

# TEORIA DA INFORMAÇÃO APLICADA À ANÁLISE DE SEQUENCIAS GENÉTICAS

**Bolsista:** Igor Ruan Kosarczuk da Silva (Estudante de Bacharelado em Física)

**Orientador:** Prof. David Dominguez

**Área de Estudo:** Física Interdisciplinária

## TEORIA DA INFORMAÇÃO

A Teoria da Informação é um ramo interdisciplinar envolvendo probabilidade, estatística e física com a finalidade de lidar com sistemas de comunicação, transmissão de dados, criptografia, compressão de dados e etc. Foi desenvolvida por Claude E. Shannon da *Bell Telephone Laboratories* cuja ideia central é de que a comunicação pode ser considerada como um problema matemático, ocorrendo através de um meio (o canal) entre um emissor e um receptor. Desenvolve também o conceito de capacidade do canal, e com paralelo à descrição de entropia de Boltzmann, onde o termo *entropia de informação*, *incerteza*, ou simplesmente *informação* é definido como uma grandeza matemática em bits que descreve principalmente a capacidade que um canal de informação necessita na transmissão de dados.

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

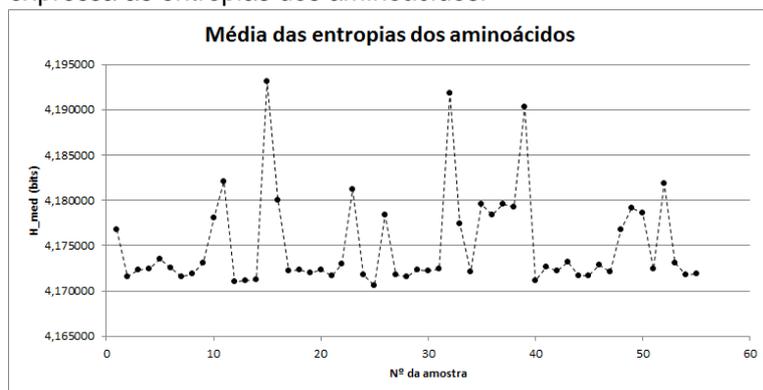
(A equação mais famosa da teoria: entropia de informação)

## APLICAÇÕES À GENÉTICA

A teoria de Shannon é comumente usada em criptografia e compressão de dados. Dentro da genética, o DNA pode ser considerado como uma fonte de informação criptografada onde cada base nitrogenada correspondente a uma dupla de bits (00, 01, 10, 11) está arranjada de tal forma (ordenada propositalmente pela evolução biológica) que podemos determinar o quão comprimido a informação está contida no segmento de DNA. Expandindo a aplicação para códon (trincas de bases nitrogenadas que contém a *instrução* para a polimerização proteica) somos capazes não só de determinar uma compressão, mas sim um ordenamento (ou desordem) correspondente de suas quantidades e estabelecer que existe uma instrução codificada (não aleatória; selecionada) neste segmento de DNA. O terceiro passo é incluir os *conjuntos equivalentes de instrução* que são essencialmente códon diferentes que possuem a mesma instrução, assim reduz-se ainda mais a informação compactada no DNA, mas também gera uma espécie de estabilidade local frente a mutações das bases nitrogenadas com o uso do segundo conceito da entropia, a *incerteza*.

## DETERMINAÇÃO DE VARIAÇÃO VIRAL

O trabalho apresentado consiste em determinar variações virais (estirpes) com o uso da teoria da informação. Dentro da análise realizada fomos capazes de verificar variações do vírus da gripe, a Influenza A (H1N1) (o vírus da pandemia de 2009, com variação H275Y no sexto segmento) durante os anos de 2010 à 2014 para 55 amostras em cada um dos seus 8 segmentos de RNA virais. Fizemos a análise para bases nitrogenadas, códon, e para os aminoácidos (conjuntos equivalentes de instrução) e verificamos através deste último, padrões mais expressivos que nos mostram as diferenças não triviais entre as amostras e seus segmentos. O gráfico abaixo expressa as entropias dos aminoácidos:



## PERSPECTIVA

As aplicações aqui mencionadas são apenas um primeiro contato da teoria com a genética e a totalidade de sua aplicação certamente está fora de nossa imaginação. Porém em uma perspectiva inicial, com esta certeza da possibilidade do uso da teoria na genética, podemos prever a possibilidade da utilização da mesma na determinação de *árvores genealógicas de mutabilidade*, e com isso, sendo-nos capazes de mapear a totalidade de vírus e cruzar todas as possibilidades de mutações e *prever* qual variação viral pode ser agressiva a saúde humana, sendo-nos capazes de evitar pandemias. Dentro deste mesmo contexto, a aplicação da teoria na agricultura é vantajosa na estimação da mutabilidade de bactérias e assim na prevenção de suas possíveis mutações desvantajosas. A última e mais significativa aplicação da teoria seria a de sermos capazes de determinar fisicamente que a natureza possui um *padrão entrópico-energético* que vai além da matéria e energia bruta conhecida, e se expande para sistemas de informação (e isso não é trivial), também à própria mente, pela psicologia, ao mercado de ações no ramo da economia, na administração, na arte, etc.

