Designing an algorithm to study flu viruses

Arti Khana
Department of Electrical and Computer Engineering, The George Washington University

Dr. Takis Benos
Department of Computational Biology, University of Pittsburgh

Introduction

Since 1997, the avian influenza H5N1 virus has infected over 250 people worldwide and has a fatality rate of approximately 60% [1]. Currently, the number of reported transmission of avian flu to humans is small and subsequent human-to-human transitions are of limited virulence. We are interested in studying the evolutionary aspects of flu viruses. To that extent we will develop a genetic algorithm that given: **(1)** a starting flu genomic sequence, **(2)** a set of accepted amino acid substitutions, **(3)** a mutation rate and **(4)** a fitness function, it will be able to generate new sequences over time. One of the biggest concerns regarding avian-to-human transmission is whether the avian strain can recombine with human strains in a way that will help it adapt better to humans. So our algorithm will also model this process, known as *reassortment* [2].

Method

We aim to develop an algorithm that given a fitness function and a mutation, it can emulate evolution and answer to the question: how long will it take for the virus to evolve below a given threshold in the fitness function. We will work on the genetic algorithms (GA) framework, since this appears to be the most relevant to the biology of the problem. The algorithm will start with a flu genome (all 8 segments) and will generate random mutations according to a mutation rate. Sequences will be selected

based on some fitness function and the process will be repeated. This basic algorithm will be supplemented by an option in which two different viral strains can infect the same cell and recombine randomly and the overall fitness of the recombined viruses will be taken into consideration. At first, the fitness function will be the amino acid adaptability of the virus as it is calculated from a set of aligned amino acid sequences.

Goals

1. Develop an algorithm that given an initial RNA sequence and some parameters it will *generate N* new sequences.
2. Assign probabilities for each sequence.
3. Selecting top *X* sequences (*e.g.*, *X*=100,) and repeat *K* number of times.
4. If two sequences are given in the input, then evolve both and every *T* time recombine their segments randomly.

References:

[1] Zamboon, Maria. Lessons from the 1918 influenza. *Nature Biotechnology* **25**, 433-434 (2007)

[2] Nicholls, Henry. Pandemic Influenza: The Inside Story. *PLoS Biology* **4**, 156-160 (2006)