Unlocking the Secrets of DNA: From Sequences to Disease Dynamics

DNA sequencing has become remarkably affordable, and DNA sequence data are now available on an unprecedented scale. The pressing question is: what can we actually learn from these data?

One fascinating application of pathogen DNA sequences is the estimation of key parameters in dynamic models, such as the well-known reproduction number. But we can go even further: it is possible to detect mechanisms and patterns of spread directly from DNA sequences.

The method relies on the coalescent theory: by examining pathogen samples from $n \in \mathbb{N}$ individuals, we obtain n distinct gene sequences. When comparing these sequences, mutations are almost inevitable. The greater the differences between two sequences, the less closely related they are. This information allows us to reconstruct possible ancestral trees—much like how phylogenetics reveals the evolutionary relationships among species (e.g., humans are more closely related to gorillas than to horses).

In phylodynamics, this same approach is used to build a "family tree" of pathogens. What makes this particularly exciting is that such trees also encode information about the dynamics of disease spread—and we can harness this for deeper insights.

In this project, we will explore this method hands-on, applying it to mPox (monkeypox) data. mPox is a compelling case study: historically confined mainly to Africa, infections surged globally around 2022, appearing suddenly in Europe, North America, and beyond. The key question is: Can we trace this dramatic shift in the DNA data itself?

Project plan:
(a) Understand the basic theory of phylodynamics
(b) understand the statistical tools which are available
(c) Set up a dynamic ODE model for the mPox
(e) use the phylodynamics-tools to investigate Pox DNA samples with the aim to detect traces of the recent dramatic spread of mPox

Supervisor: Johannes Müller.