NITheCS Colloquium
Monday, 23 August 2021, 16h00
Prof Erik Aurell | KTH, Stockholm, Sweden

“Lessons learned on SARS-CoV-2 genomics from GISAID”

ABSTRACT

The organization known as GISAID started as an initiative to store and share all influenza data. Since the start of the COVID19 pandemic it has become the main repository of SARS-CoV-2 genome sequences. As of writing this abstract (June 16, 2021) there are close to two million SARS-CoV-2 submissions on GISAID, many of them whole-genome sequences labelled as high quality. This makes GISAID one of the largest, richest and most easily accessible data sources in all genomics.

I will discuss some simple analysis that can be done with such data, likely a harbinger of many similar data resources to appear in the future.

First, one can compute time series of frequencies of mutations. Such time series graphically demonstrate that the variants up to recently known as 'UK" and "South African" (B.1.1.7 and B.1.351) are in fact the agglomeration of different variants, and/or that the initial formal definitions from Public Health England were faulty.

Second, the distribution of coronaviruses over genotypes is like everything else in biology shaped by evolution, a process driven by natural selection, mutations, genetic drift (finite-N noise) and recombination. Coronoviruses are in general known to have significant recombination, and there are also reports of ongoing recombination in SARS-CoV-2 in human hosts. It was discovered by Motoo Kimura in the mid-60ies that if recombination is the fastest mechanism, the stationary distribution over genotypes is of the Ising/Potts type from equilibrium statistical mechanics. One can hence apply simple machine learning techniques to infer parameters of such distributions from sequence data, and check how stable are those parameters, and what additional information they may convey.

BIOGRAPHY

Erik Aurell has a BSc (1983) and PhD (1989) from Göteborg University, Sweden. He holds since 2003 the position as professor of Biological Physics at KTH Royal Institute of Technology in Stockholm, Sweden. He has held guest professorships in Finland and China. He is still active in what was his main research interest of biological physics and the interface to computational biology and genomics; since about 2015 he has however focused on open quantum systems, and other problems of quantum science. He is Board member of European Physical Society Division for Nonlinear and Statistical Physics.

CLICK TO REGISTER

https://bit.ly/3xN4Og8

WANT TO FIND OUT MORE?

Contact our Communications Officer
T: +27 (0)87 702 9364 | E: info@nithecs.ac.za