

Поиск SNP штаммов *Mycobacterium tuberculosis*, распространенных в различных регионах мира

Анна Гончар

Итоговая презентация проекта 31.05.2014

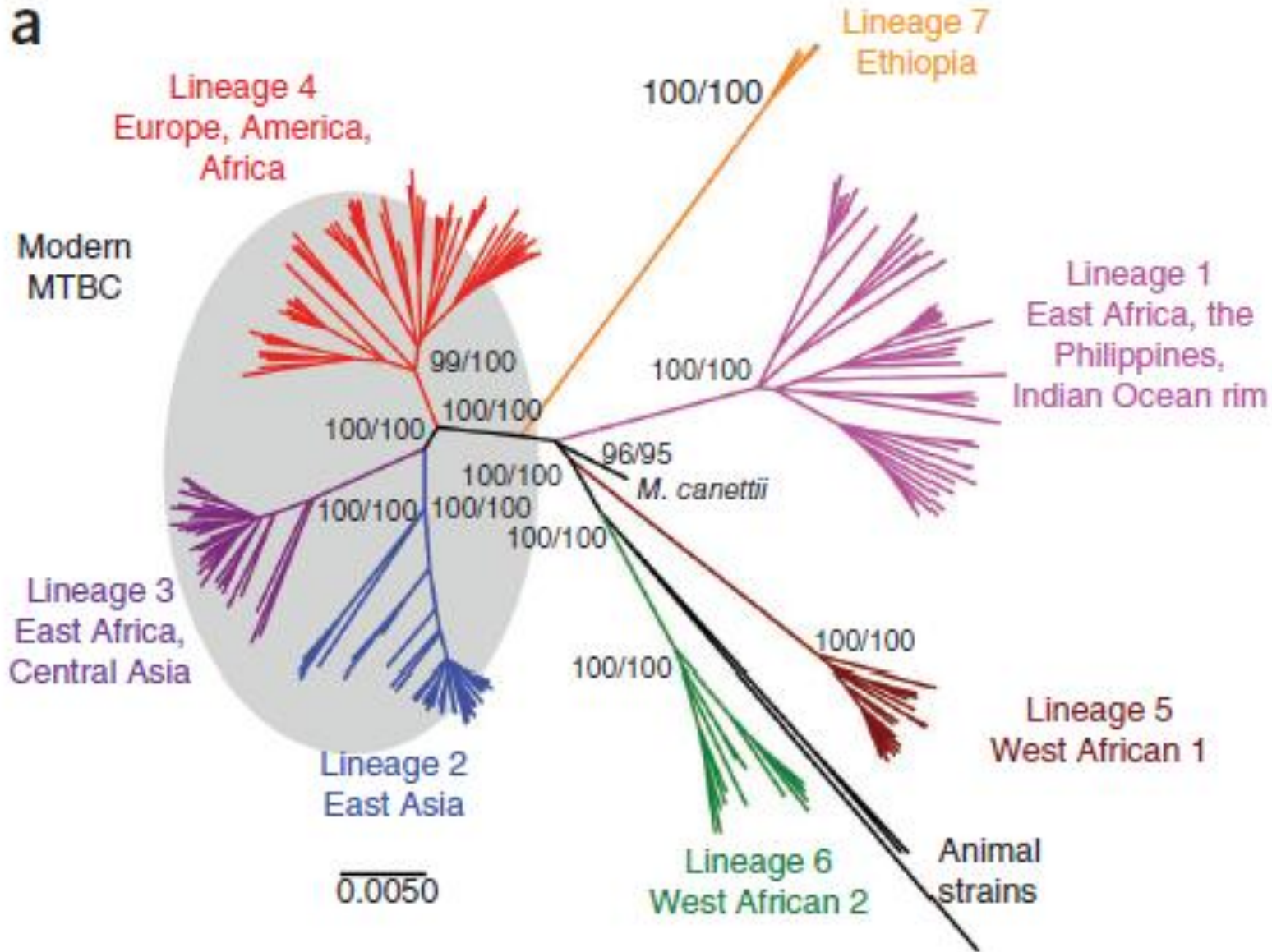
Руководитель Екатерина Черняева, Центр геномной биоинформатики им. Ф.Г. Добржанского

The project aims to discover SNPs in *M. tuberculosis* strains from all over the world using publicly available data, for further phylogenetic analysis.

Goals:

- search for WGS data in open access archives and databases;
- collect information about sequenced samples from publications;
- find nucleotide variations.

Seven major lineages of *M. tuberculosis*



Workflow: SNP calling with bowtie2 / samtools

Paper

- Min info – accession numbers, geographical locations; + info – spoligotype, DR.

.SRA

- Download data; sequencing info (single / paired, read length etc.)

→

.FASTQ

- fastq-dump (script 1)

QC

- Trimmomatic SLIDINGWINDOW:10:20 (script 2)

.SAM

- Align to reference H37Rv – bowtie2 (script 3)

.BAM

- Samtools (scripts 4 and 5); .bam → .sorted.bam

.VCF

- Mpileup (scripts 6 and 7); .bcf → .vcf

concat

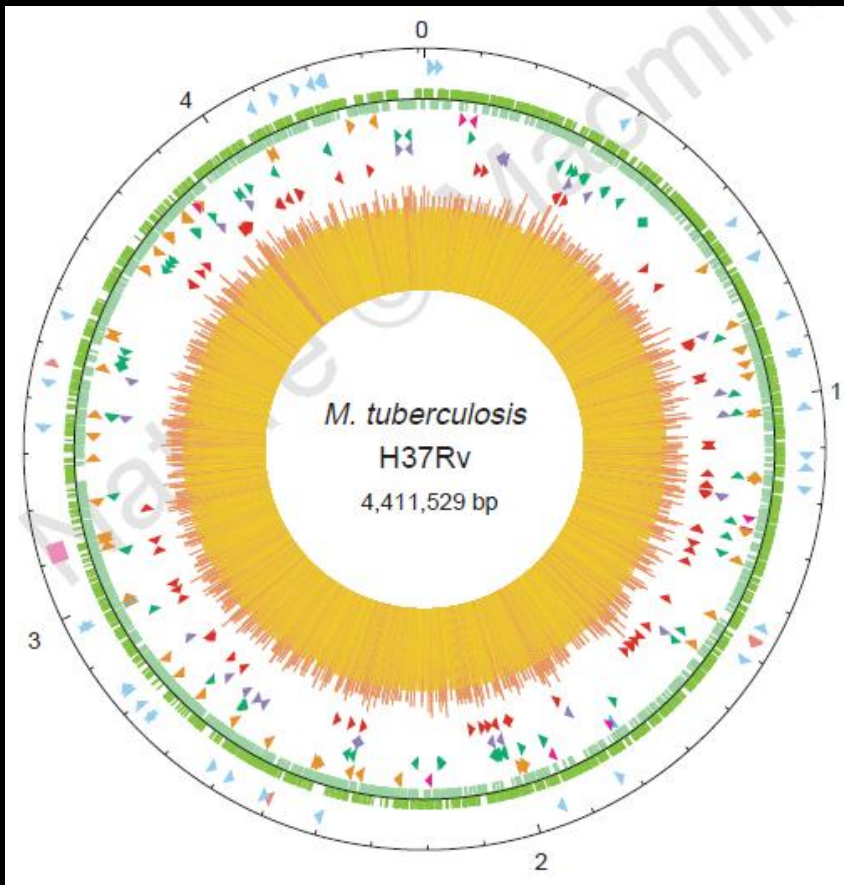
- SNP concatenation (script 8, by M. Rotkevich)

tree

- Neighbour-joining (simple clusterization)

First *M. tuberculosis* genome – 1998 год

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence



- H37Rv
- First isolated in 1905
- Most frequent reference
- 4'411'529 bp
- ~ 4047 genes

total of >1000 .sra

Paper references

1. Zhang, Hongtai, et al. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nature genetics* 45.10 (2013): 1255-1260.
2. Gardy, Jennifer L., et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine* 364.8 (2011): 730-739.
3. Merker, Matthias, et al. Whole Genome Sequencing Reveals Complex Evolution Patterns of Multidrug-Resistant *Mycobacterium tuberculosis* Beijing Strains in Patients. *PloS One* 8.12 (2013): e82551.
4. Walker, Timothy M., et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet infectious diseases* 13.2 (2013): 137-146.
5. Clark, Taane G., et al. Elucidating Emergence and Transmission of Multidrug-Resistant Tuberculosis in Treatment Experienced Patients by Whole Genome Sequencing. *PloS One* 8.12 (2013): e83012.
6. Bryant, Josephine M., et al. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC infectious diseases* 13.1 (2013): 1-12.
7. Coll, Francesc, et al. PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. *Tuberculosis* 94.3 (2014): 346-354.
8. Comas, Iñaki, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature genetics* 45.10 (2013): 1176-1182.

Very interesting *M. tuberculosis* WGS paper

Sample info
(lineage, spoligotype,
DR etc.)

Sample sequences

Sample names

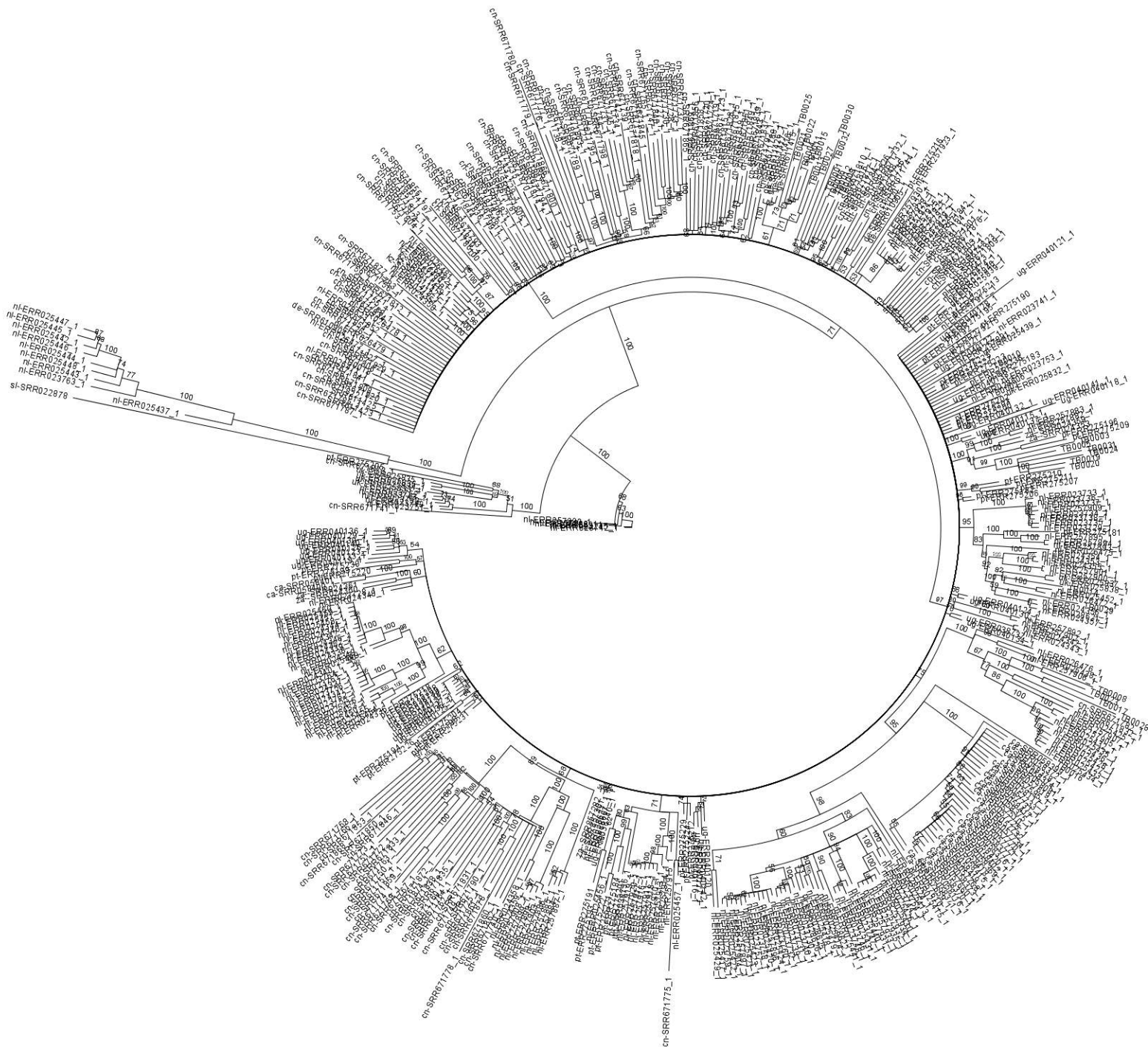
SRA accession
numbers

L7_BTBS746	GuangZ33
L5_1001003	SH400
L6_541504	HEN06035
L1_V232IO	XJ06017
L-xz09102	NM11008
FJ07036	...

ERR036185	ERR036191
ERR036186	ERR036192
ERR036187	ERR036193
ERR036188	ERR036194
ERR036189	ERR036195
ERR036190	...

Results

- 1101 .vcf
- **Countries** — Russia, UK, USA, China, Canada, Portugal, Germany, Georgia, Uzbekistan, Netherlands, Malawi, Uganda, South Africa, Global collection (Ethiopia, Vietnam, Mexico, South Korea, Pakistan, Senegal, Cambodia, Gambia, Malaysia, Sri Lanka, Nepal, India, Ghana, Sierra Leone, Tanzania, Iran, Afghanistan, Turkey, Singapore, Burkina Faso, Turkmenistan, Colombia, Puerto Rico, Nicaragua, Mongolia, Indonesia, Thailand, Burma, Laos, The Philippines, Guatemala, Salvador, Somalia)
- **DR, spoligotype, lineage info** ~1/3 samples
- <Overall tree>



Ongoing and future work

- Add more data
- Run with all available Russian sequences (GMTV)
- Annotate SNPs (SnpEff?)
- DR-related SNPs

Thank you