clades on a tree have grown or shrunk through time

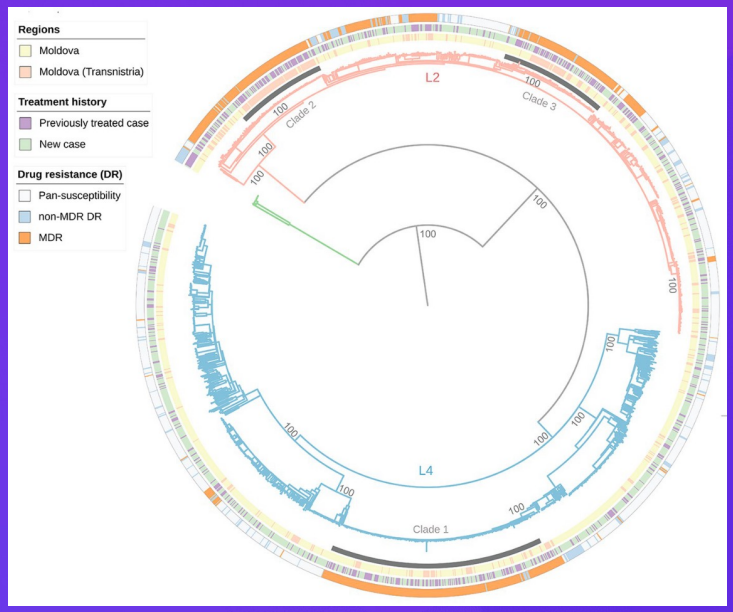


From endeavor.moffitt.org

3. Phylodynamics

Identify rapidly spreading *Mtb* strains

- Large clades of MDR-TB strains identified in Moldova on an un-timed phylogeny
- MDR-TB concerning as difficult to treat
- Previously shown one clade is distributed throughout the country
- But is it spreading rapidly – reason to be more concerned!

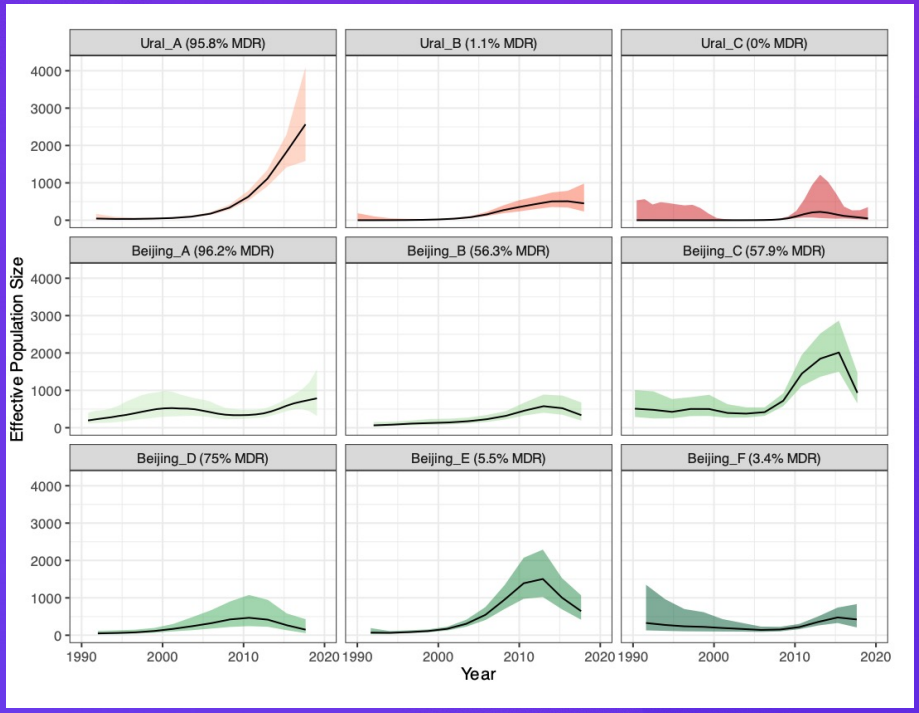


Yang et al. 2022

3. Phylodynamics

Identify rapidly spreading *Mtb* strains

- Reconstructed past population dynamics of 9 clusters of *Mtb* in Moldova.
- Found evidence of recent, rapid expansion of one MDR-TB cluster – strains contain multiple novel mutations
- Concerning and leads to more questions, what's driving this expansion (mutations, behaviour etc.)?



Chitwood et. al. 2024, Nat. Comms (accepted)

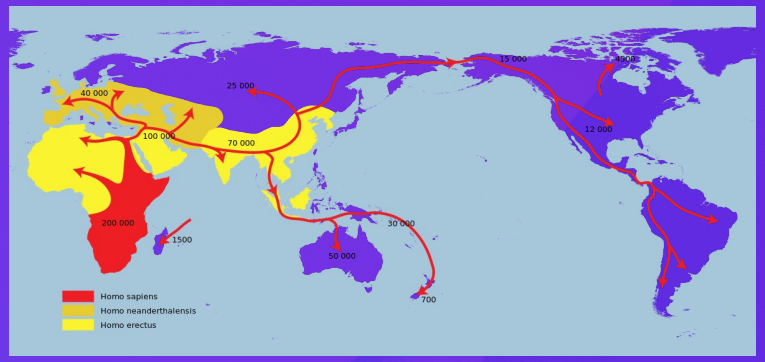
Outline of this lecture:

1. The history of phylogenetics in biology and epidemiology
2. Inferring transmission from phylogenies
3. Phylodynamics with BEAST
4. **Further applications of phylogenetics – Phylogeography, detecting selection and identifying associated variants**

4. Further applications: Phylogeography

What is phylogeography?

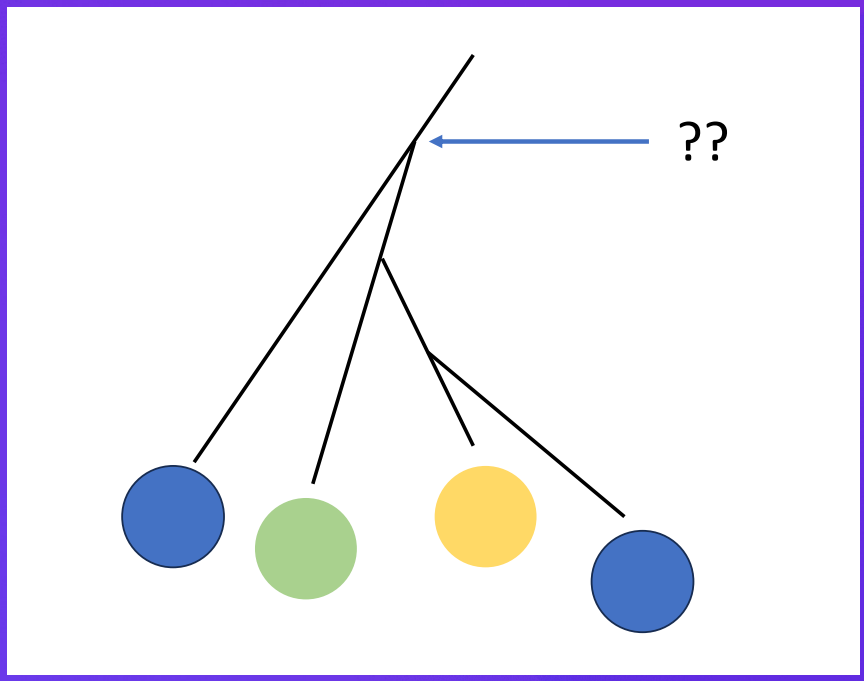
- Phylogeography allows us to reconstruct when and where lineages or clades were present
- Estimate the location of the emergence of new strains or lineages
- Track the migration of different strains or the flow of particular genes or traits
- Packages in BEAST2 - BREAK_AWAY, GEO_SPHERE



4. Further applications: Phylogeography

Phylogeography

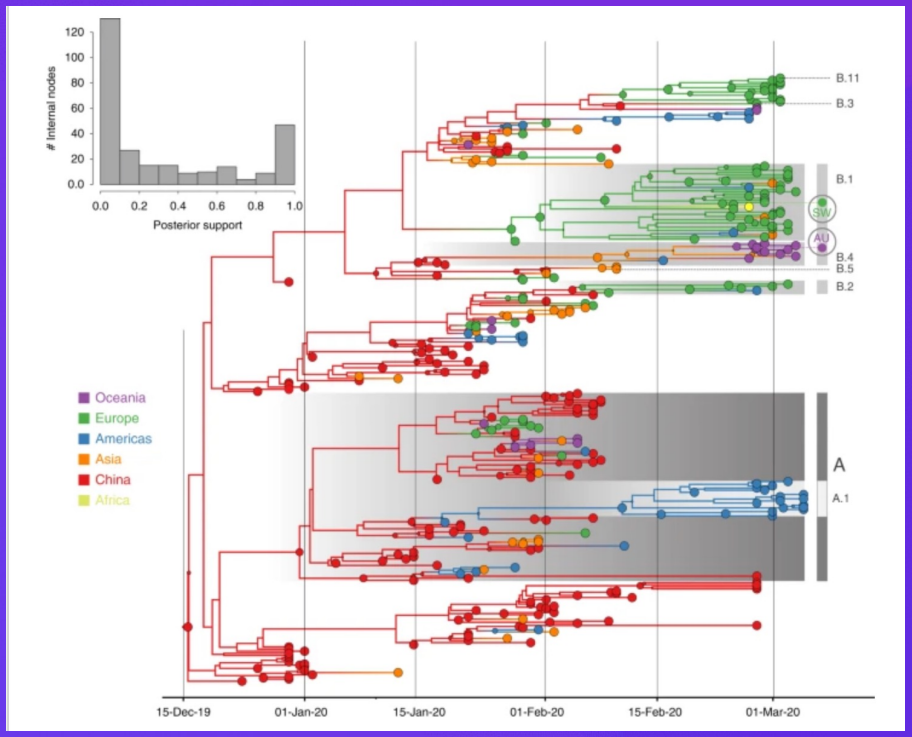
- In which location were past ancestors?
- Inferred using phylogenetic trees and the location data at tips using ancestral state reconstruction
- Employs probabilistic models to infer the most likely states of ancestral nodes, considering:
  - the observed character at tips
  - the topology of the phylogenetic tree
  - evolutionary processes governing the character evolution



4. Further applications: Phylogeography

Phylogeography

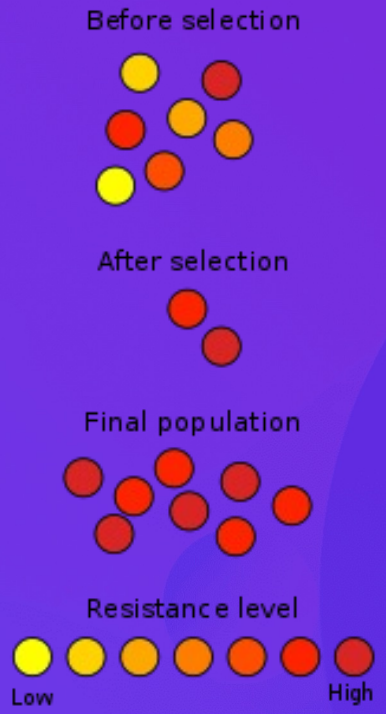
- Lemay et. al. 2020 reconstructed the location of the emergence of the COVID-19 pandemic
- Also inferred the dates in which there were first introductions to other regions



4. Further applications: Selection

Testing for sites under selection

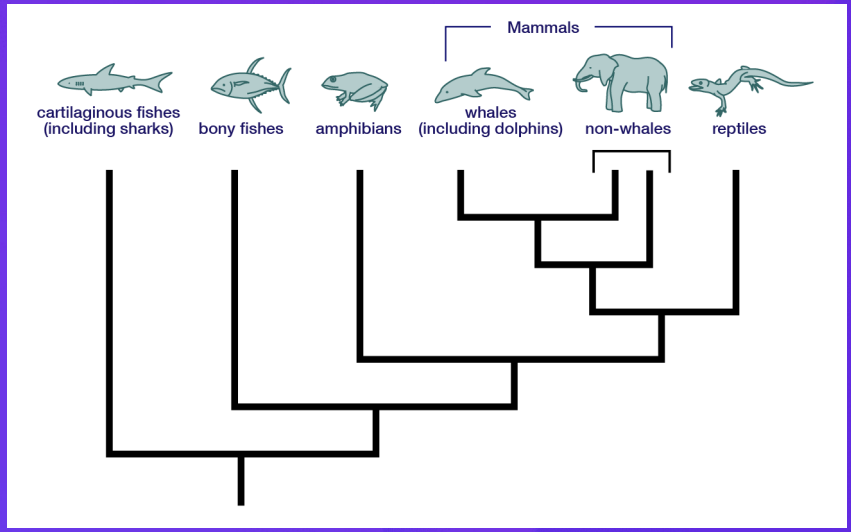
- Selection refers to the process by which certain heritable traits become more or less common in a population over time
- This process occurs because individuals with advantageous traits are more likely to survive and reproduce
- Selection can lead to adaptation and the evolution of new traits or the fixation of particular mutations



4. Further applications: Selection

Convergent evolution of genomic variants

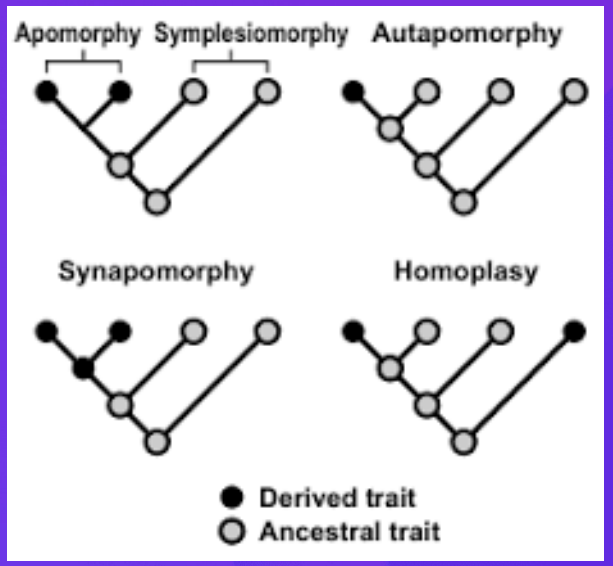
- The independent evolution of mutations or traits in distantly related individuals due to selective pressures or environmental constraints
- In genetics, sites under convergent evolution may represent instances of positive selection
- Mutations that enhance the fitness or adaptability of organisms to similar conditions



4. Further applications: Selection

Homoplasy

- Convergent evolution can cause homoplasies on a phylogenetic tree
- Shared traits or mutations among taxa that do not accurately reflect their evolutionary relationships
- Most methods use parsimony to infer the most likely ancestral states at internal nodes of the tree
- Identify branches where reversals or independent acquisitions occur, indicating homoplasy



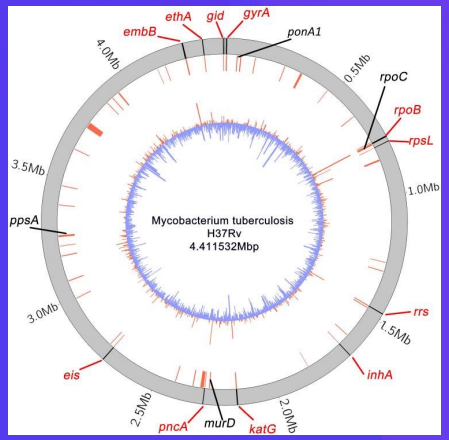
4. Further applications: Selection

Novel mutations in AMR strains

Nat Genet. Author manuscript; available in PMC 2014 Apr 1.
PMCID: PMC3887553
Published in final edited form as:
NIHMSID: NIHMS514223
Nat Genet. 2013 Oct; 45(10): 1183–1189.
PMID: 23995135
Published online 2013 Sep 1. doi: 10.1038/ng.2747

Genomic Analysis Identifies Targets of Convergent Positive Selection in Drug Resistant Mycobacterium *tuberculosis*

Maha R Farhat,[*,1,†] B Jesse Shapiro,[2,3,4,5,†] Karen J Kieser,[6] Razvan Sultana,[7] Karen R Jacobson,[8,9] Thomas C Victor,[9] Robin M Warren,[9] Elizabeth M Streicher,[9] Alistair Calver,[10] Alex Sloutsky,[11] Devinder Kaur,[11] Jamie E Posey,[12] Bonnie Plikaytis,[12] Marco R Oggioni,[13] Jennifer L Gardy,[14] James C Johnston,[15] Mabel Rodrigues,[16] Patrick K C Tang,[16] Midori Kato-Maeda,[17] Mark L Borowsky,[18,19] Bhavana Muddukrishna,[18,19] Barry N Kreiswirth,[20] Natalia Kurepina,[20] James Galagan,[21,22,23,2] Sebastien Gagneux,[24,25] Bruce Birren,[2] Eric J Rubin,[6] Eric S Lander,[2] Pardis C Sabeti,[2,3,4,6] and Megan Murray[*,26,27]

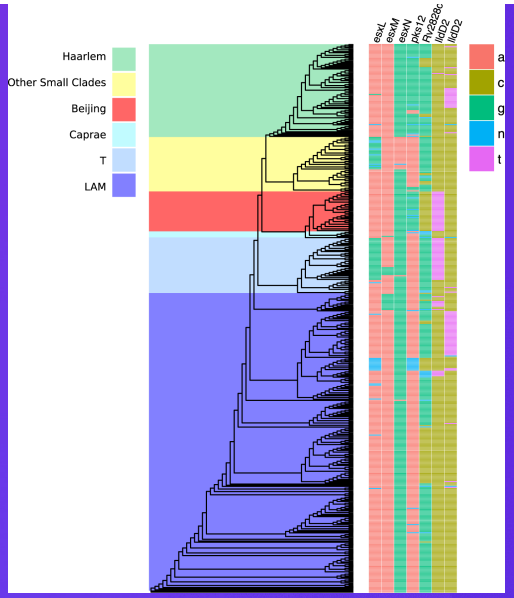


PLOS ONE

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

Convergent evolution and topologically disruptive polymorphisms among multidrug-resistant tuberculosis in Peru

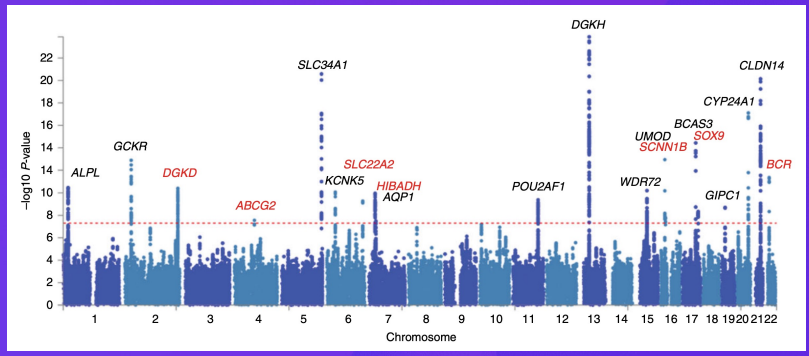
Louis Grandjean, Robert H. Gilman, Tomatada Iwamoto, Claudio U. Köser, Jorge Coronel, Mirko Zimic, M. Estee Török, Diepreye Ayabina, Michelle Kendall, Christophe Fraser, Simon Harris, Julian Parkhill, Sharon J. Peacock, David A. J. Moore, Caroline Colijn



4. Further applications: Association between traits and mutations

Genome Wide Association Studies

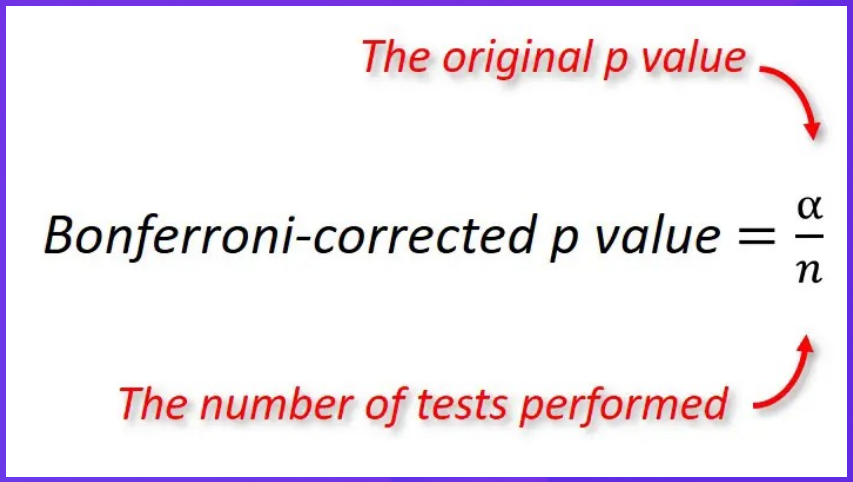
- GWAS is a test to identify genetic variants that are associated with a particular trait
- This may be mutations that cause antibiotic resistance, increase virulence or transmissibility, or evolve with host adaptation
- Advancements in sequencing technologies and bioinformatics have significantly enhanced the ability to conduct GWAS in microbial populations



4. Further applications: Association between traits and mutations

Genome Wide Association Studies

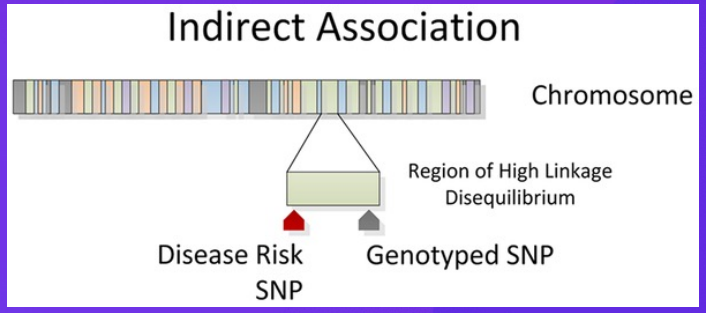
- GWAS statistically analyzes genetic variants to identify any correlations with specific traits or diseases
- Multiple testing correction (such as Bonferroni) is applied to account for potentially thousands of sites being tested
- Also need to account for population structure to remove the confounding effect of genetic substructure in the population



4. Further applications: Association between traits and mutations

Genome Wide Association Studies

- GWAS can be complicated in microbial populations for the following reasons:
    - Complex population structures, including clonal lineages, recombination events
    - Linkage disequilibrium, where alleles at different loci are inherited together due to limited recombination leading to spurious associations
    - Causal Inference, traits can be caused by multiple genes and environmental factors

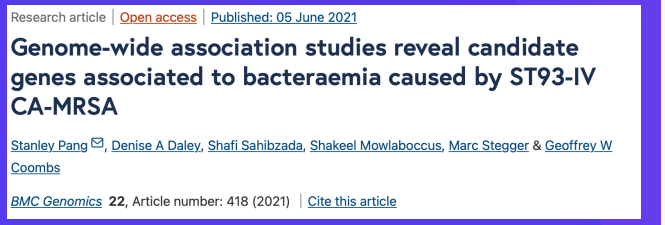


From Bush & Moore, 2012

4. Further applications: Association between traits and mutations

Genome Wide Association Studies

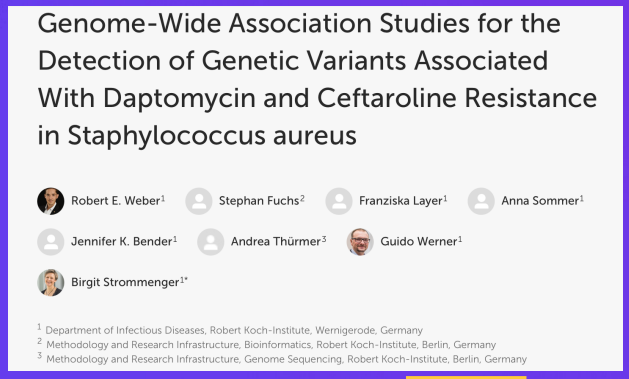
- GWAS has led to the discovery of novel variants associated with a variety of traits in a range of pathogens


Research article | Open access | Published: 05 June 2021

Genome-wide association studies reveal candidate genes associated to bacteraemia caused by ST93-IV CA-MRSA

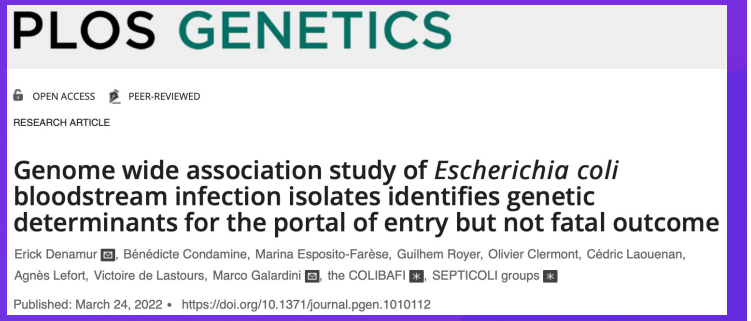
Stanley Pang, Denise A Daley, Shafi Sahibzada, Shakeel Mowlaboccus, Marc Stegger & Geoffrey W Coombs

BMC Genomics **22**, Article number: 418 (2021) | Cite this article


Genome-Wide Association Studies for the Detection of Genetic Variants Associated With Daptomycin and Ceftaroline Resistance in Staphylococcus aureus

Robert E. Weber[1] Stephan Fuchs[2] Franziska Layer[1] Anna Sommer[1]
Jennifer K. Bender[1] Andrea Thürmer[3] Guido Werner[1]
Birgit Strommenger[1*]

[1] Department of Infectious Diseases, Robert Koch-Institute, Wernigerode, Germany
[2] Methodology and Research Infrastructure, Bioinformatics, Robert Koch-Institute, Berlin, Germany
[3] Methodology and Research Infrastructure, Genome Sequencing, Robert Koch-Institute, Berlin, Germany

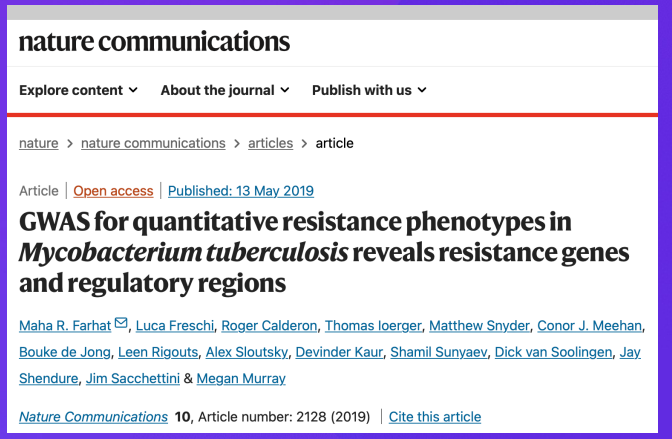

PLOS GENETICS

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

Genome wide association study of *Escherichia coli* bloodstream infection isolates identifies genetic determinants for the portal of entry but not fatal outcome

Erick Denamur, Bénédicte Condamine, Marina Esposito-Farèse, Guilhem Royer, Olivier Clermont, Cédric Laouenan, Agnès Lefort, Victoire de Lastours, Marco Galardini, the COLIBAFI, SEPTICOLI groups

Published: March 24, 2022 • https://doi.org/10.1371/journal.pgen.1010112


nature communications

Explore content About the journal Publish with us

nature > nature communications > articles > article

Article | Open access | Published: 13 May 2019

GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions

Maha R. Farhat, Luca Freschi, Roger Calderon, Thomas Ioerger, Matthew Snyder, Conor J. Meehan, Bouke de Jong, Leen Rigouts, Alex Sloutsky, Devinder Kaur, Shamil Sunyaev, Dick van Soolingen, Jay Shendure, Jim Sacchettini & Megan Murray

Nature Communications **10**, Article number: 2128 (2019) | Cite this article

The BEAST interactive session

1. Full Bayesian reconstruction of timed phylogenies from sequence data with BEAST2
2. Phylodynamic analysis with BEAST2

Link to practical: bensobkowiak.github.io/VanML