



Evento	Salão UFRGS 2015: SIC - XXVII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2015
Local	Porto Alegre - RS
Título	INFORMAÇÃO GENÉTICA BASEADA NA TEORIA DA INFORMAÇÃO
Autor	IGOR RUAN KOSARCZUK DA SILVA
Orientador	DAVID RENATO CARRETA DOMINGUEZ

INFORMAÇÃO GENÉTICA BASEADA NA TEORIA DA INFORMAÇÃO

Bolsista: Igor Ruan K. Silva

Orientador: Prof. David Dominguez

Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.

Sequências genéticas podem ser analisadas em um contexto informativo a partir da teoria da informação. Determinando densidades de bases nitrogenadas e até mesmo triplas de bases, conhecidas como códon, responsáveis pela informação referente à síntese proteica, pode-se obter a entropia de informação em cada caso que auxilia na obtenção de padrões na própria sequência genética e até mesmo entre duas sequências distintas.

Como base teórica, utilizamos a Teoria da Informação, desenvolvida por Claude E. Shannon do Bell Labs. A teoria expressa, em poucas palavras, que toda informação gerada por um transmissor e obtida pelo receptor possui um ruído, e que tal ruído pode ser minimizado de forma otimizada conhecendo-se a entropia de informação desse sistema. No caso de uma transmissão telefônica, a qual era o motivo inicial da pesquisa, quantas vezes era necessário multiplicar o sinal enviado para garantir que a mensagem entregue estivesse dentro de um limite de confiabilidade desejável? Na verdade, conhecendo a entropia da informação podem-se gerar métodos de entregar a mesma mensagem com o nível de confiança desejável sem a necessidade de multiplicar tanto o sinal. Com isto, pode-se aplicar o mesmo princípio da base de informação conhecida como DNA até a síntese completa de uma proteína. Entre o DNA e uma proteína existem muitos processos de transmissão e recepção de informação, em cada processo pode ocorrer um ruído, e é neste cenário em que a teoria pode ser aplicada.

Utilizamos a teoria sobre amostras de vírus do tipo Influenza A, determinando densidades de bases, códon e aminoácidos, objetivando a informação mútua contida em cada amostra e relacionando-as entre si, buscando obter o *grau de proximidade* entre estas. Com esses graus de proximidade, será mostrado que a teoria pode ser usada para classificar estirpes para um determinado grupo genético (vírus, bactérias, archeas, etc.).

O futuro da pesquisa se baseia em determinar rotas geográficas de mutabilidade de agentes biológicos, tais como vírus. Em outro contexto, estipular árvores genealógicas prováveis com base na menor variação entrópica e, juntamente com este, estipular as novas variações/mutações que sejam mais prováveis para os agentes biológicos em questão. Neste último caso, determinar as variações mais possíveis é motivador, pois a título de prevenção de novas pandemias (causadas por novas mutações de agentes biológicos) é muito viável.