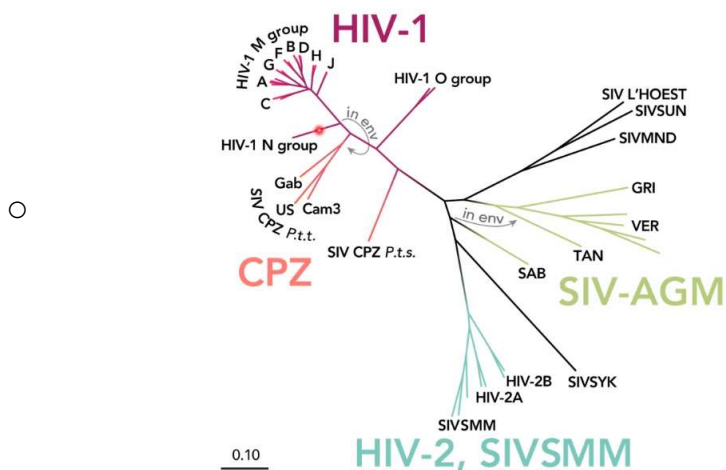


- Prabhav Kalaghatgi
 - Schwerpunkte seiner Arbeit: Prob theory, algorithm design, theoretical software development, recently hardware side of software development
- Orga

Kein Whiteboard, sondern Google-Classroom	https://classroom.google.com/c/NTk3MzI4MDA3MjYx?cjc=74ey5xb
Same link (zoom) for all meetings	https://us02web.zoom.us/j/2107279597

 - Meetings
 - Dienstags und Freitags um 14:00
 - Nächstes Meeting 10.03.
 - Danach erst 21ster wegen Softwarepraktikum
- Expections for course:
 - Appreciation of viruses evolution, to some extent (it can get messy)
 - What can we learn from virus sequences
 - Evolutionary analysis
 - DNA samples from infected persons
 - RNA sample from virus
 - Which virus variant is growing, declining, taking over, ...?
- Pandemic is associated with species transgressions of viruses
- There is some work on "evolutionary proof" vaccines
- Lectures: Phylodynamic analysis of SARS-CoV-2
 - First trees based on protein-sequences
 - Longer branches => Higher amount of change

Evolutionary origin of viruses



- Two types of HIV that have successfully transferred from chimpanzees to humans
- HIV 1 is closely related to SIV in CPZ (and more severe to humans)
- HIV 2 is closely related to SIV in Humans (and less severe to humans)
- CPZ are also infected with HIV-like viruses
- CPZ evolved to survive the most HIV-infections
- Humans will usually die of HIV-infections, if not medicated

Topology with leaf names & branch lengths:

- **(use this)**

$$(((l_1: t_1, l_2: t_2): t_6, l_3: t_3): t_7, (l_4: t_4, l_5: t_5): t_8);$$
- Newick format in that specific style, can be computed by performing a post-order traversal (visit children before you visit the parents, from the leafs to the root)

Project overview

- Part I
 - Phylogeny construction
 - 3 weeks (March 7 to March 24)
 - Part II
 - Phylodynamic inference
 - 3 weeks (March 27 to April 14)
 - Part III
 - Trouble-shooting
 - Preparing report and presentation
 - 3 weeks (April 18 to May 5)
 - Report and presentation
 - Week of May 16 to 19
- Shannon Entropy: Shows a lot of evolution in the spike protein
- Spike system is evolving to evade immune system
- Start Analysis with subsampled sequences and intermediate files
Nextstrain.org/ncov/open/global/6m

Subsampled sequences and intermediate files

The files below exist for every region (`global`, `africa`, `asia`, `europa`, `north-america`, `oceania` and `south-america`) and correspond to each region's 6 month timespan build (e.g. `global/6m`, `africa/6m`, `asia/6m`, etc). Files for the `2m` and `all-time` builds (e.g. `global/2m`, `global/all-time`, etc) are not yet available. The links below refer to the `global` region; substitute `global` with another region name in the links if desired.
 - - * `global/6m metadata.tsv.gz`
 - * `global/6m sequences.fasta.gz`
 - * `global/6m aligned.fasta.gz`
 - * `global/6m Auspice tree`
 - * `global/6m Auspice root sequence`
 - * `global/6m Auspice tip frequencies`
 - Metadata.tsv.zst => für phylogenetic interference
 - Aligned.fasta.zst => für den Start
 - No alignment necessary
- **ToDo:**
 - Familiarize with website: Nextstrain.org/ncov/open/global/6m
 - Download data, read fast sequence, built initial trees with Neighbor Joining or similar tool. Visualize tree. Start with existing libraries. This will be a benchmark to later self-made software for building trees and co.