



Evento	Salão UFRGS 2018: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
Ano	2018
Local	Campus do Vale - UFRGS
Título	Script para a otimização de análises comparativas utilizando o módulo Biopython
Autor	TOBIAS WEBER MARTINS
Orientador	FABIANA HORN

RESUMO

[máximo duas páginas]

TÍTULO DO PROJETO: Script para a otimização de análises comparativas utilizando o módulo Biopython

Aluno: Tobias Weber Martins

Orientadora: Fabiana Horn

RESUMO DAS ATIVIDADES

1. Introdução:

O progresso no campo da bioinformática vêm crescendo exponencialmente nos últimos anos; de descobertas inovadoras à invenção de novos métodos de análise, é notável perceber que um novo profissional está começando a surgir no nicho recém-criado pela genômica e por avanços computacionais.

Além do conhecimento essencial sobre a inerente relação entre todos os organismos vivos e os ambientes que os cercam, os biólogos que se aventuram por essas áreas precisam estar munidos com conhecimentos diversos de estatística, lógica, análise de dados e programação.

Como implicação direta desse aumento de complexidade, as quantidades de análises, pesquisas e ensaios necessários para se obter um argumento robusto também disparou, se tornando inevitável o dispêndio de tempo com trabalhos mecânicos e repetitivos tanto *in vivo* quanto *in silico*.

Foi visando acelerar alguns desses processos que a mescla entre biologia e programação provou sua eficácia, pois elevou a capacidade do pesquisador de realizar uma grande quantidade de tarefas *in silico* em um tempo substancialmente reduzido. Dentre as linguagens mais utilizadas entre os bioinformatas, o Python é uma escolha recorrente, seja por sua facilidade de aprendizado ou pela versatilidade de seus módulos e funções. Além disso, outra vantagem do Python é sua popularidade, por ser uma das linguagens mais utilizadas no mundo, e a participação ativa da comunidade na construção de novos módulos catalisa sua versatilidade de uso. Um desses módulos, o Biopython, além de trazer novos tipos de dados, também possibilita a utilização de uma variedade de ferramentas voltadas para genômica e biologia computacional em geral.

Valendo-se disso e objetivando a agilização no processamento e interpretação de dados obtidos de análises genômicas simples e cotidianas, construí um script em Python que, ao passar uma sequência como entrada, extrai os primeiros resultados do BLAST para essa sequência e faz um alinhamento global entre a sequência passada como parâmetro e cada um dos resultados obtidos separadamente utilizando o módulo *pairwise2*, também do Biopython. Como saída o script retorna três arquivos texto, um contendo os resultados do BLAST, outro com os alinhamentos globais e um terceiro, ainda em construção, que quantifica quantas sequências diferentes foram analisadas de acordo com o alinhamento obtido no *pairwise2*.

2. Atividades realizadas:

- Construção dos códigos responsáveis por pesquisar e registrar os resultados do BLAST
- Construção dos códigos responsáveis por alinhar e registrar os resultados do *pairwise2*
- Atualmente construindo os trechos responsáveis pela interpretação dos dados do alinhamento

1. Objetivos atingidos:

Com exceção do trecho do programa (atualmente em construção) que quantifica quantos polimorfismos diferentes foram encontrados das execuções do *pairwise2*, todo ele está funcional, sendo possível passar tanto uma sequência de nucleotídeos quanto de aminoácidos como parâmetro.

1. Resultados obtidos:

O programa, embora não finalizado, vem correspondendo às expectativas nos diversos testes feitos: ele acelerou o processo repetitivo de pesquisar uma sequência, salvar as de interesse e alinhá-las globalmente, procurando identificar polimorfismos. Apesar de necessária conexão com a internet para utilizar o BLAST e os bancos de dados da NCBI, o programa compensa esses processos de busca demorados com a utilização do módulo de alinhamento pertencente ao próprio Biopython.

2. Conclusão:

O programa apresentou uma performance satisfatória ante os casos testados, quando finalizado ele será postado no portal GitHub, que centraliza algoritmos de usuários, empresas e autônomos em um só lugar, sendo disponibilizado para todos baixarem e fazerem suas próprias otimizações.