



Aalto-yliopisto
Perustieteiden
korkeakoulu

Analyzing Viruses via Hidden Markov Models

Tomas Laamanen

22.10.2021

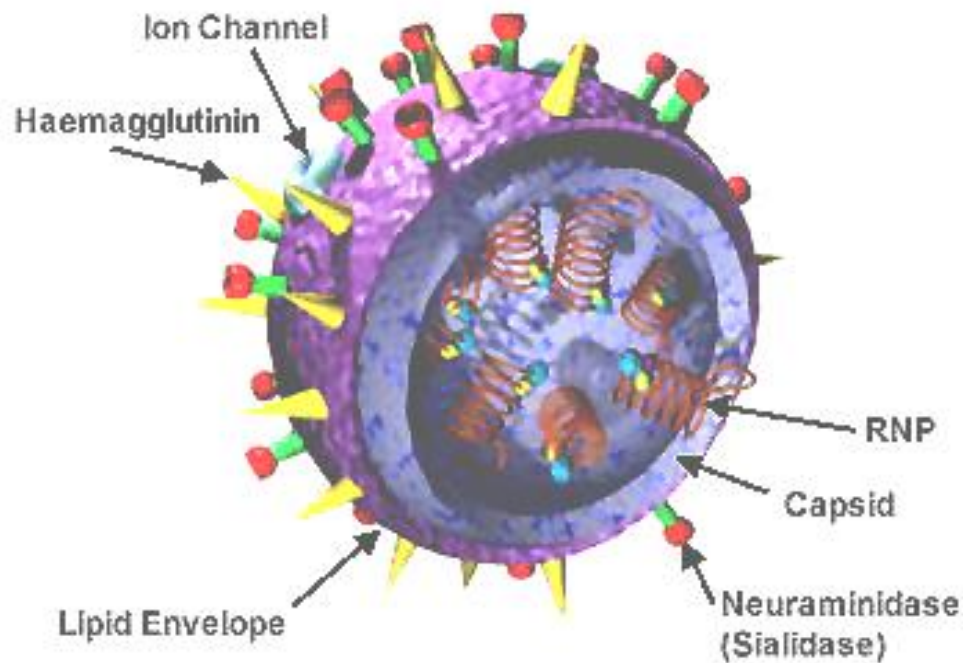
Supervisor: *Kai Virtanen*

Director: *Kai Virtanen*

This work may be saved and published on Aalto University's open websites. All other rights reserved.

Background 1/2

Viruses are made from genetic material and a shell

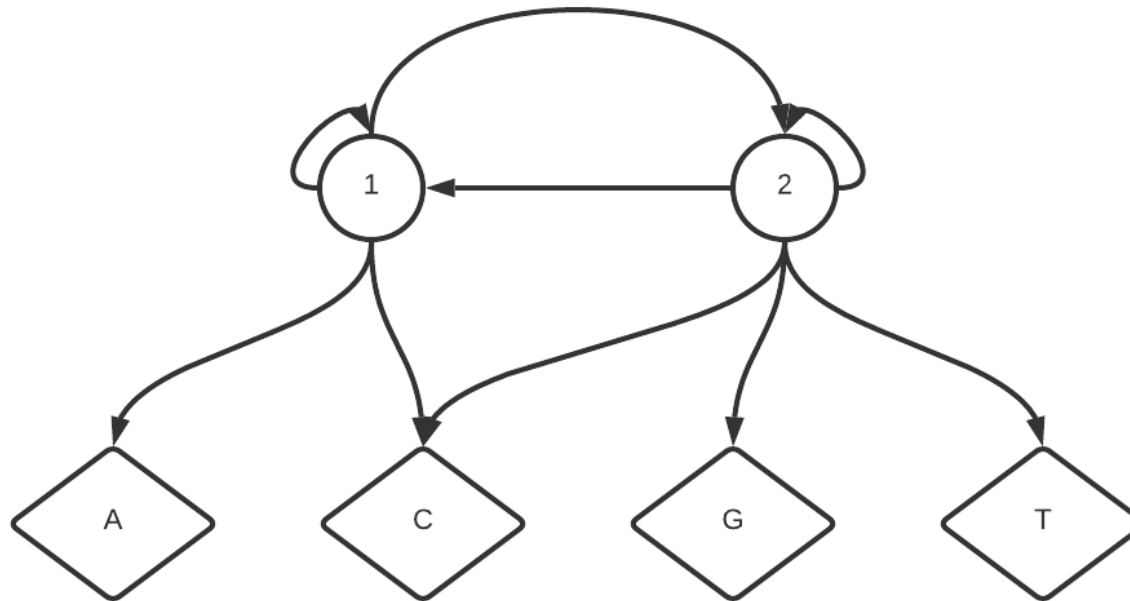


Background 2/2

- Genetic material determines properties of viruses
- Understanding mutations means understanding viruses
- Stochastic models have been used for
 - analyzing structures of viruses
 - predicting mutations of viruses

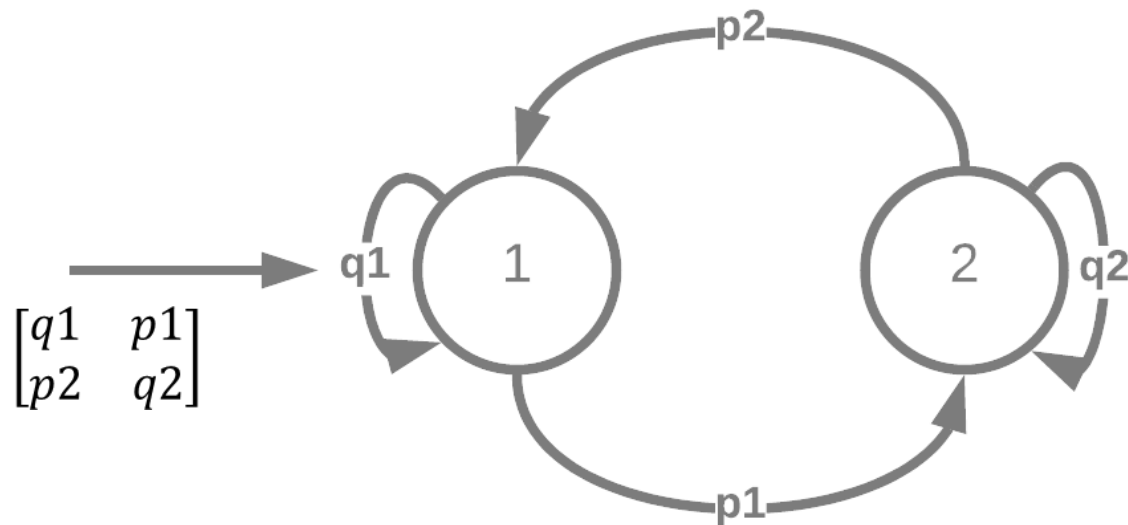
Aims 1/2

- Create a Hidden Markov Models for predicting mutations
- These models are trained with the genomes of a virus family



Aims 2/2

- Use the Hidden Markov Models to analyze how common different mutations are
- Study transmission and emission matrices in order to understand patterns in viruses



Methods and tools

- Matlab
- The training data is obtained from National Center for Biotechnology Information Nucleotide
- Information criteria and Cross-validation
 - Used for evaluating how useful the trained models are

Limitations

- Convergence of the Baum-Welch algorithm
- The training of the models can be computationally intensive
 - Limits how many states the models can have

References

- Byung-Jun, Y. "Hidden Markov Models and their Applications in Biological Sequence Analysis." *Current Genomics*, vol. 10, no. 6, 2009, pp. 402-415
- Pierce, Benjamin A. *Genetics: A Conceptual Approach*. 5th ed., Freeman, 2013.
- Weinan, E, Tiejun, Li, and Vanden-Eijnden, Eric. *Applied Stochastic Analysis*. American Mathematical Society. 2019.

Timetable

- Research 1.7.2021 – 5.8.2021
- Writing 6.8.2021 – 10.9.2021
- Proofreading 11.9.2021 – 1.11.2021
- Introductory presentation 22.10.2021
- Final presentation of results 2.12.2021