



ANÁLISE DE CÓDONS, DOWNLOAD DE SEQUÊNCIAS FASTA E COMPARAÇÃO DE ARQUIVOS FASTA A PARTIR DO NCBI

Resumo expandido análise genética

Pedro Rezende¹, Rodrigo Curvello²

1 Estudante de Graduação em Ciência da Computação, IFCC—*Campus* Rio do Sul. E-mail: rezende.ifcc@gmail.com.

2—Orientador, Professor Rodrigo Curvello, IFCC—*Campus* Rio do Sul. E-mail: rodrigo.curvello@ifcc.edu.br

RESUMO

Este resumo expandido apresenta três ferramentas bioinformáticas desenvolvidas para análise de sequências de DNA: o CodonAnalyzer, para análise da frequência de códons; a classe FastaBaixar, para download de sequências FASTA a partir do NCBI; e a classe FastaComparar, para comparação de sequências FASTA. O CodonAnalyzer calcula a frequência de cada códon em uma sequência de DNA e gera um relatório detalhado. A classe FastaBaixar utiliza a API E-utilities do NCBI para buscar e recuperar sequências FASTA, verificando previamente se o arquivo já existe localmente. A classe FastaComparar realiza a comparação de duas sequências FASTA, identificando a similaridade e as regiões compatíveis entre elas. Essas ferramentas são fundamentais para estudos na área da biologia molecular e genética, permitindo a otimização da expressão gênica, análises filogenéticas e o desenvolvimento de novas biotecnologias.

Palavras-chave: Análise de códons; Sequências FASTA; Comparação de sequências; NCBI; Bioinformática.

INTRODUÇÃO

O DNA (Ácido Desoxirribonucleico) é a molécula responsável por armazenar e transmitir as informações genéticas em todos os seres vivos. Essa molécula é composta por uma fita dupla em forma de espiral, formada por nucleotídeos que se ligam de forma complementar, sendo as bases nitrogenadas adenina (A), citosina (C), guanina (G) e timina (T) os principais componentes.

A sequência de nucleotídeos no DNA carrega a informação necessária para a síntese de proteínas, processo conhecido como “código genético”. Nesse código, cada conjunto de três nucleotídeos, chamado de códon, corresponde a um aminoácido específico ou a um sinal de parada da tradução. O estudo do uso preferencial de códons, ou “codon bias”, é essencial para entender a expressão gênica e a evolução dos organismos.

Além disso, a comparação de sequências de DNA, geralmente no formato FASTA, é fundamental para a identificação de genes conservados, detecção de mutações e construção de árvores filogenéticas. O acesso a bancos de dados de sequências biológicas, como o NCBI (National Center for Biotechnology Information), é crucial para a realização dessas análises comparativas.

Nesse contexto, o presente trabalho propõe o desenvolvimento de três ferramentas bioinformáticas: o CodonAnalyzer, para a análise da frequência de códons em sequências de DNA; a classe FastaBaixar, responsável por realizar o download de sequências FASTA a partir do NCBI; e a classe FastaComparar, que realiza a comparação de duas sequências FASTA, identificando a similaridade e as regiões compatíveis entre elas. Essas ferramentas têm aplicações importantes em diversas áreas da biologia molecular e genética.

PROCEDIMENTOS METODOLÓGICOS

O desenvolvimento do CodonAnalyzer, da classe FastaBaixar e da classe FastaComparar seguiu os seguintes procedimentos metodológicos:

Revisão Bibliográfica: Foi realizada uma ampla revisão da literatura sobre o DNA, códons, o formato FASTA, a utilização do NCBI e a comparação de sequências, a fim de fundamentar teoricamente o desenvolvimento das ferramentas.

Desing patterns: Esse projeto foi utilizado o Facade para Controlar todas as classes a seguir.

Implementação do CodonAnalyzer: A classe CodonAnalyzer foi desenvolvida para analisar a frequência de códons em arquivos FASTA. Ela possui métodos para ler a sequência de DNA, calcular a frequência de cada códon e gerar um relatório com os resultados.

Implementação da Classe FastaBaixar: A classe FastaBaixar foi criada para realizar o download de sequências FASTA a partir do NCBI. Ela utiliza as E-utilities do NCBI para buscar e recuperar as sequências, verificando previamente se o arquivo já existe localmente.

Implementação da Classe FastaComparar: A classe FastaComparar foi desenvolvida para comparar duas sequências FASTA, identificando a similaridade entre elas e as regiões compatíveis.

Integração das Ferramentas: Foi desenvolvida uma classe GeneticaFacade que integra as funcionalidades do CodonAnalyzer, da FastaBaixar e da FastaComparar, oferecendo uma interface simplificada para a realização de análises completas.

RESULTADOS E DISCUSSÃO

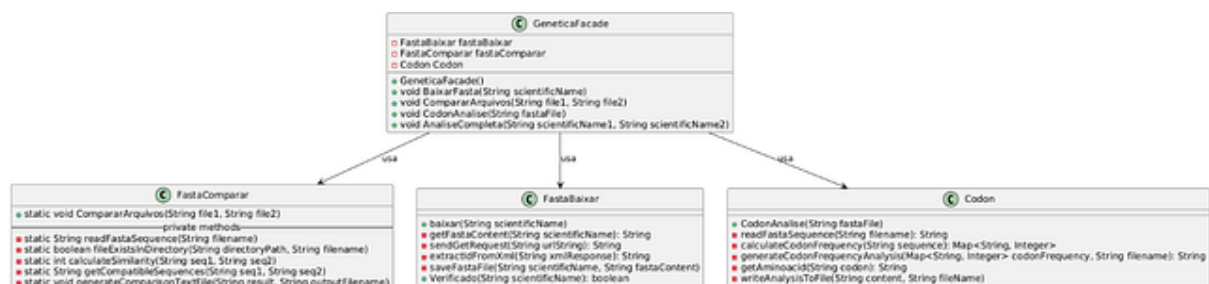
CodonAnalyzer: Análise de Frequência de Códons A classe CodonAnalyzer foi desenvolvida para analisar a frequência de códons em sequências de DNA contidas em arquivos FASTA. Ela calcula e gera relatórios detalhados sobre a frequência de cada códon, auxiliando na compreensão do uso preferencial de códons em diferentes organismos, o que é crucial para estudos de expressão gênica e evolução. O relatório permite identificar

padrões de seleção natural e preferências no processo de tradução em determinadas espécies.

FastaBaixar: Download de Sequências FASTA do NCBI A classe FastaBaixar automatiza o download de sequências FASTA diretamente do NCBI, utilizando as E-utilities para busca e recuperação de dados biológicos. Ela garante a obtenção eficiente de sequências genômicas, verificando a existência local do arquivo antes de baixar. Isso facilita a condução de estudos comparativos e a obtenção de dados genômicos, especialmente ao permitir downloads pelo nome científico da espécie.

FastaComparar: Comparação de Sequências FASTA A classe FastaComparar compara duas sequências FASTA e gera um relatório detalhado, destacando similaridades e divergências. Essa comparação é essencial para identificar genes conservados, detectar mutações e construir árvores filogenéticas. O relatório facilita a visualização de regiões importantes para estudos genéticos e evolutivos.

GeneticaFacade: Integração das Ferramentas de Análise Genômica A classe GeneticaFacade integra as funcionalidades das classes CodonAnalyzer, FastaBaixar, e FastaComparar, simplificando a realização de análises genômicas. Essa integração facilita análises complexas de maneira mais intuitiva e eficiente, aumentando a produtividade em estudos genômicos e filogenéticos.



Figura¹- Diagrama de classes

Fonte: Lucena Biologia Natural

		Segunda Base					
		U	C	A	G		
Primeira Base 5'	U	UUU } Fenil-alanina UUC } UUA } Leucina UUG }	UCU } UCC } Serina UCA } UCG }	UAU } Tirosina UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine UGC } UGA } Stop codon UGG } Tryptophan	Terceira Base 3'	U C A G
	C	CUU } Leucina CUC } CUA } CUG }	CCU } CCC } Prolina CCA } CCG }	CAU } Histidina CAC } CAA } Glutamina CAG }	CGU } CGC } Arginina CGA } CGG }		U C A G
	A	AUU } Isoleucina AUC } AUA } AUG } Metionina start codon	ACU } ACC } Treonina ACA } ACG }	AAU } Asparagina AAC } AAA } Lisina AAG }	AGU } Serina AGC } AGA } Arginina AGG }		U C A G
	G	GUU } Valina GUC } GUA } GUG }	GCU } GCC } Alanina GCA } GCG }	GAU } Ácido GAC } Aspártico GAA } Ácido GAG } Glutâmico	GGU } GGC } Glicina GGA } GGG }		U C A G

Figura²- Tabela codon

Discussão

A implementação das classes descritas resultou em um conjunto robusto de ferramentas para análise genômica. O CodonAnalyzer fornece insights valiosos sobre o uso de códons em sequências de DNA, o que pode influenciar diretamente estudos sobre expressão gênica e evolução. A capacidade de baixar sequências FASTA de forma automatizada através do FastaBaixar simplifica a obtenção de dados genômicos, enquanto o FastaComparar permite a comparação detalhada de sequências, essencial para a identificação de semelhanças e diferenças entre organismos.

A GeneticaFacade, ao integrar essas funcionalidades, mostrou-se uma solução eficiente para pesquisadores, facilitando a execução de análises completas de forma simplificada. A interface unificada proposta pela GeneticaFacade representa um avanço significativo, impactando positivamente pesquisas nas áreas de biologia molecular, genética e evolução. A abordagem modular e integrada do sistema atende às necessidades práticas da análise genômica, oferecendo uma plataforma flexível e poderosa para estudos biológicos avançados.

CONSIDERAÇÕES FINAIS

O presente trabalho desenvolveu três ferramentas bioinformáticas relevantes para a análise de sequências de DNA: o CodonAnalyzer, para a análise da frequência de códons; a classe FastaBaixar, para o download de sequências FASTA a partir do NCBI; e a classe FastaComparar, para a comparação de sequências FASTA.

O CodonAnalyzer permite identificar padrões de uso preferencial de códons em diferentes organismos, o que é essencial para a otimização da expressão gênica, estudos evolutivos e o desenvolvimento de novas biotecnologias. A classe FastaBaixar facilita o acesso a dados genômicos, possibilitando a realização de análises comparativas e filogenéticas. A classe

FastaComparar, por sua vez, identifica a similaridade entre sequências de DNA e as regiões compatíveis, o que é fundamental para a detecção de genes conservados e a construção de árvores filogenéticas.

A integração dessas ferramentas na classe GeneticaFacade simplifica o uso e a aplicação das funcionalidades desenvolvidas, tornando-as acessíveis e fácil usabilidade.

REFERÊNCIAS

BRASIL ESCOLA. *DNA: resumo, função, estrutura, composição, DNA x RNA*. Disponível em: <https://brasilecola.uol.com.br/biologia/codigo-genetico.htm>. Acesso em: 20 ago. 2024. NATIONAL HUMAN GENOME RESEARCH INSTITUTE. *Codon*. Disponível em: <https://www.genome.gov/genetics-glossary/Codon>. Acesso em: 20 ago. 2024. NATIONAL GEOGRAPHIC BRASIL. *O que é o DNA e do que ele é composto*. Disponível em: <https://www.nationalgeographicbrasil.com/ciencia/2024/04/o-que-e-o-dna-e-do-que-ele-e-composto>. Acesso em: 20 ago. 2024. KHAN ACADEMY. *Nucleic acids*. Disponível em: <https://pt.khanacademy.org/science/ap-biology/gene-expression-and-regulation/dna-and-rna-structure/a/nucleic-acids>. Acesso em: 20 ago. 2024. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. *Genetic Code*. Disponível em: <https://brasilecola.uol.com.br/biologia/codigo-genetico.htm>. Acesso em: 20 ago. 2024.

