Skoltech

Skolkovo Institute of Science and Technology

MSc Program

Phylogeographic analysis of hepatitis A virus in Russia

Student: *Alevtina Koreshova*

Research Advisor: *Georgii Bazykin*

Co-Advisor: *Alexey Neverov*

January, 2019

Introduction

Skoltech
Skolkovo Institute of Science and Technology

Hepatitis A virus (HAV)

- is a positive-stranded RNA virus, 7.5 kb genome
- causes acute hepatitis in adults
- spreads via fecal-oral route
- vaccination is very effective, but nearly nobody does it
- comprises 6 genotypes, from which I and III are most frequent in humans and are both divided into A and B subgenotypes
- genotyping is carried out on short variable fragments
- in Russia, IA and IIIA subgenotypes predominate
- has low mutation rate $\approx 1.99 \times 10^{-4}$ (Kulkarni et al., 2009)

Aim and objectives

Skoltech
Skolkovo Institute of Science and Technology

The aim of the work is to reconstruct the phylogenetics, geographic structure and migration events of HAV in Russia.

The hypothesis is that IIIA subgenotype originates from India while IA is expanding from Europe.

Objectives:

- Phylogeographic analysis using Bayesian approach (BEAST)
- Phylogeographic analysis with ML approach (augur) and GenGis
- Obtaining full genomes of HAV
- Have fun with full genomes

The dataset

Skoltech
Skolkovo Institute of Science and Technology

- Central Research Institute of Epidemiology kindly provided us with the unique collection of HAV isolates obtained all over Russian Federation and CIS countries.
- Samples (more than 500 isolates) were collected from 1999 to 2015 and characterized by one of the two most variable fragments of HAV genome (VP1 and 2C regions) or by both of them.
- The dataset includes 124 unique sequences of 2C/3A region, length ≈ 650 bp, and 217 sequences of VP1/2A region, length ≈ 400 bp. For each sample, date and location of collection is indicated.
- Whole-genome sequences of HAV from GenBank were used (accession numbers KC182588, KC182589, HM769724, LC049342, LC049339, LC049338, AB909123, AB623053, JQ655151, AB973400, FJ360735, FJ360734, FJ360733, FJ360732, FJ360731, FJ360730, AB279734, AB279733, AB279732).

Methods

Skoltech
Skolkovo Institute of Science and Technology

Software:

MEGA – alignment, tree reconstruction

IQTree – model selection

BEAST – phylogeography, model selection

Spread3 – visualization of phylogeography

Chelper – primer design for sequencing of whole genomes

Planned:

GenGis – estimating directions of migration

Augur – phylogeography based on ML, not Bayes

This python package which name I don't know – geographic clusterization

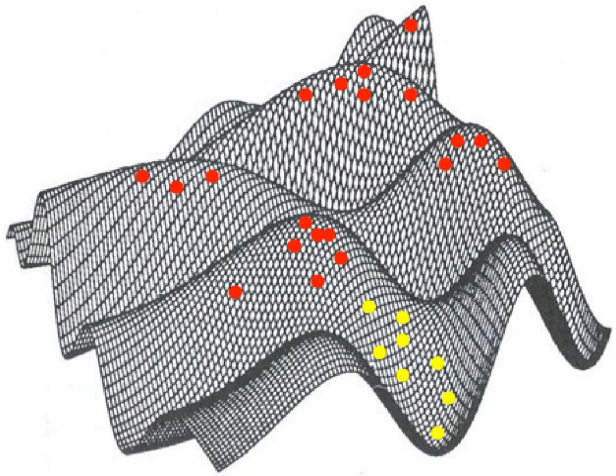
Whole genome sequencing will be performed with Mi-seq

BEAST and VEME2018

BEAST
Bayesian Evolutionary Analysis Sampling Trees

The 23rd International Workshop on Virus Evolution and Molecular Epidemiology (**VEME2018**) was hosted in Berlin, Germany. It was devoted to phylogenetic inference and BEAST.

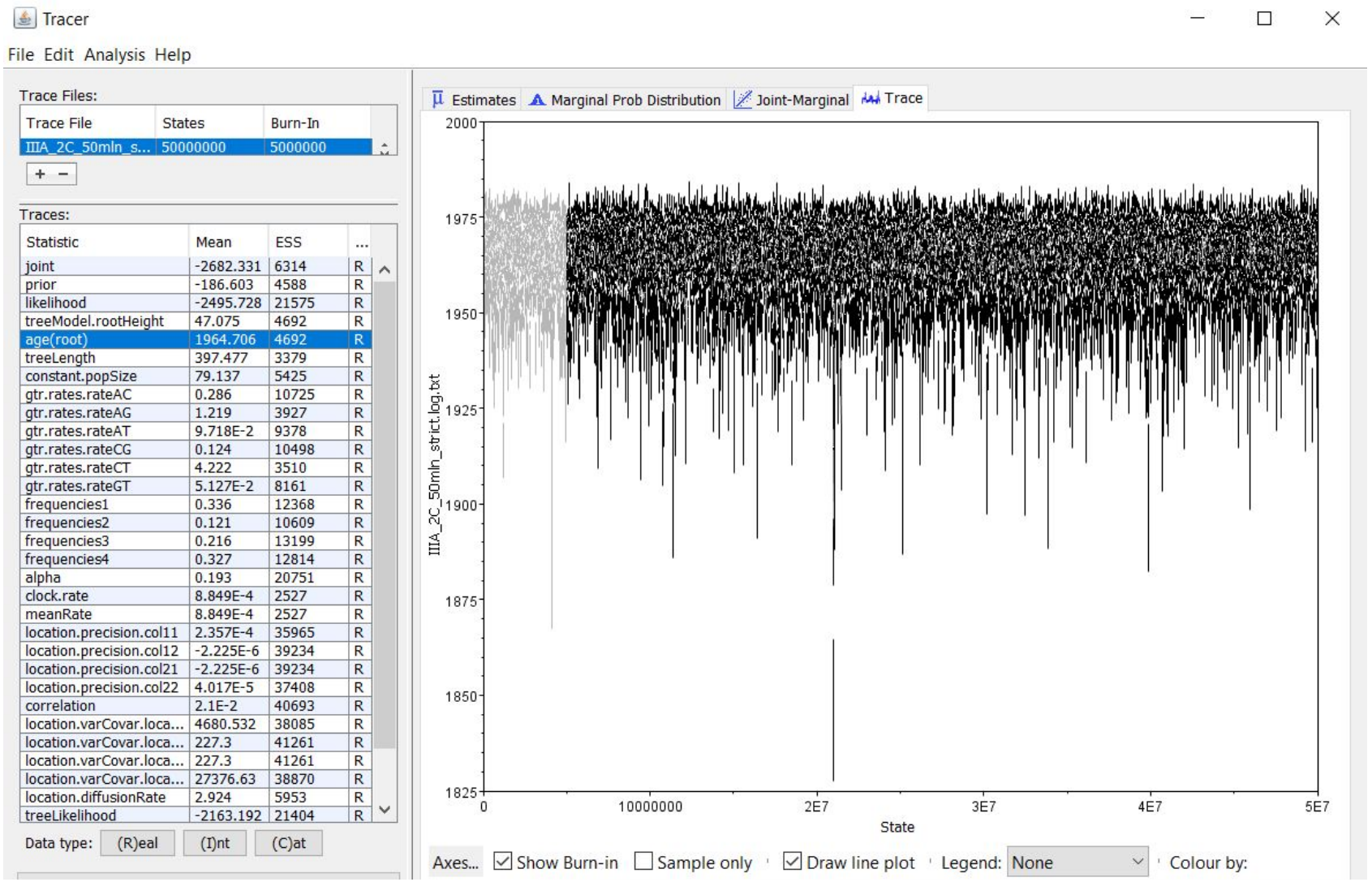
$$\Pr(M|D) = \frac{\Pr(D|M)\Pr(M)}{\Pr(D)}$$



Markov chain Monte Carlo

Results

Skoltech
Skolkovo Institute of Science and Technology



Skoltech
Skolkovo Institute of Science and Technology

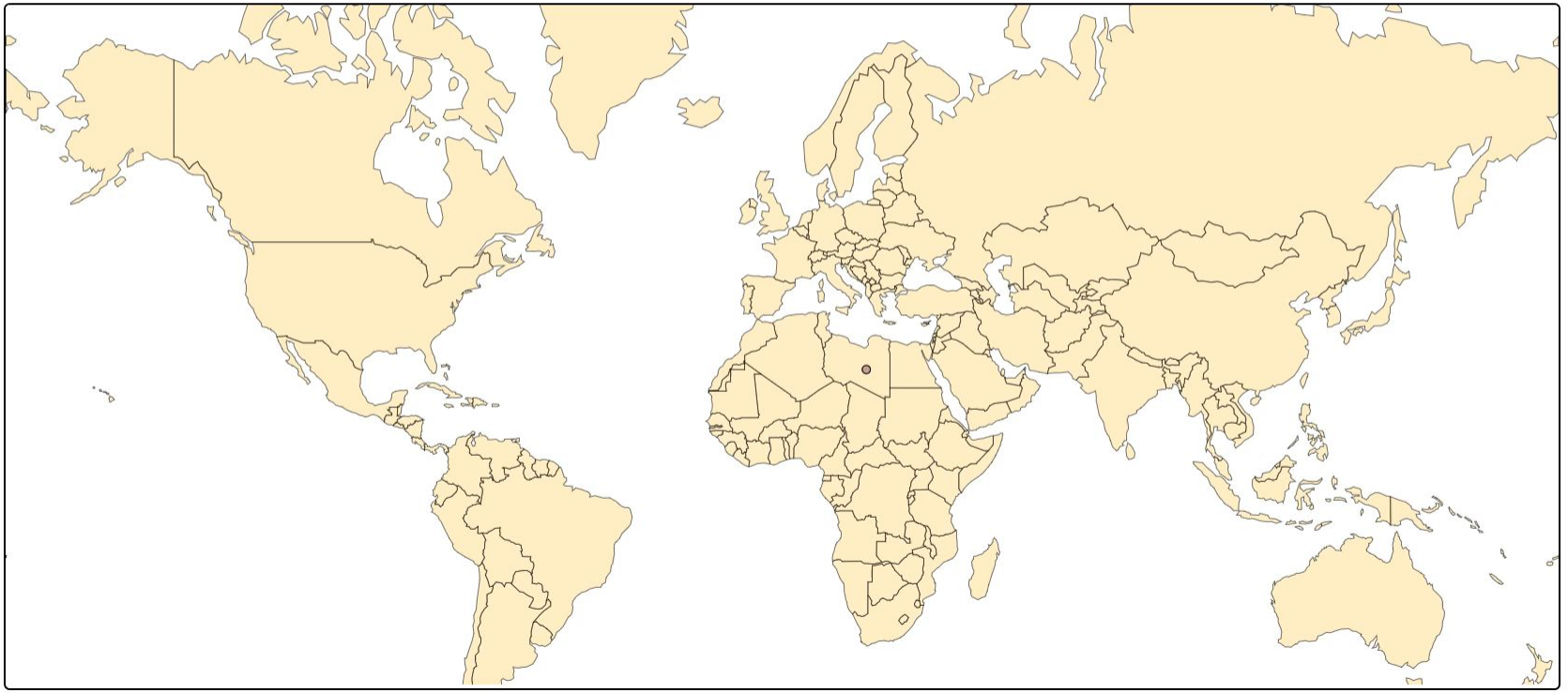
Results: IA phylogeography

Current date: 1798/07/19

1800 1820 1840 1860 1880 1900 1920 1940 1960 1980 2000

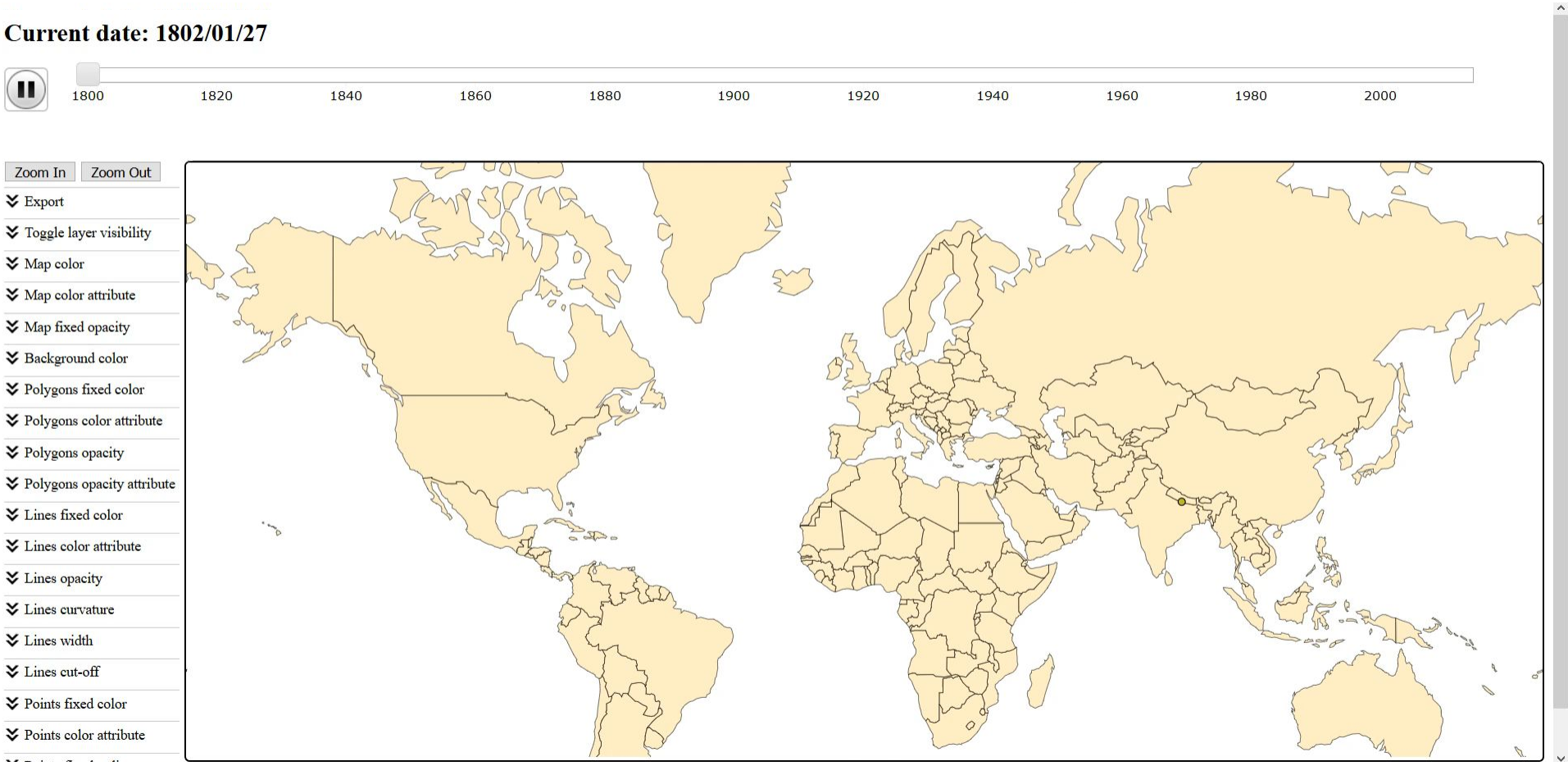
Zoom In Zoom Out

- Export
- Toggle layer visibility
- Map color
- Map color attribute
- Map fixed opacity
- Background color
- Polygons fixed color
- Polygons color attribute
- Polygons opacity
- Polygons opacity attribute
- Lines fixed color
- Lines color attribute
- Lines opacity
- Lines curvature
- Lines width
- Lines cut-off
- Points fixed color
- Points color attribute
- Points fixed radius



Skoltech
Skolkovo Institute of Science and Technology

Results: IIIA phylogeography



Discussion and problems

- Sampling issues
- Uncertainty in time and place of sample collection (current solution – setting random jitter in BEAST)
- Low mutation rate => low resolution

Plans

Skoltech
Skolkovo Institute of Science and Technology

- Verify BEAST results with other approaches (augur, GenGis)
- Obtain whole genomes of HAV
- Analyze the whole genomes and add partial sequences from Genbank if possible

Conclusions (obtained, desired)

Skoltech
Skolkovo Institute of Science and Technology

So far, the results of BEAST estimation of phylogeography agree with literature data

Desired :

Acknowledgements

List of colleagues who considerably contributed in the research.