ИНСТИТУТ ПО МИКРОБИОЛОГИЯ
“СТЕФАН АНГЕЛОВ”
БЪЛГАРСКА АКАДЕМИЯ на НАУКИТЕ 1869
ЧЛЕН НА МЕЖДУНАРОДНАТА АСОЦИАЦИЯ НА ИНСТИТУТИТЕ ПАСТЬОР
international network Institut Pasteur

Приложение на пълното геномно секвениране и подходите на биоинформатиката за епидемиологичен анализ на туберкулозата при хора и животни

Гл.ас. д-р Виолета Вълчева

Департамент по Инфекциозна микробиология
Лаборатория по зоонози и бактериална вирулентност

МИНИСТЕРСТВО НА ЗЕМЕДЕЛИЕТО, ХРАНИТЕ И ГОРИТЕ
ЦЕНТЪР ЗА ОЦЕНКА НА РИСКА ПО ХРАНИТЕЛНАТА ВЕРИГА

НАУЧЕН КОЛОКВИУМ, София, 20.11.2019г.

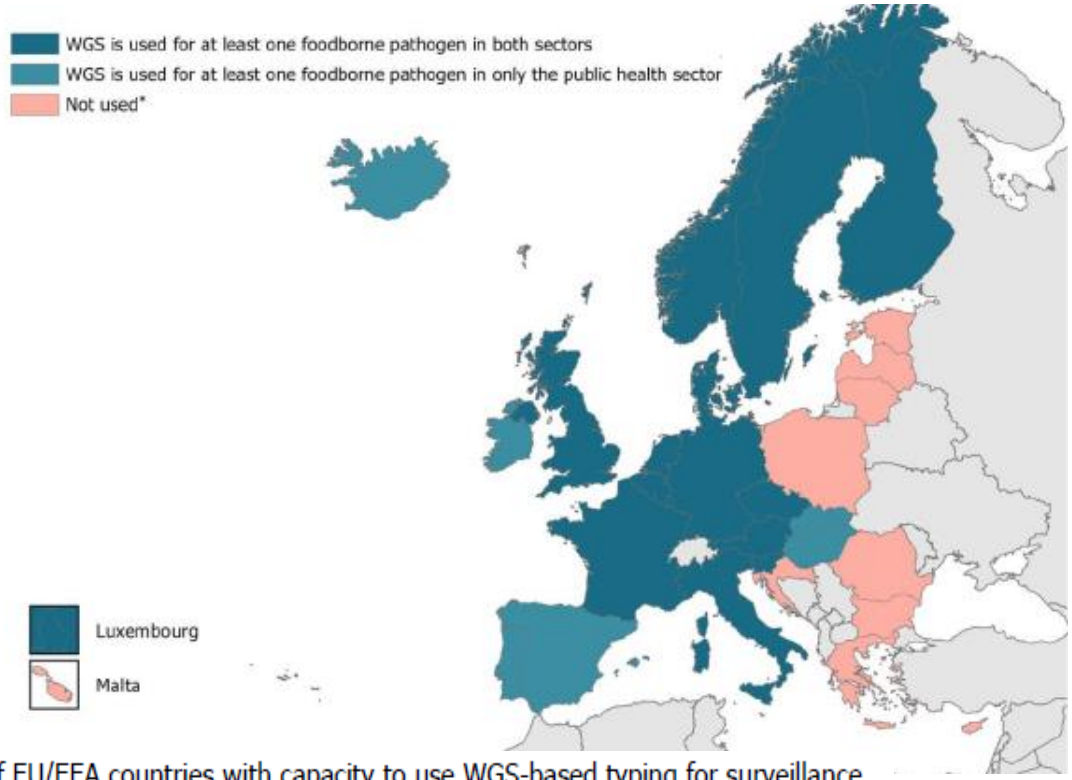
„Приложение на методите на пълен геномен секвентен анализ и ДНК-баркодинг при диагностиката и контрола на патогенни микроорганизми във ветеринарната медицина и общественото здраве – Готова ли е България за предизвикателствата на бъдещето?“

- В ерата на **бурно развиващите се геномни технологии**, специалистите по молекулярна диагностика ежедневно се сблъскват с редица предизвикателства

- **Нашите познания** за човешкия, бактериалния или вирусен геном, **за взаимодействието между генетични и епигенетични фактори**, за това **как да интерпретираме всеки резултат** и да **анализираме генотип-фенотипните корелации** се променят **ежедневно**

- Това до голяма степен е свързано с **огромните възможности на съвременните технологии,** което също **изисква големи усилия** от страна на специалистите **постоянно да надграждат квалификационните си умения и да бъдат в крак с технологиите от новата генерация**

Приоритетни патогени, определени от ECDC за интегриране на молекулярно/геномно секвениране в постоянния (рутинен) надзор на инфекциозни заболявания в страните от EU 2019-2021



Number of EU/EEA countries with capacity to use WGS-based typing for surveillance and/or outbreak investigations and data completeness, by sector

	Development of genomic typing
ECDC programme: Food- and Waterborne Diseases and	
Listeria monocytogenes	Integration of PFGE/WGS quality collection and analysis system ext
	Development and testing of WGS core genome allele nomenclature 18)
	Implementation of WGS-enhance
Salmonella enterica	Integration of PFGE/MLVA/WGS analysis system extended to EFSA
	WGS: focus on data sharing and testing of open access analysis pl
	Contribution to design and testing nomenclature
Shiga toxin-producing *E. coli* (STEC)	Integration of PFGE/WGS quality collection and analysis system ext
	WGS: focus on data sharing and testing of open access analysis pl
	Contribution to design and testing nomenclature [14]
ECDC programme: Emerging and Vector-Borne Diseases	
West Nile virus	Development of molecular typing
ECDC programme: HIV, AIDS, STIs and Viral Hepatitis (H	
Antibiotic resistant *Neisseria gonorrhoeae*	Development of business case ba (2016)
	Pilot WGS-based EU survey based
HIV drug resistance	Map capacity among EU/EEA cou
	Development of antiviral resistanc
Hepatitis C virus	Map capacity
	Development of antiviral resistanc
ECDC programme: Influenza and other Respiratory Virus	
Influenza virus	Implementation of business case pandemic (including zoonotic) str
	Implementation of sequence-base
ECDC programme: Tuberculosis (TB)	
Multidrug-resistant *Mycobacterium tuberculosis*	Maintenance of MIRU-VNTR exter and analysis (2016–18)
	WGS: monitoring of progress of in identification, drug resistance pre
ECDC programme: Vaccine-Preventable Diseases (VPD)	
Neisseria meningitidis	Development of business case ba analysis and pilot survey (2016)
	Implementation of case based-W
	Implementation of WGS based su

Източник: European Centre for Disease Prevention and Control. ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations –2019–2021. Stockholm: ECDC; 2019.

КАК ПЪЛНОТО ГЕНОМНО СЕКВЕНИРАНЕ И БИОИНФОРМАТИКАТА МОГАТ ДА ПРОМЕНЯТ ЖИВОТА НИ

- Възможност да се използва богатството на новата информация и да се разработят нови техники и методология за интелигентен информационен анализ, обработка и откриване на знания

- Тези ресурси ще позволят ефективен достъп на учените до геномни данни, позволяващи нови типове анализи чрез компютърно моделиране и експериментално потвърждаване на комплексни геномнни мрежи, които обуславят причините за много болести при хора и животни

- Геномен аналитичен подход, който позволява изследване на генома, интерпретация на получената необятна база данни и на тази основа молекулярно-генетично верифициране на диагнозата при хора и животни

- Нови биомаркери навлизат в клиничната практика, създаващи нови предизвикателства за медици, генетици, биолози и биохимици.

European Commission › Strategy › Digital Single Market › News ›

Digital Single Market

DIGIBYTE | 10 April 2018

EU countries will cooperate in linking genomic databases across borders

13 European countries have signed a declaration for delivering cross-border access to their genomic information. This is a game changer for European health research and clinical practice: sharing more genomic data will improve understanding and prevention of disease, allowing for more personalised treatments (and targeted drug prescription), in particular for rare diseases, cancer and brain related diseases.

The signatory Member States commit to collaborate on the secure and authorised access to national and regional banks of genetic data

https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders

2018 г. – България е присъединена към проекта за Единен цифров пазар

«Достъп до най-малко 1 милион геноми в EU до 2022 г.«
който включва следните основни моменти

✓ Изработване на съвместна декларация за развитие на съществуващи национални и регионални инициативи за персонализирана медицина

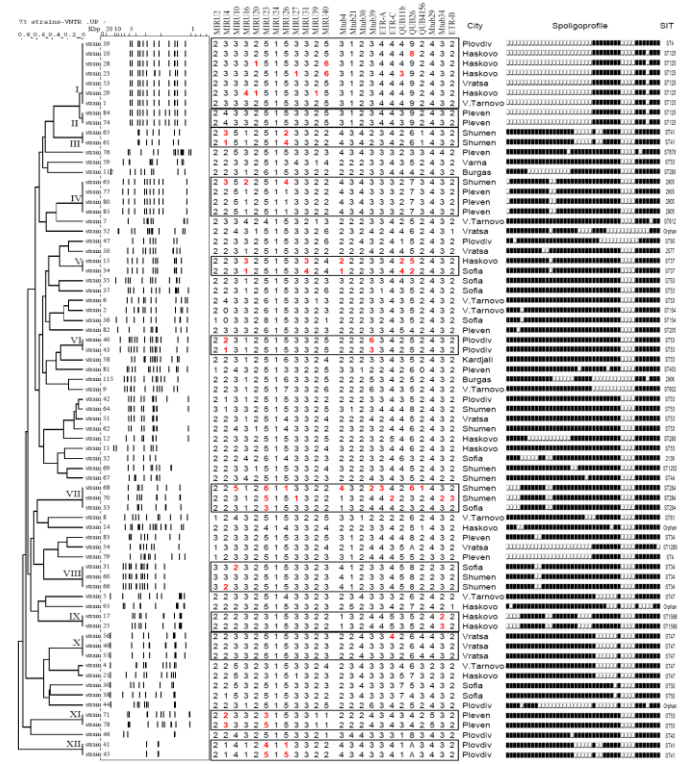
✓ Укрепване на сътрудничеството между страните от EU и на Европейското икономическо пространство

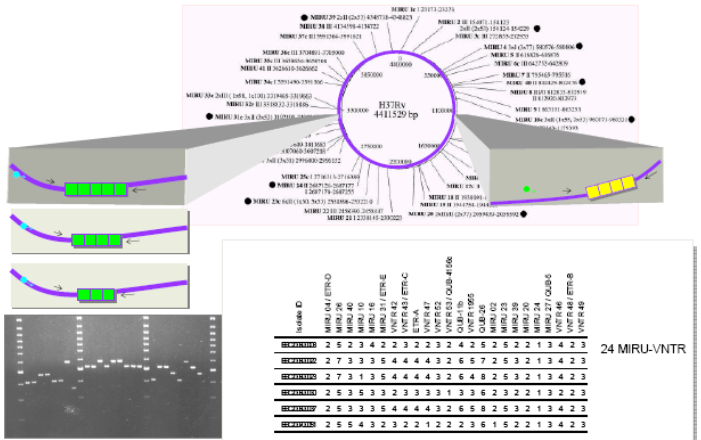
✓ **Реализиране на ползите от трансграничния достъп до и обмен на геномни данни**

✓ Постигане на напредък в персонализираната медицина

Въвеждането на молекулярно-генетичните маркери
за сравнителни генетични изследвания и
щамова диференциация
поставиха основата на **нова концепция** за:

- ❑ Проследяване динамиката на епидемиологичния и епизоотологичния процес
- ❑ Трансмисията, еволюцията и филогенетиката на бактерии и вируси
- ❑ Осъществяване на надзора и оценка на риска от възникването и разпространението на инфекциозните заболявания при хора и животни

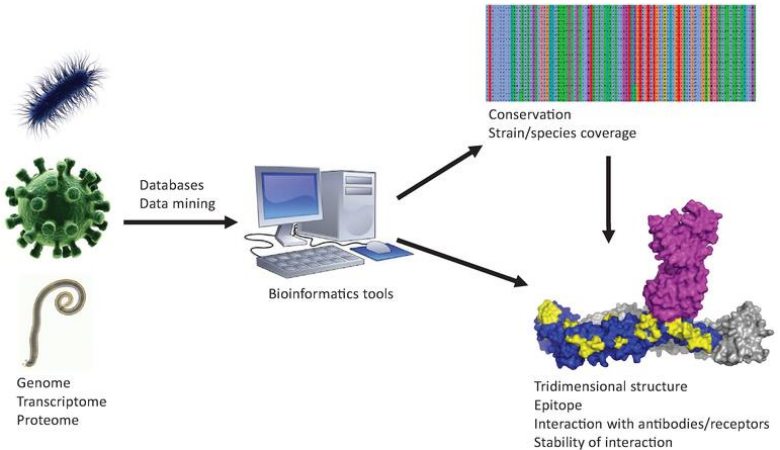




Реконструкцията на еволюцията обаче изисква модели, алгоритми, статистически и изчислителни методи със все по-голяма сложност

- С въвеждането на пълното секвениране на ДНК (WGS) или секвениране от следващо поколение (NGS), вече е възможно секвениране на цели геноми на патогени в реално време

- За разлика от генотипирането, при което се използва само малка част от генома на патогена за определяне на филогенетични взаимоотношения, целият геном на патогена може да се използва за отчитане динамиката на предаване на епидемията с много по-големи подробности.



Технологията за секвенирането значително разшири кръгозора ни за това как биологичната информация се кодира в последователността на ДНК,

но също така доведе до нови въпроси:

- Какво е значението на финитe промени в последователността, особено в некодиращите региони?
- Кога взаимодействието на регулаторните елементи влияе върху експресията на гена?
- Как информацията, базирана на последствия - епигенетична информация - допринася за програмирането на развитието и диференцирането на клетките?
- Как отговорите на тези въпроси могат да бъдат от полза за приложните науки?

Характеризирането на тази информация **изисква изследователски инструменти от ново поколение,** които дават **възможност за гъвкавост на приложението и събиране на количествени данни**

Биоинформатиката се занимава основно с:

- Моделиране на биологични системи и функции;
- Анализиране на лабораторни данни

(напр. появата и предаването на мутации за резистентност);

- Генериране на модели на база натрупани данни от експерименти;
- Изследване на нови данни с помощта на математически модели;
- Разпознаване на мотиви в експериментални данни;
- Предсказване на функции на гени и белтъци;
- *In silico* експерименти

Основни средства и методи:

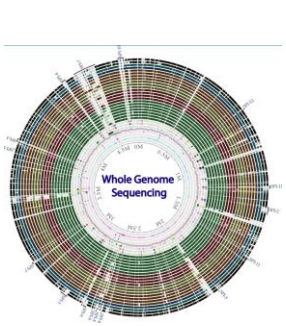
- ДНК
- Протеинни секвенции и структури
- Геноми и протеоми
- Бази данни и извличане на информация
- WWW и компютърното програмиране
- Архиви с бази данни, както и начини

за извличане на информация от тях

WGS / NGS

процес на генериране на файлове, съдържащи ДНК последователности (секвенции)

за постоянно /рутинно приложение изисква



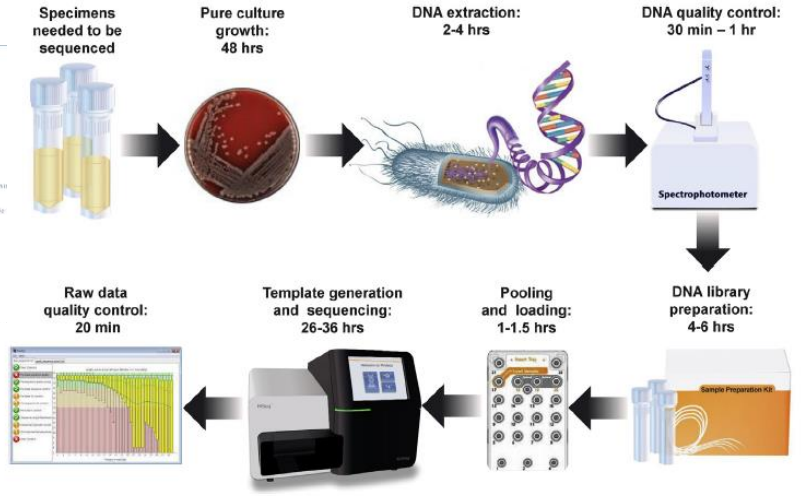


Fig. 2. Typical whole genome sequencing workflow in a clinical or public health laboratory.

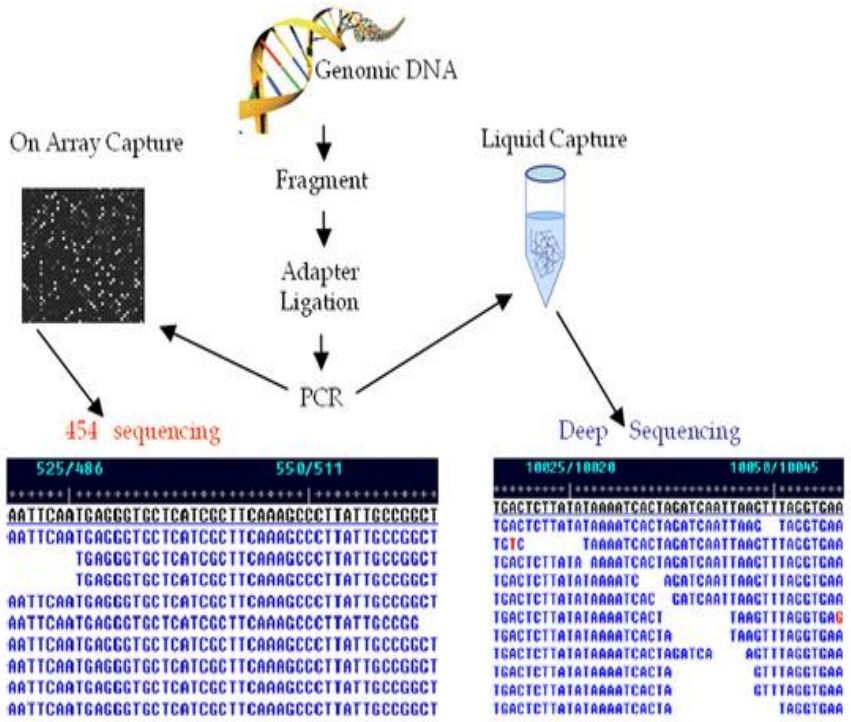
- **Финансиране –** Национални и международни програми

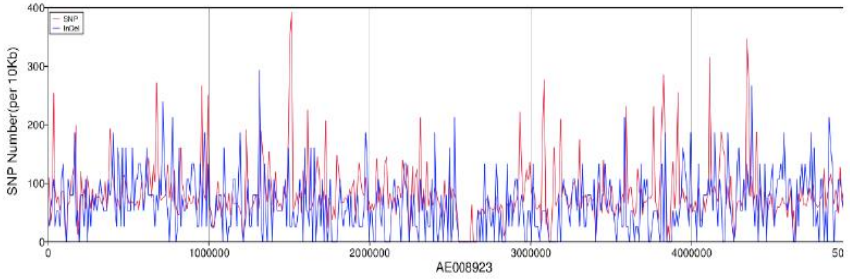
- Налични щамове
- WGS платформи/ апаратура
- Биоинформатични софтуерни програми, базаданни
- Експерти, които да могат да интерпретират геномните данни

✓ Приготвяне на ДНК библиотеки за секвениране

✓ Платформите за NGS, продуцират „секвенционни прочити” на
- къси ДНК фрагменти (35-100bp) - Illumina и ABI SOLiD
- по-дълги фрагменти (250-500bp) - Roche 454 GS FLX

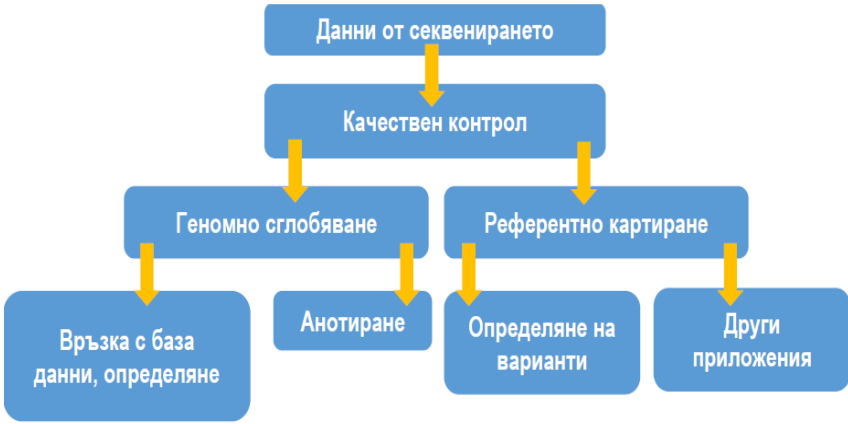
Платформа	Предимства	Недостатъци
Roche/454's GS FLX Titanium	По-дълги секвенции, Подобряване на възможността за разчитане на хомополимерни участъци, бързина	Скъпоструващи консумативи, висока честота на грешки в хомополимерни участъци
Illumina/ Solexa's GAII	Най-широко използваната понастоящем платформа	Ниско ниво на мултиплексиране на пробите – едновременен анализ
Life/APG's SOLiD 3	Двубазовият код позволява коригиране на на грешките	Продължително време на секвениране
Helicos BioSciences HeliScope	Отсъствие на отклонение в представителността на матриците за геномни и секвенционни анализи	Високо ниво на грешки в сравнение с обратимите терминатори или химични реагенти използвани в другите платформи
Pacific Biosciences (target release: 2010)	Има най- висок потенциал за получаване на секвенции с дължина над 1кб	Най-високо ниво на грешки в сравнение с другите NGS химични реакции





Биоинформатичен работен подход за обработка и анализ на данни от геномното секвениране

в 8 стъпки , означен като “*pipeline*” последователна обработка на данните:



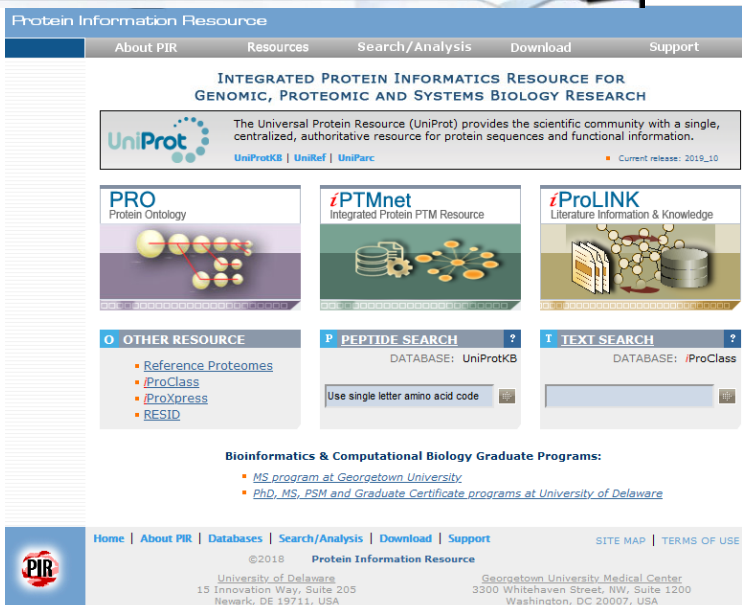
1) Секвенционни формати и контрол на качеството на прочетената секвенция/цял геном;
2) Изравняване на секвенциите, получени от секвенирането с референтни последователности;
3) Сравняване, четене и картографиране на референтната ДНК последователност и тази, която е секвенирана;
4) Вариации в нуклетидните бази т.е откриване на еднонуклеотиден полиморфизъм (SNP), мутации (малка инсерция или делеция (indels)) в секвенираната верига;
5) *De novo* „сглобяване“на секвенциите/ цял геном;
6) Геномна анотация - процес на идентифициране на местонахождението на гените и всички кодиращи региони в генома и определяне на функцията на тези гени т.е анализ на генната експресия; регулация; метаболитни процеси, откриване на нови лекарства;
7) Създаване на филогенетично дърво;
8) Филогенетичен анализ;

Биологични Бази Данни
Съдържание, Приложение

- Множество библиотеки на информация за науките за живота, които са събрани от научни експерименти, публикувана литература, високопропускателна екпериментална технология и изчислителни анализи

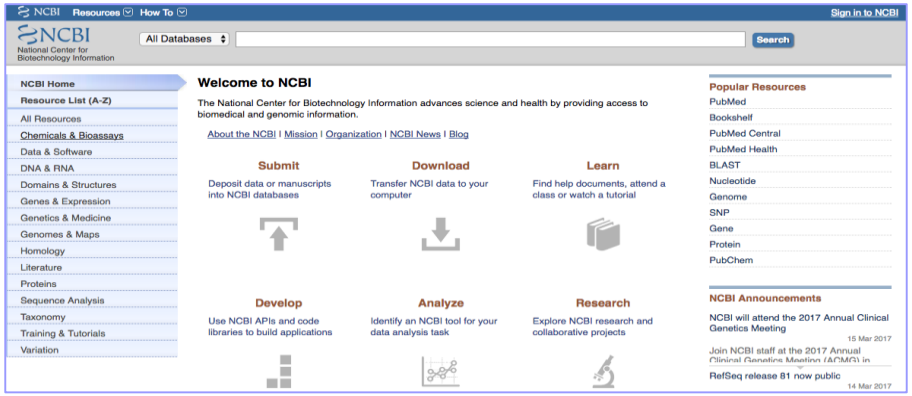
- Съдържат информация от изследователски области, включително геномика, протеомика, метаболомика

- Важен инструмент, който помага на учените да разберат и да обяснят множеството биологични явления, от структурата на биомолекулите и тяхното взаимодействие до целият метаболизъм на организмите и разбирането на еволюцията на видовете



Биологични бази данни

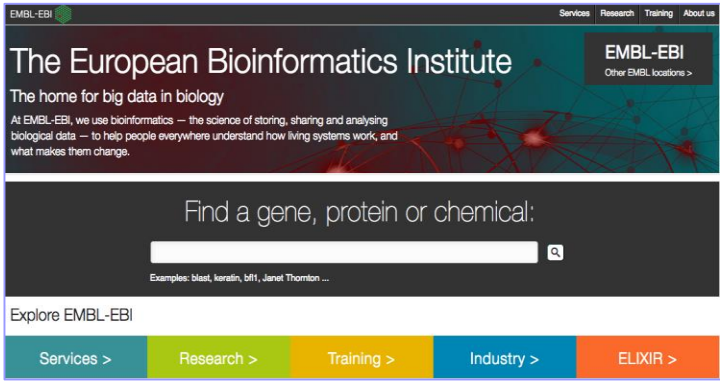
- Типове биологични бази данни според съдържанието им:
  - Първични бази данни
    - GenBank
    - Protein Data Bank (PDB)
  - Вторични бази данни
    - SWISS-Prot
    - Protein Information Resources – PIR
  - Специализирани
    - https://www.hiv.lanl.gov/content/index - база данни за секвенции на HIV
    - https://www.gisaid.org/ ; https://www.fludb.org/ - бази за секвенции на грипни вируси
    - http://symbiont.iis.sinica.edu.tw/evidence/static/html/ - база за секвенции на ентеровируси



Първични бази данни

- GenBank
  - (National Center for Biotechnology Information (NCBI): http://www.ncbi.nlm.nih.gov
- European Molecular Biology Laboratory (EMBL):
  - http://embl.org/
- DNA Databank of Japan (DDBJ):
  - http://www.ddbj.nig.ac.jp/
- Протеинна банка
  - http://www.wwpdb.org/

Биоинформатични програми в EMBL-EBI



Sequence Similarity Search/Търсене на подобие на последователностите

- NCBI BLAST+, FASTA, PSI-Search/BLAST, HMMER 3 hmmscan, HMMER 3 phmmer и др .

Pairwise and Multiple Sequence Alignment/ популярни методи за подравняване по двойки и последователно множествено подравняване

- Clustal Omega, MAFFT, MUSCLE, T-coffee, Kalign, Needle, Water, Matcher, Lalign, и др.

Protein Functional Analysis/ Анализ на белтъчната функция

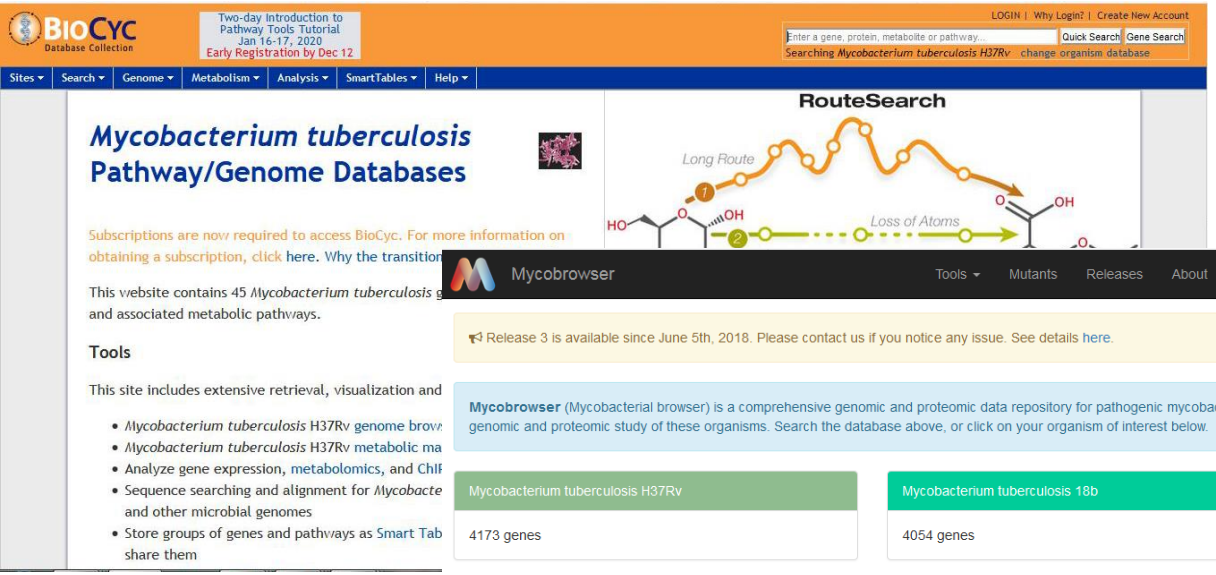
- InterProScan5, PfamScan, Phobius, Infernal cmscan,и др.

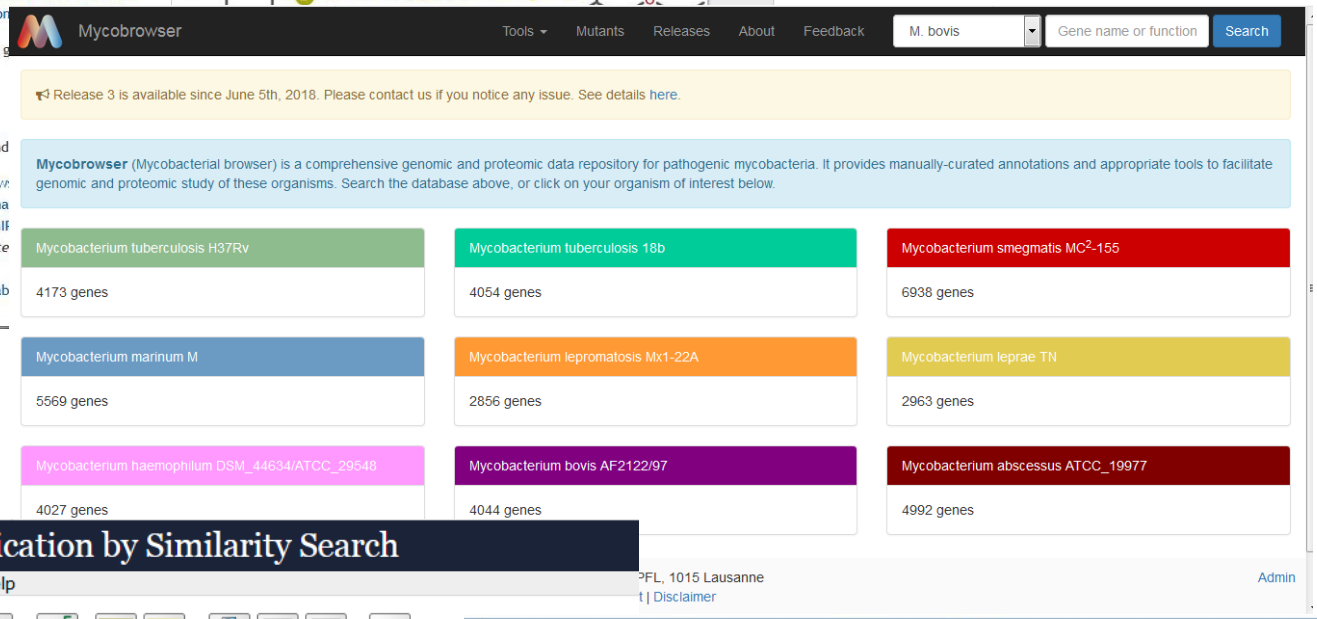
Други секвенционни приложения и статистика

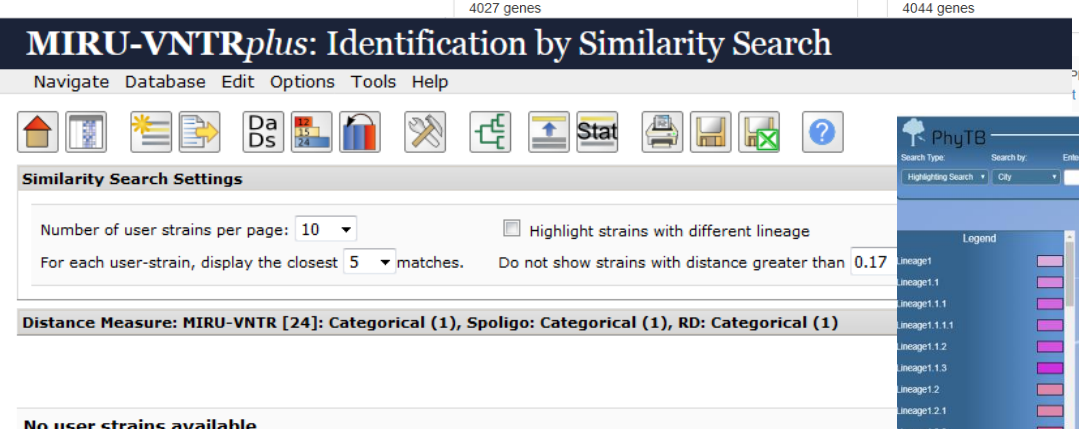
- EMBOSS приложения, Simple Phylogeny, и др.

https://www.ebi.ac.uk/services

База данни за микобактерии
подробна информация за гени и техните продукти







Молекулярна филогенетика и анализ на данни за епидемиологичен анализ на туберкулозата при хора и животни

- Защо са ни нужни филогенетичните дървета от биологична гледна точка
  - Класификация на видовете в подгрупи
  - Наследствени реконструкции
  - Изучаване на древни геноми
  - Филогеография и динамика на епидемии

Как да си постоим едно филогенетично дърво

- ❑ Трябва да си изберем молекулярен маркер
- ❑ Да си подравним последователностите
- ❑ Да си определим метода за построяване на филогенетичното дърво

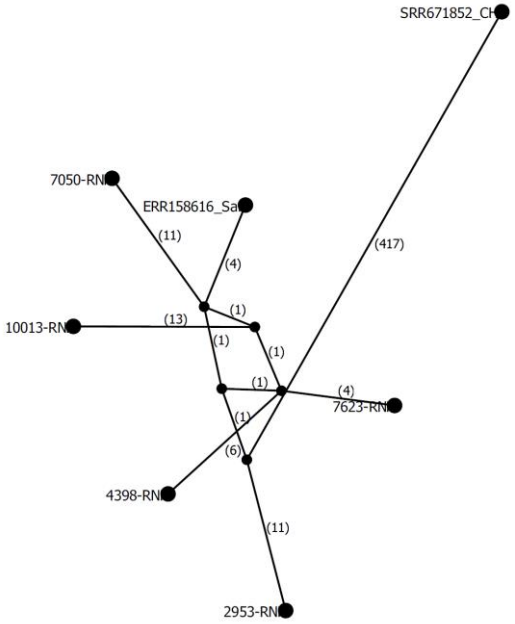
✓ чрез определяне на разстоянията “Distance methods” – UPGMA и Neighbour-Joining алгоритми - определяне на разстоянието между клоновете и двойките разстояния/ Матричен подход за определяне дължините на разклоненията

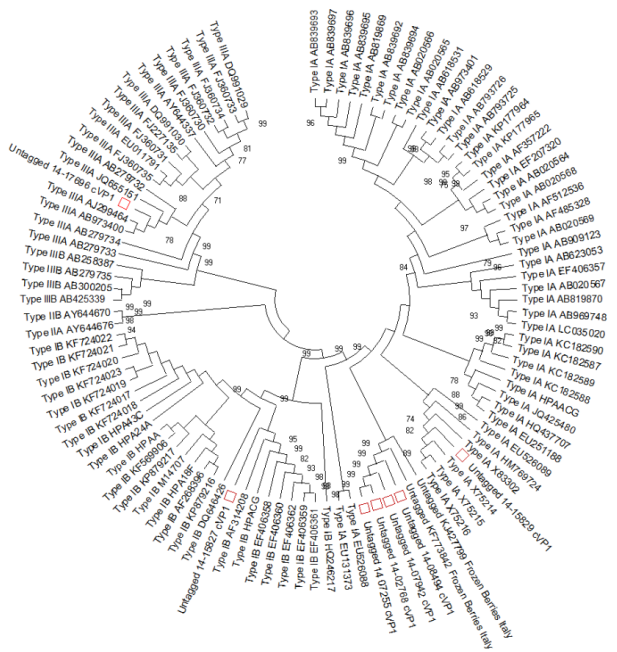
✓ чрез опростеност/парсимония “Maximum Parsimony” - което изисква най-малко еволюционни стъпки или което най-кратко обяснява данните

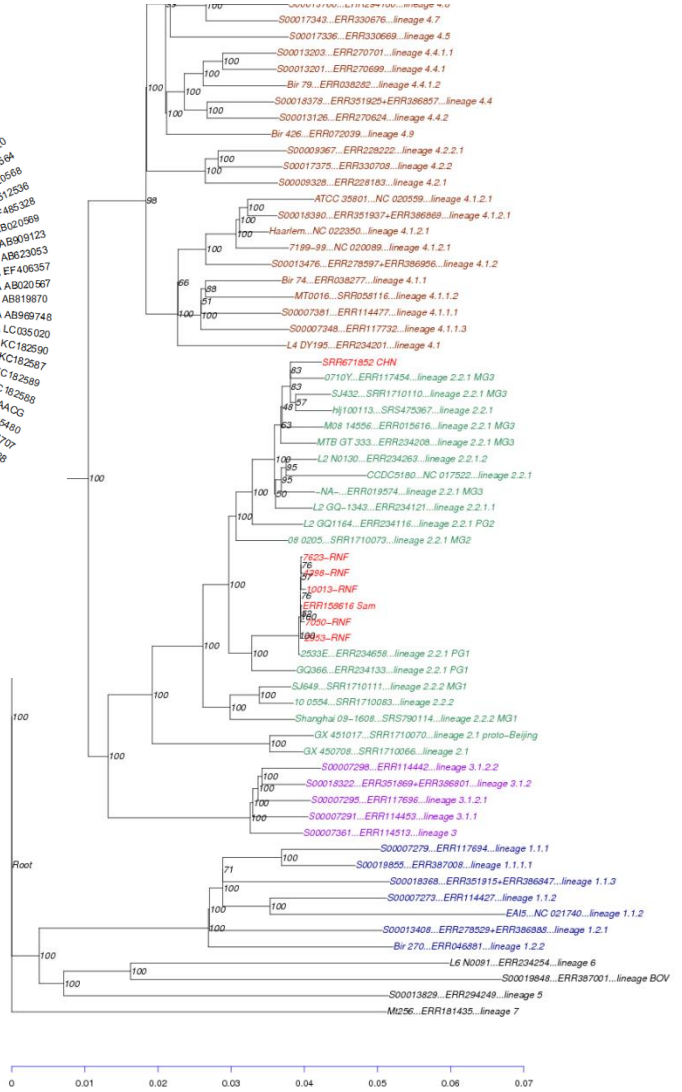
✓ Чрез еволюционни модели - “Maximum likelihood”/Методи на максимална вероятност- статистическа обработка, обобщение на най-малките квадрати

✓Чрез математически модели - преподреждания на секвенциите
- “Markov chain Monte Carlo (MCMC) sampling” - клас от алгоритми за разпределение на вероятностите, чрез конструиране на верига от няколко стъпки
- “Bayesian inference” - метод на статистическа обработка, чрез използване на хипотеза, с повече доказателства или информация за вероятностите

Как изглежда едно филогенетично дърво

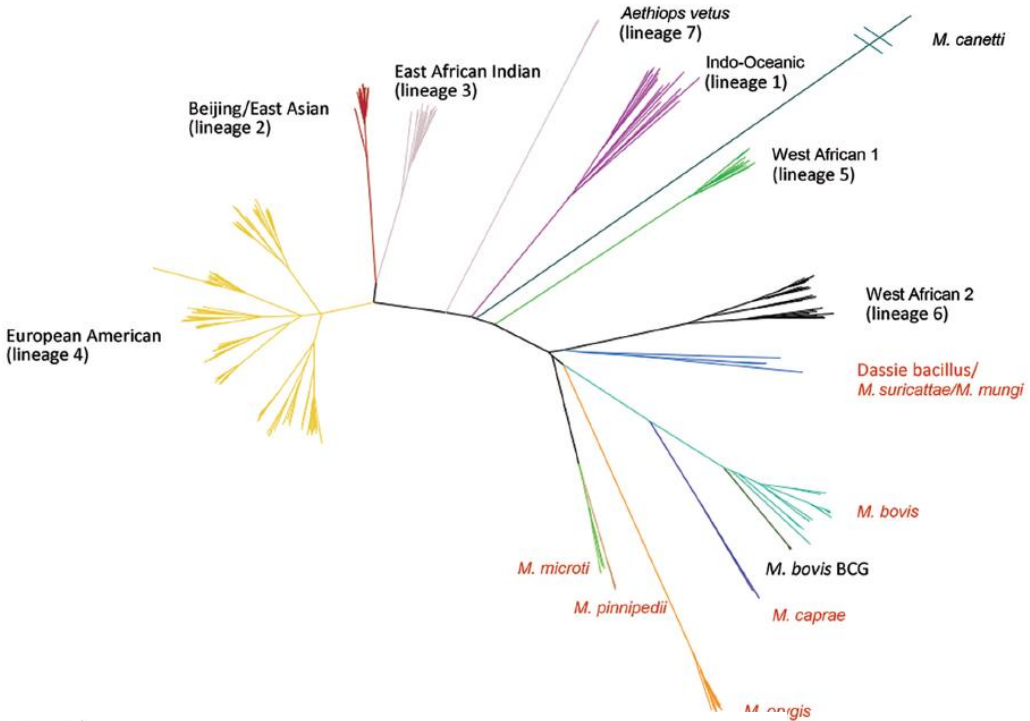






- Филогенетичното дърво е диаграма на разклоняване, която показва изведените еволюционни връзки между различни биологични видове или други обекти - тяхната филогения - въз основа на приликите и разликите в техните физически или генетични характеристики

Идентификация на видове/ подвидове и генотипи в *Mycobacterium tuberculosis* complex след WGS на 323 проби



- базира се на множество SNP
- (под) видовете в комплекса могат да бъдат ефективно разпознати
- няма нужда да се използват други методи за идентифициране на (под) видове и генотипни семейства като Beijing

PLOS ONE

RESEARCH ARTICLE

Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study

Rana Jajou[1]*, Albert de Neeling[1], Rianne van Hunen[1,2], Gerard de Vries[1,2], Henrieke Schimmel[1], Arnout Mulder[1], Richard Anthony[1], Wim van der Hoek[1], Dick van Soolingen[1,3]

1 National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, 2 KNCV Tuberculosis Foundation, The Hague, The Netherlands, 3 Radboud University Medical Centre, Department of Medical Microbiology, Nijmegen, The Netherlands

* rana.jajou@rivm.nl

Check for updates

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 November 2018
doi: 10.3389/fmicb.2018.02820

A New Phylogenetic Framework for the Animal-Adapted *Mycobacterium tuberculosis* Complex

Daniela Brites[1,2], Chloé Loiseau[1,2], Fabrizio Menardo[1,2], Sonia Borrell[1,2], Maria Beatrice Boniotti[3], Robin Warren[4], Anzaan Dippenaar[4], Sven David Charles Parsons[4], Christian Beisel[5], Marcel A. Behr[6], Janet A. Fyfe[7], Mireia Coscolla[1†] and Sebastien Gagneux[1,2†]*

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland, [2] University of Basel, Basel, Switzerland, [3] Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna: Centro Nazionale di Referenza per la Tubercolosi Bovina, Brescia, Italy, [4] SAMRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa, [5] Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland, [6] McGill International TB Centre, Infectious Diseases and Immunity in Global Health, McGill University Health Centre and Research Institute, Montréal, QC, Canada, [7] Mycobacterium Reference Laboratory, Victoria Infectious Diseases Reference Laboratory, Peter Doherty Institute, Melbourne, VIC, Australia, [8] Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

OPEN ACCESS

Edited by:
Haiwei Luo,
The Chinese University of Hong Kong, Hong Kong

Reviewed by:
Ana M. S. Guimaraes,
Universidade de São Paulo, Brazil
Djaltou Aboubaker Osman,
Center of Study and Research

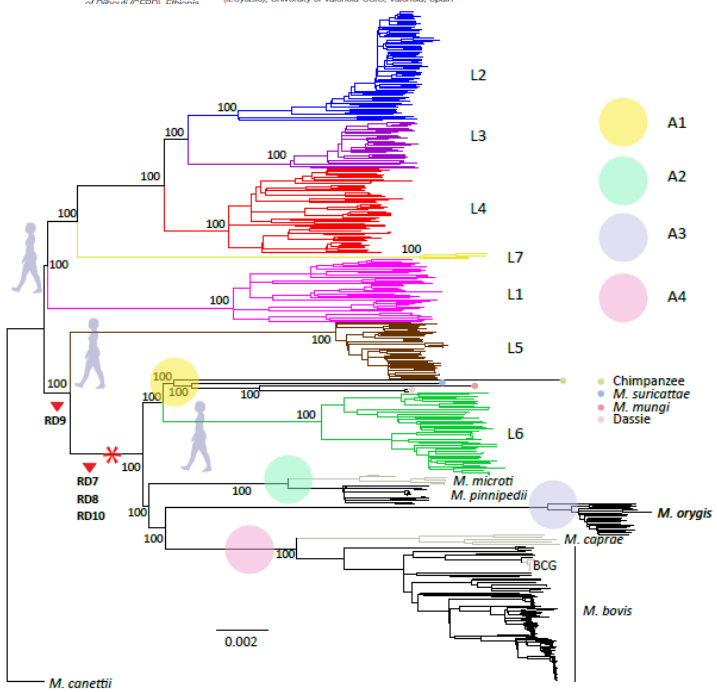


FIGURE 1 | Maximum Likelihood topology of 322 human-adapted and 529 animal-adapted MTBC members. Branch lengths are proportional to nucleotide substitutions and the topology is rooted with *Mycobacterium canettii*. Support values correspond to bootstrap values. Main large deletions defining the animal-adapted MTBC are indicated by red arrow heads. The asterisk marks host range expansion within the MTBC.

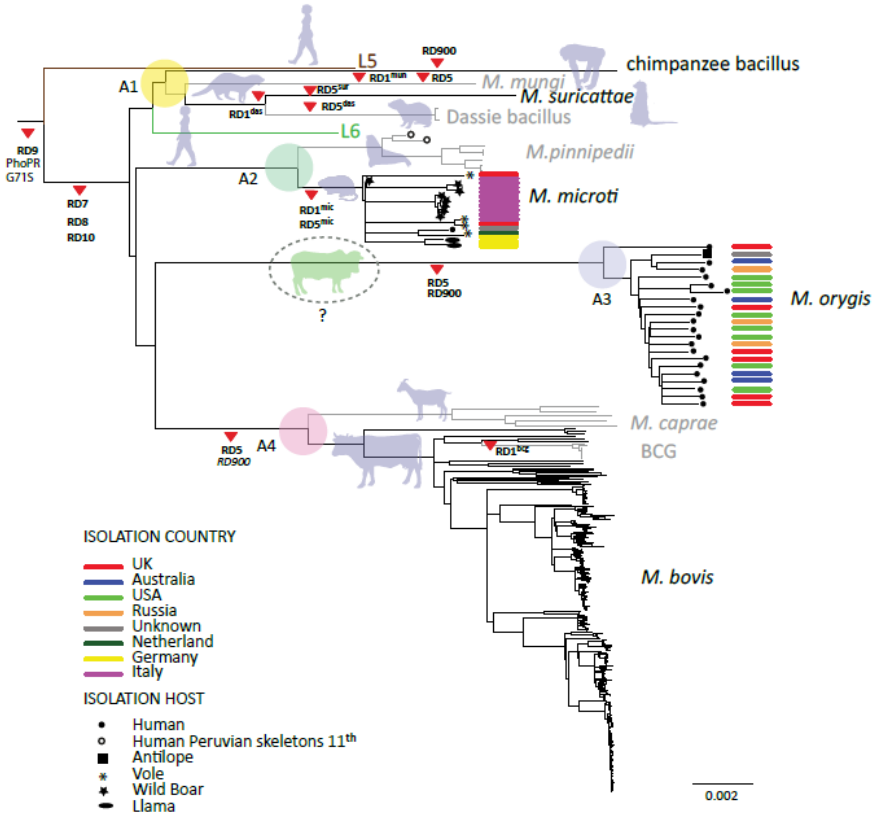
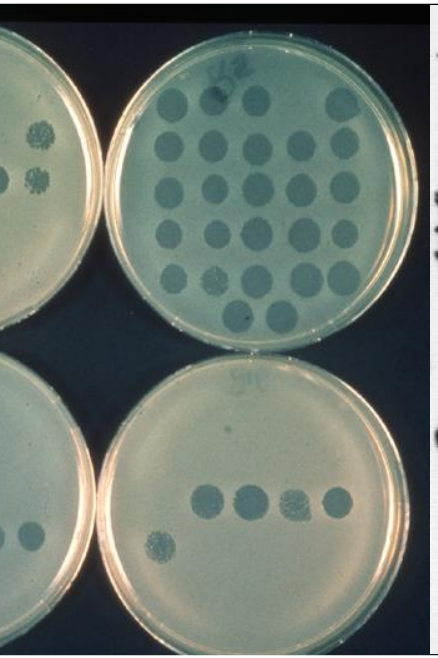
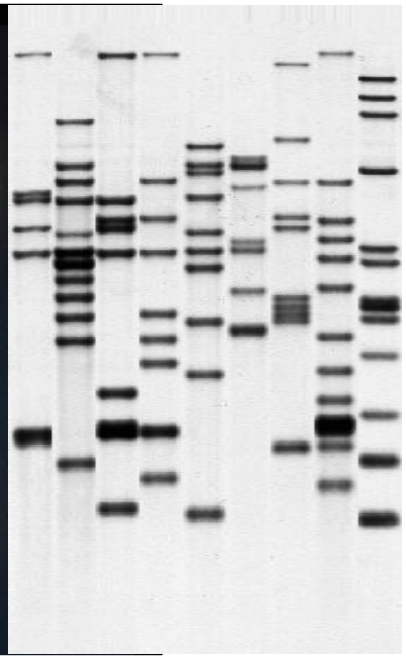
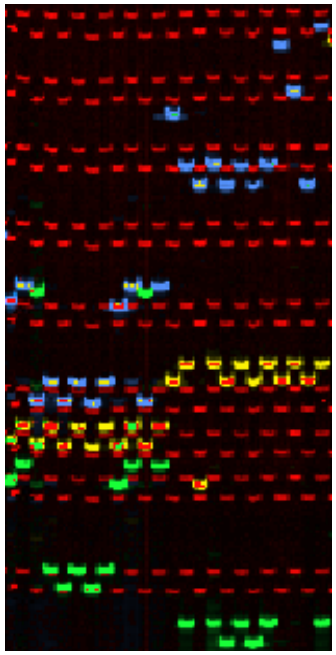
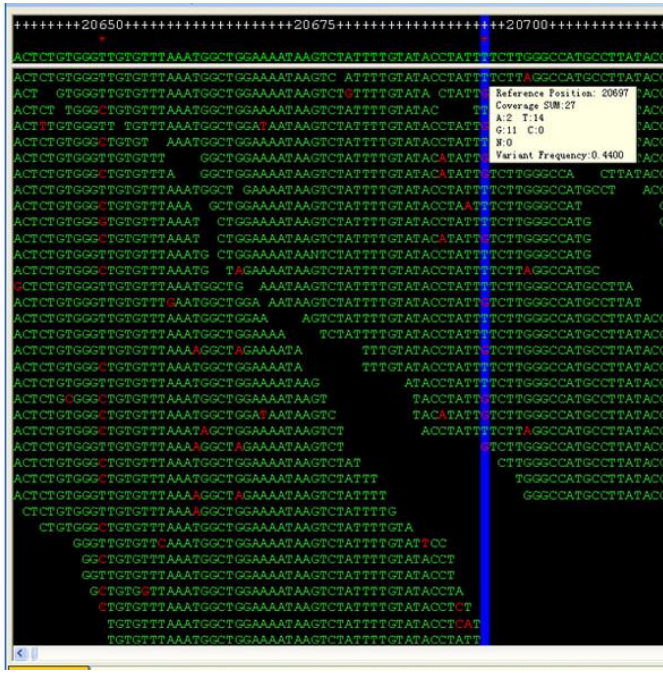
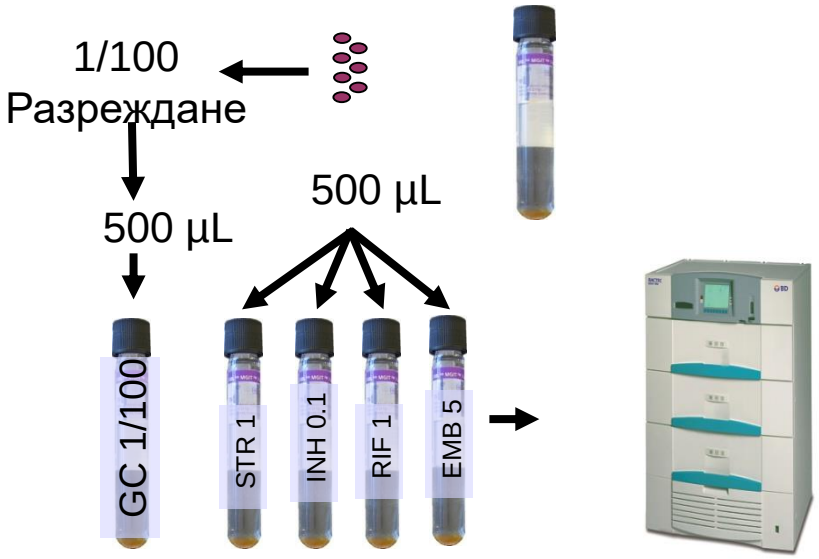


FIGURE 2 | Topology showed in **Figure 1** after collapsing all human-adapted branches. Branch lengths are proportional to nucleotide substitutions and the topology is rooted with *Mycobacterium canettii*. Support values are those of **Figure 1**. Main large deletions discussed in the text are indicated by red arrow heads and RD specific nomenclatures are indicated when available (Brodin et al., 2002; Mostowy et al., 2004; Alexander et al., 2010; Parsons et al., 2013; Dippenaar et al., 2015). Deletions which are polymorphic in terms of their presence or absence within main clades are indicated in italics.

Развитие на методите за проследяване на епидемиологията на туберкулозата

1970	1993	2009	2016
			
phage typing	IS*6110* RFLP typing	24 locus VNTR typing	whole genome sequencing

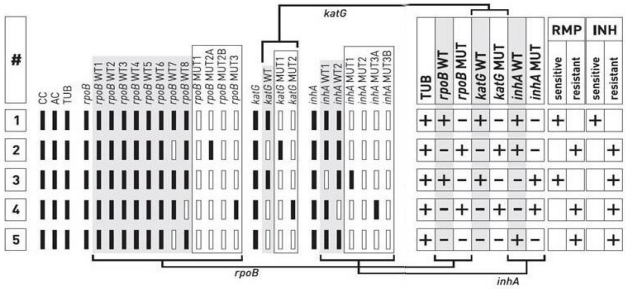
MGIT метод за фенотипна резистентност



Отнема време, не винаги е надеждно ...

ДНК метод (Reverse Line Blot) за RIF/INH резистентност

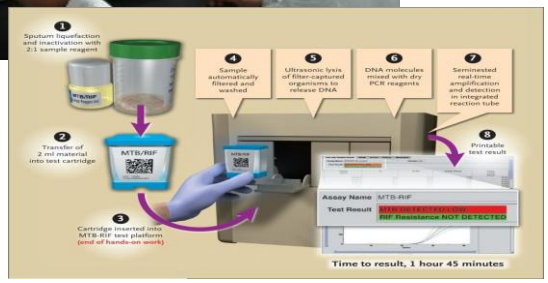
- Амплификация на *katG*, *inhA*, *rpoB* гени
- Хибридизация върху стрипове с див тип и мутантни проби
- надеждно прогнозиране на резистентност
- Обаче, само няколко мутации могат да се установят
- negative predictive value е ниска



GeneXpert® MTB/RIF тест

2ч от пробовземане до резултат

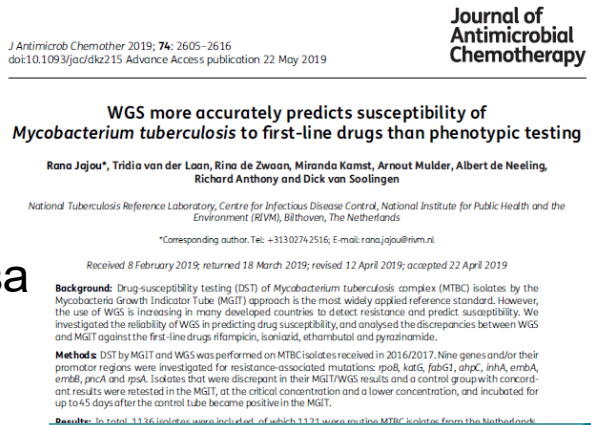
- Интегрирана и автоматизирана подготовка на проби, откриване чрез PCR система в реално време



- Крайният продукт е напълно автоматизиран, затворен, изискващ само основни умения на работниците и без специализирана структура на лабораторията

Предимство на пълното геномно секвениране за установяване на резистентност при туберкулоза

- ❑ Изключително подходящ метод за надеждна идентификация на (под) видове / генотипове
- ❑ Разкрива висока точност при прогнозиране на чувствителност към лекарства от първа линия, което дава възможност да заместване по-голяма част от тестовете за фенотипна резистентност
- ❑ Има много по-висока специфичност като епидемиологичен маркер в сравнение с класическите методи DNA fingerprinting и VNTR typing
- ❑ Може да бъде финансово достъпен, защото замества много други лабораторни изследвания и подобрява диагностиката
- ❑ Все още не може да се прилага директно върху клиничен материал, а към ранни положителни MGIT култури



J Antimicrob Chemother 2019; **74**: 2605–2616
doi:10.1093/jac/dkz215 Advance Access publication 22 May 2019

Journal of Antimicrobial Chemotherapy

WGS more accurately predicts susceptibility of *Mycobacterium tuberculosis* to first-line drugs than phenotypic testing

Rana Jajou*, Tridia van der Laan, Rina de Zwaan, Miranda Kamst, Arnout Mulder, Albert de Neeling, Richard Anthony and Dick van Soolingen

National Tuberculosis Reference Laboratory, Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

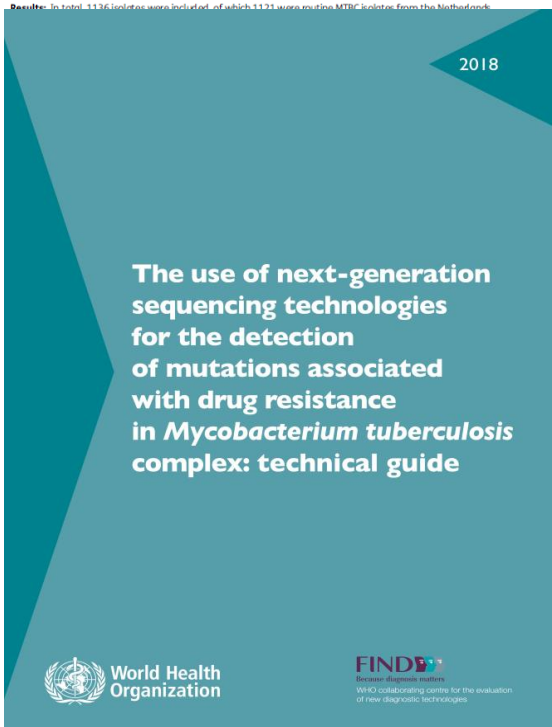
*Corresponding author. Tel: +31302742516; E-mail: rana.jajou@rivm.nl

Received 8 February 2019; returned 18 March 2019; revised 12 April 2019; accepted 22 April 2019

Background: Drug-susceptibility testing (DST) of *Mycobacterium tuberculosis* complex (MTBC) isolates by the Mycobacteria Growth Indicator Tube (MGIT) approach is the most widely applied reference standard. However, the use of WGS is increasing in many developed countries to detect resistance and predict susceptibility. We investigated the reliability of WGS in predicting drug susceptibility, and analysed the discrepancies between WGS and MGIT against the first-line drugs rifampicin, isoniazid, ethambutol and pyrazinamide.

Methods: DST by MGIT and WGS was performed on MTBC isolates received in 2016/2017. Nine genes and/or their promotor regions were investigated for resistance-associated mutations: *rpoB*, *katG*, *fabG1*, *ahpC*, *inhA*, *embA*, *embB*, *pncA* and *rpsA*. Isolates that were discrepant in their MGIT/WGS results and a control group with concordant results were retested in the MGIT, at the critical concentration and a lower concentration, and incubated for up to 45 days after the control tube became positive in the MGIT.

Results: In total, 1136 isolates were included, of which 1121 were routine MTBC isolates from the Netherlands



Пълното геномно секвениране гарантира ли перфектни резултати?

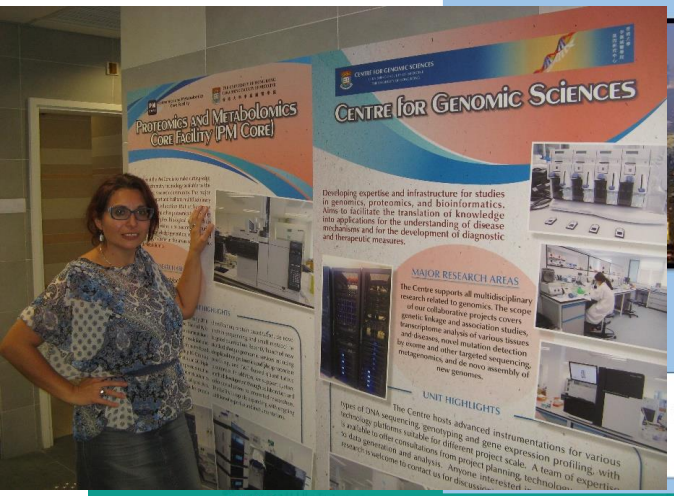
Данните от качествения контрол **на ECDC 2017/2018**, показват, че:

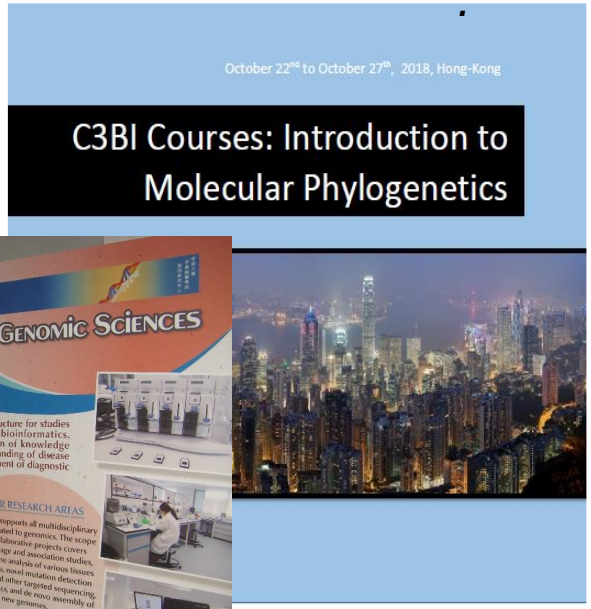
- Критериите за качество варират според геоположението на страните и биоинформатичния подход
- Няма стандартен подход за определяне чистотата на получените секвенции, поради възможност за замърсяване с бактериална флора или човешка ДНК
- Информация за набора от щамове и биоинформатични подходи са налични само в около половината от европейските лаборатории
- В много лаборатории не успяват да идентифицират филогенетичните връзки между щамовете, поради липса на достатъчно пълна информация за пациентите
- Установено е наличие на малки промени в генетичното разстояние на епидемиологично свързаните случаи

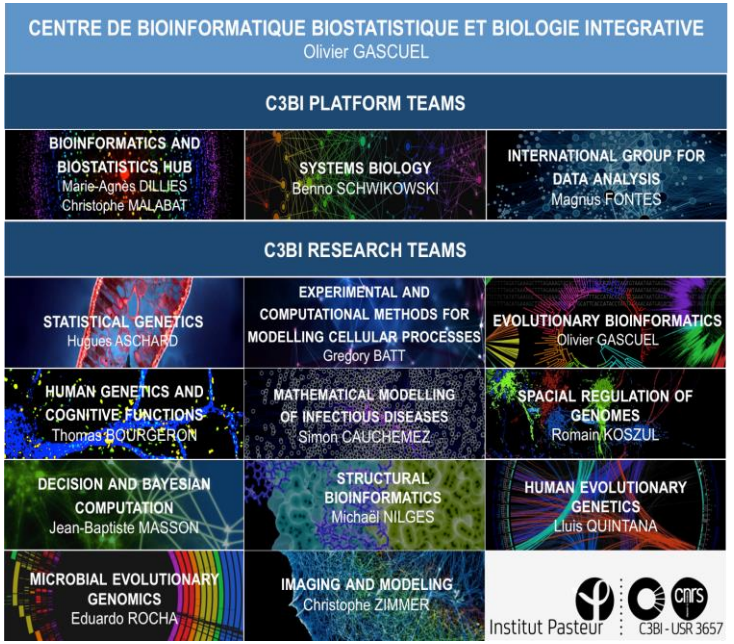
Пълното геномно секвениране гарантира ли перфектни резултати?

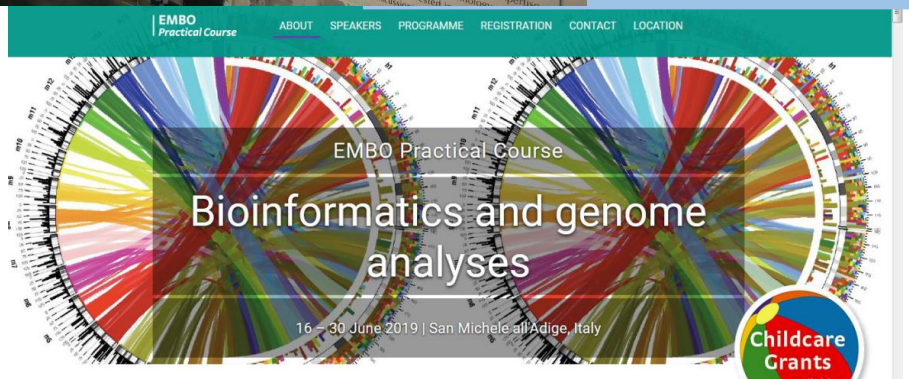
- Установено е наличие на основни разлики в средните генетични разстояния между неепидемиологично свързани случаи
- Много лаборатории пропускат “смислените” делеции в резистентните гени
- Интерпретацията и отчитането на резистентните мутации не е стандартизирана, особено за някои SNPs
- Понастоящем използваният референтен щам не е идеален за щамове от други линии/родове

Благодарности за Повишаване на квалификационните и ключови знания и умения по WGS/ NGS и Биоинформатика









ПРОГРАМА НА НЦЗПБ

За обучение на специализанти, докторанти
и продължителна квалификация на медицински и
немедицински кадри
(курсове и индивидуално обучение) през 2019 г.

Rijksinstituut voor Volksgezondheid
en Milieu
*Ministerie van Volksgezondheid,
Welzijn en Sport*

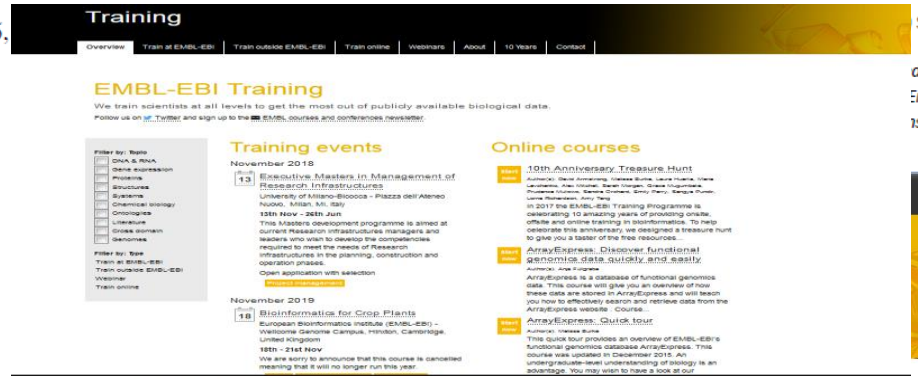
ТЕМА: ВИРУСНА ФИЛОГЕНИЯ И МОЛЕКУЛЯРНО-ЕПИДЕМИОЛОГИЧНИ МЕТОДИ ЗА НАДЗОР И КОНТРОЛ НА ВИРУСНИТЕ ЗАБОЛЯВАНИЯ
ДАТИ ЗА ПРОВЕЖДАНЕ: 14-17.05.2019 г.
ХОРАРИУМ: 24 часа (4 дни)
ЛЕКТОРИ: доц. д-р Любомира Николаева-Гломб, Ася Стоянова, Савина Стоицова, д.м.

Workshop

"Next generation sequencing and bioinformatics tools for *Mycobacterium tuberculosis* drug resistance detection and epidemiological analysis"

September 2019

d by
Environment, RIVM (Netherlands)
ıstitute (Russia)



Molecular Phylogenetics and Evolution 99 (2016) 133–143

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: Human traces across pathogen's phylogeography

Igor Mokrousov [a,*], Anna Vyazovaya [a], Tomotada Iwamoto [b], Yuriy Skiba [c], Ilva Pole [d,e], Svetlana Zhdanova [f], Kentaro Arikawa [b], Viacheslav Sinkov [f], Tatiana Umpeleva [g], Violeta Valcheva [h], Maria Alvarez Figueroa [i], Renate Ranka [d], Inta Jansone [d], Oleg Ogarkov [f], Viacheslav Zhuravlev [j], Olga Narvskaya [a,j]

[a] St. Petersburg Pasteur Institute, 14 Mira Street, St. Petersburg 197101, Russia
[b] Kobe Institute of Health, 4-6 Minatojima-nakamachi, Chuo-ku, Kobe 650-0046, Japan
[c] Aitkhozhin Institute of Molecular Biology and Biochemistry, 86, Dosmuhamedov str., Almaty 050012, Kazakhstan
[d] Latvian Biomedical Research and Study Centre, Ratsupites Street 1, Riga LV-1067, Latvia
[e] Center of Tuberculosis and Lung Diseases, Riga East University Hospital, Stopinu p., Riga LV-2118, Latvia
[f] Scientific Center of Family Health and Reproductive Problems, Irkutsk 664003, Russia
[g] Ural Research Institute of Phthisiopulmonology, 50 22go Partsiezda str., Ekaterinburg 620039, Russia
[h] The Stephan Angeloff Institute of Microbiology, Bulgarian Academy of Sciences, 26 Acad. G Bonchev str., Sofia 1113, Bulgaria
[i] Central Research Institute for Epidemiology, 3A Novogireevskaya str., Moscow 111123, Russia
[j] Research Institute of Phthisiopulmonology, 2-4 Ligovsky prospect, St. Petersburg 191036, Russia

ARTICLE INFO

Article history:
Received 23 November 2015
Revised 10 March 2016
Accepted 15 March 2016
Available online 19 March 2016

Keywords:
Latin-American-Mediterranean lineage
Mycobacterium tuberculosis
Phylogeny
Phylogeography

ABSTRACT

Currently, *Mycobacterium tuberculosis* isolates of Latin-American Mediterranean (LAM) family may be detected far beyond the geographic areas that coined its name 15 years ago. Here, we established the framework phylogeny of this geographically intriguing and pathobiologically important mycobacterial lineage and hypothesized how human demographics and migration influenced its phylogeography. Phylogenetic analysis of LAM isolates from all continents based on 24 variable number of tandem repeats (VNTR) loci and other markers identified three global sublineages with certain geographic affinities and defined by large deletions RD115, RD174, and by spoligotype SIT33. One minor sublineage (spoligotype SIT388) appears endemic in Japan. One-locus VNTR signatures were established for sublineages and served for their search in published literature and geographic mapping. We suggest that the LAM family originated in the Western Mediterranean region. The most widespread RD115 sublineage seems the most ancient and encompasses genetically and geographically distant branches, including extremely drug resistant KZN in South Africa and LAM-RUS recently widespread across Northern Eurasia. The RD174 sublineage likely started its active spread in Brazil; its earlier branch is relatively dominated by isolates from South America and the derived one is dominated by Portuguese and South/Southeastern African isolates. The relatively most recent SIT33-sublineage is marked with enigmatic gaps and peaks across the Americas and includes South African clade F11/RD761, which likely emerged within the SIT33 subpopulation after its arrival to Africa. In addition to SIT388-sublineage, other deeply rooted, endemic LAM sublineages may exist that remain to be discovered. As a general conclusion, human mass migration appears to be the major factor that shaped the *M. tuberculosis* phylogeography over large time-spans.

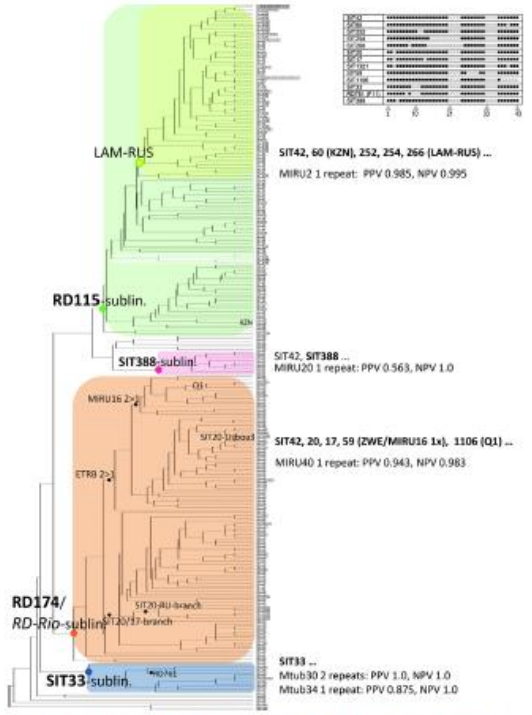


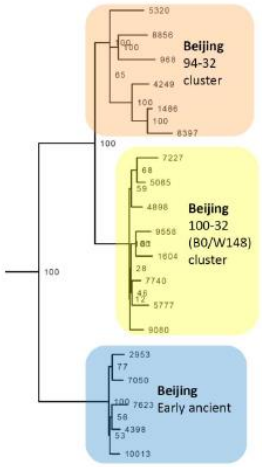
Fig. 1. UPGMA tree of 236 different 24-MIRU-VNTR profiles of the global dataset of 382 *M. tuberculosis* LAM isolates (see enlarged tree in Fig. S1). The same color code of the LAM sublineages is used through all figures. The VNTR signatures of the particular LAM sublineages concern only LAM family isolates. PPV and NPV: positive and negative predictive values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

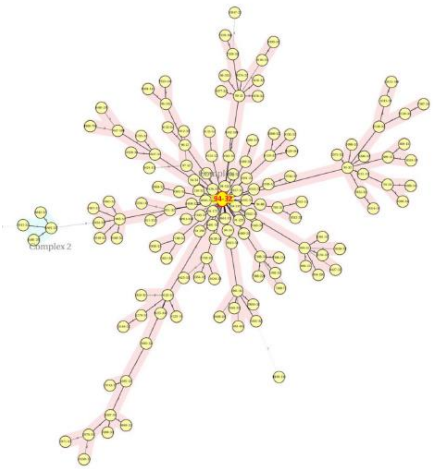
AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

Rapid Assay for Detection of the Epidemiologically Important Central Asian/Russian Strain of the *Mycobacterium tuberculosis* Beijing Genotype

Igor Mokrousov,[a] Ekaterina Chernyaeva,[b] Anna Vyazovaya,[a] Yuriy Skiba,[c] Natalia Solovieva,[d] Violeta Valcheva,[e,f] Klavdia Levina,[g] Natalya Malakhova,[c] Wei-Wei Jiao,[h,i,j,k] Lia Lima Gomes,[l] Philip N. Suffys,[l] Marge Kütt,[g] Nagima Aitkhozhina,[c] A-Dong Shan,[h,i,j,k] Olga Narvskaya,[a,d] Viacheslav Zhuravlev[d]

[a] Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg, Russia
[b] Oshuharsky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia
[c] Laboratory of the Genome, Aitkhozhin Institute of Molecular Biology and Biochemistry, Almaty, Kazakhstan
[d] St. Petersburg Research Institute of Phthisiopulmonology, St. Petersburg, Russia
[e] Department of Infectious Microbiology, The Stephan Angeloff Institute of Microbiology, Bulgarian Academy of Sciences, Sofia, Bulgaria
[f] "In vitro" Laboratory for Evaluation of Biological Activity and Toxicity, Biopharmaceutical Laboratory Complex, Sofia Tech Park, Sofia, Bulgaria
[g] North Estonian Medical Centre Foundation, Tallinn, Estonia
[h] Beijing Key Laboratory of Pediatric Respiratory Infection Diseases, Beijing Pediatric Research Institute, Beijing





f 19 *M. tuberculosis* Beijing genotype isolates, based on NGS/WGS data.

.. 24-MIRU-VNTR minimum spanning tree of the global dataset of the CC1 clonal complex *M. tuberculosis* taken from Merker et al. (2015), built with MIRU-VNTRplus.org.

Благодаря за вниманието!