

Genomic signatures in viruses

Akademisk avhandling

Som för avläggande av medicine doktorsexamen vid Sahlgrenska akademien, Göteborgs universitet kommer att offentligen försvaras i Mikrobiologen Föreläsningssalen våning 3, Guldhedsgatan 10, den 16 juni, klockan 13:00

av Joel Gustafsson

Fakultetsopponent:

Prof. Jan Albert

Institutionen för mikrobiologi, tumör- och cellbiologi, Karolinska Institutet, Stockholm

Avhandlingen baseras på följande delarbeten

- I. Gustafsson, J., Norberg, P., Qvick-Wester, J.R., Schliep, A.
Fast parallel construction of variable-length Markov chains
BMC Bioinformatics 22, 487 (2021)
- II. Gustafsson, J., Edwards, S., Schliep, A., Norberg, P.
Estimating phylogenies from raw sequencing reads using variable-length Markov chains
Manuscript
- III. Holmudden, M.*, Gustafsson, J.*, Schliep, A., Norberg, P.
Species-specific genomic signatures in viruses
Manuscript * Denotes shared first-authorship
- IV. Gustafsson, J., Schliep, A., Norberg, P.
Virus-host similarities in genomic signatures
Manuscript
- V. Gustafsson, J., Schliep, A., Norberg, P.
Detection of Herpes simplex type 1 and 2 recombination in clinical samples
Manuscript

SAHLGRENSKA AKADEMIN
INSTITUTIONEN FÖR BIOMEDICIN



Genomic signatures in viruses

Joel Gustafsson

Avdelningen för infektionssjukdomar, Institutionen för biomedicin, Sahlgrenska akademien, Göteborgs universitet, Sverige, 23.

Abstract

In an age of global pandemics, studying how viruses and their genomes evolve is of great importance. It has previously been found that genomes of many eukaryotes and prokaryotes have specific preferences for nucleotides, dinucleotides and codons. Such preferences are characterized by the selective pressure acting on the genomes, and are referred to as specific genomic signatures. The presence of such signatures has, to our knowledge, not been studied in viruses, and it is therefore the aim of this thesis to thoroughly investigate genomic signatures in viruses.

In the first two papers of this thesis, new algorithms for the study of genomic signatures were developed. Here, such genomic signatures were based on variable-length Markov chains of a genome. Compared to pre-existing methods, our new algorithms are a thousand times faster, and compared to the state-of-the-art, the algorithms are up to 600 times faster while also requiring less memory. These methods enable computationally efficient analysis of genomic signatures, even on laptops.

In the subsequent two papers, we thoroughly analyzed the genomic signatures of viruses and compared such signatures to those of the viruses' hosts. The results illustrate that a majority of viruses have specific genomic signatures. In addition, in most cases, the signatures of viruses are not similar to the signatures of their hosts other than in GC content. This dissimilarity indicates that viruses' signatures are independent of their host's signature, despite viruses' dependence on their host's genetic and protein-expression machinery.

In the final paper, we illustrated an application of the genomic signatures by applying them to identify recombination events between *Human alphaherpesvirus 1* and *Human alphaherpesvirus 2*.

We thus demonstrate that genomic signatures of variable length are an important property of virus genomes. They hint at the importance of the evolution of specific patterns of the nucleotide sequence of viruses. These patterns can likely identify even remotely related viruses in collections of unknown sequences, thus helping detect and classify novel viruses. In addition, it might be possible to use and modify the genomic signatures to, e.g., attenuate viruses to create vaccine candidates.

Keywords: Virus evolution, Virology, Bioinformatics, Alignment-free

ISBN: 978-91-8069-269-4 (TRYCK)

<http://hdl.handle.net/2077/75906>

ISBN: 978-91-8069-270-0 (PDF)