



INSTITUTO FEDERAL
CATARINENSE

Instituto Federal Catarinense
Sistema Integrado de Gestão de Atividades Acadêmicas
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação

Emitido em 02/08/2019 às 14:04

Projeto de Pesquisa

Dados do Projeto Pesquisa	
Código:	PVD482-2019
Título do Projeto:	FERRAMENTA COMPUTACIONAL WEB APLICADA À ANÁLISE DE SEQUÊNCIAS GENÔMICAS
Tipo do Projeto:	EXTERNO (Projeto Novo)
Natureza do Projeto:	Projeto de Pesquisa
Tipo de Pesquisa:	Pesquisa Aplicada
Situação do Projeto:	AGUARDANDO VALIDAÇÃO
Unidade de Lotação do Coordenador:	RIO DO SUL - UU - COORD. GERAL DE ENSINO (11.01.05.12)
Unidade de Execução:	RIO DO SUL - COORD GERAL DE ENSINO (11.01.05.11)
Centro:	CAMPUS RIO DO SUL (11.01.05)
Palavra-Chave:	Bioinformática; Arroz; Linguagem Livre de Contexto
E-mail:	marcela.leite@ifc.edu.br
Período do Projeto:	01/08/2019 a 31/12/2019
Área de Conhecimento, Grupo e Linha de Pesquisa	
Área de Conhecimento:	Linguagem Formais e Autômatos
Grupo de Pesquisa:	.
Linha de Pesquisa:	Paradigmas de programação
Comitê de Ética	
Nº do Protocolo:	Não possui protocolo de pesquisa em Comitê de Ética.
Resumo	
<p>Atualmente, ferramentas de cunho tecnológico são essenciais nas mais diversas áreas do conhecimento. Dessa forma, quando aplicadas ao desenvolvimento de pesquisas, são utilizadas principalmente pelo fato de simplificarem processos, assim auxiliando profissionais na resolução de situações problemáticas. Este projeto tem por objetivo genérico, auxiliar profissionais da área de genética a traduzirem cadeias de ácidos nucleicos para aminoácidos, e destes últimos para proteínas. Além disso, a ferramenta proposta tem como objetivo permitir que pacotes oriundos de sequenciamentos genéticos possam ser unidos e consequentemente analisados, e também que cadeias de proteínas possam ser comparadas. Com isso, para testar a capacidade da ferramenta, o projeto será aplicado em um contexto onde é necessário comparar os genomas de dois indivíduos de uma variedade de arroz, com o detalhe de que um deles foi exposto a radiação e por consequência sofreu mutações. Assim sendo, a ferramenta agirá neste cenário como uma solução para a identificação de variações nas sequências, permitindo desse modo, a análise das relações dessas variações com seu impacto no desempenho da planta mediante a condições de clima extremo. Na primeira etapa do projeto, será desenvolvido o tradutor de sequências genômicas. Assim, será necessário aplicar conceitos da ciência da computação, por exemplo, gramáticas livres de contexto. Na segunda etapa de desenvolvimento, será necessário realizar um estudo sobre técnicas de alinhamento e algoritmos necessários para o processo de comparação de duas ou mais sequências genômicas. Devido ao suporte web que a ferramenta disponibilizará, durante o desenvolvimento do projeto serão utilizadas a linguagem de marcação HTML, a linguagem de estilo CSS, a linguagem de script PHP e também a linguagem Javascript. Além disso, para a definição da gramática será utilizado o software GALS. Portanto, acredita-se que no contexto onde será efetivamente testada, a ferramenta desenvolvida será fundamental para a descoberta do ponto da mutação exata a qual foi capaz de interferir no conjunto da expressão gênica. Ademais, em decorrência do desenvolvimento web, haverá maior disponibilidade aos usuários para a realização de consultas as funcionalidades do sistema.</p>	
Introdução/Justificativa	
(incluindo os benefícios esperados no processo ensino-aprendizagem e o retorno para os cursos e para os professores da IFC em geral)	
<p>O arroz é um dos alimentos mais importantes para a nutrição humana. É o segundo cereal mais cultivado no mundo, ocupando área aproximada de 158 milhões de hectares, além de ser consumido por dois terços da população mundial. O Brasil é o segundo maior produtor de arroz no mundo (USDA, 2010 e MAPA, 2017). A cadeia produtiva do arroz irrigado é uma das mais importantes para a agricultura catarinense, já que ocupa a quarta posição quanto ao valor bruto da produção, dentre os produtos de origem vegetal.</p> <p>As áreas catarinenses de cultivo de arroz irrigado estão distribuídas em quatro regiões distintas, são elas: Baixo/Médio Vale do Itajaí e Litoral Norte, Litoral Centro, Litoral Sul e Alto Vale do Itajaí. Além disso, no estado o sistema de cultivo do arroz é predominante de sementes pré-germinadas. Desta forma, quando se pensa em sustentabilidade dos sistemas produtivos, é de suma importância que as tecnologias desenvolvidas sejam cada vez mais específicas para cada realidade. Os fatores climáticos mais importantes para o cultivo do arroz irrigado são a temperatura, o fotoperíodo e a radiação solar. Estes elementos são fundamentais para o comportamento da fisiologia da cultura e, sua consequência produtividade. As plantas, por serem organismos sésseis, estão constantemente submetidas a condições adversas, como estresses bióticos e abióticos, que podem alterar aspectos morfológicos, fisiológicos, bioquímicos e/ou moleculares, limitando o crescimento, desenvolvimento e a produtividade (BUCHANAN, 2000; TRAN et al., 2009; TRAN et al., 2010).</p> <p>De maneira geral, as plantas têm mecanismo de resistência a estresses, que podem ser agrupadas em duas categorias gerais: mecanismos de evitação, no qual as plantas previnem a exposição ao estresse; e mecanismos de tolerância, no qual as plantas suportam o estresse. Além disso, quando as plantas estão submetidas a condições de estresses, elas adquirem mecanismos de resistência a determinados estresses, caracterizados como adaptação e aclimatação (BUCHANAN, 2000). Os estresses abióticos e bióticos podem reduzir, significativamente, rendimentos em lavouras (IPCC, 2007; EEA, 2011). Alguns estudos apontam os possíveis impactos das mudanças climáticas na produtividade do arroz (AGGARWAL & MALL, 2002; LI et al., 2015; MATTHEWS et al., 1995; MATTHEWS et al., 1997; TAO et al., 2008). Estes estudos apontam que a produção de arroz poderia ser comprometida diante dos eventos de mudanças climáticas globais previstas, como o aumento do calor excessivo, o que está relacionado ao aumento da temperatura e condições de déficit hídrico o que poderá causar uma esterilidade em grãos de arroz (IPCC, 2014).</p> <p>Devido à crescente demanda mundial por alimentos, observa-se a necessidade de aumentar o investimento em biotecnologia, sendo esta, uma alternativa para a geração de cultivares mais tolerantes aos estresses abióticos e resistentes aos estresses bióticos, que atendam os arranjos produtivos locais, além de alavancar a economia do Estado. Neste contexto, técnicas de engenharia genética, em especial a utilização de agentes mutagênicos químicos, pode ser uma alternativa para o desenvolvimento de sementes de arroz geneticamente modificadas e, que gerem plantas adaptadas às condições do Alto Vale do Itajaí, e que possa viabilizar o setor produtivo de arroz irrigado em todo o Estado de Santa Catarina.</p> <p>O projeto proposto está vinculado ao projeto de pesquisa intitulado Utilização de mutação induzida em sementes de arroz visando obtenção de plantas tolerantes aos estresses abióticos e bióticos, diante do cenário de mudanças climáticas globais aprovado no Edital 267-2017. Este projeto tem como objetivo principal acompanhar o desenvolvimento de plantas de arroz irrigado, submetidas à radiação gama, em condições climáticas do Alto Vale do Itajaí, e identificar plantas que sejam tolerantes aos estresses abióticos e bióticos, prevendo cenários de mudanças climáticas globais. Porém, após a identificação das plantas de arroz, será necessário descobrir a variabilidade entre as plantas, e qual região do DNA foi alterada com a mutação induzida. Sendo assim, profissionais treinados na área de Bioinformática serão fundamentais para a montagem dos contigs e para a identificação, através de técnicas computacionais, o ponto exato de mutação por meio de Bioinformática.</p>	

técnicas computacionais, o ponto exato de mutação por meio de Bioinformática.

A comparação de genomas de indivíduos possibilita a identificação e análise de variações nas sequências, como as mutações ou polimorfismos, assim como relações entre estas variações e suas consequências. A comparação é realizada através da técnica de bioinformática conhecida como Alinhamento (VERLI, 2014).

A tarefa de realizar o alinhamento de sequências biológicas é complexa e computacionalmente trata-se de um problema insolúvel, para o qual, são empregados algoritmos de heurísticas que permitem encontrar soluções próximas à ótima (VERLI, 2014).

O propósito deste trabalho será realizar um estudo sobre as técnicas e algoritmos empregados para realizar o alinhamento e análise das similaridades. Na primeira fase, será desenvolvido um interpretador capaz de reconhecer uma sequência biológica de DNA ou RNA e identificar os aminoácidos e proteínas gerados a partir da sequência dada. Tal ferramenta, possibilitará o entendimento dos conceitos biológicos envolvidos no projeto e técnicas de programação necessárias para atingir o objetivo deste. Numa segunda fase serão avaliados os algoritmos e ferramentas que podem ser integrados a ferramenta desenvolvida na primeira fase, para que seja possível analisar as variações ocorridas no DNA do arroz tratado neste projeto e permitir a análise das relações dessas variações com seu impacto no desempenho da planta mediante a condições de clima extremo.

A ferramenta desenvolvida na primeira fase será construída utilizando gramáticas livre de contexto para interpretação da sequência de DNA ou RNA, permitindo gerar através de sua árvore de derivação o conjunto de aminoácidos. Gramáticas livre de contexto são comumente aplicadas para análise sintática de textos e identificação de padrões, sendo usadas como uma abordagem para o processamento de sequências genômicas (PIO et al., 2006). O conjunto de aminoácidos gerado será a entrada para uma gramática que irá produzir as proteínas. Desta forma, a ferramenta desenvolvida funcionará como um tradutor da sequência de RNA para a cadeia de proteínas sintetizadas a partir deste código.

Objetivos

Desenvolver uma ferramenta web para alinhamento e análise de sequências de DNA/RNA.

Metodologia

Devido a Bioinformática ser uma área de pesquisa relativamente pouco difundida, inicialmente será realizado um estudo sobre aplicações, ferramentas e técnicas utilizadas nesta área. Um detalhamento inicial do projeto, técnicas utilizadas e resultados esperados será apresentado e discutido com um profissional envolvido com a área de Genética que elucidará os principais conceitos envolvidos e permitirá o início do projeto com o desenvolvimento do interpretador da sequência genômica.

Será desenvolvida uma ferramenta web que receberá como entrada um texto formado por uma sequência genômica e gerará como resultado um vetor de aminoácidos e proteínas sintetizadas. Para construção desta ferramenta serão utilizados conceitos da ciência da computação para desenvolvimento de ferramentas de processamento de texto, interpretação e compilação de programas de computador (HOPCROFT, ULLMAN, MOTWANI, 2002 e AHO et al, 2008). Essa abordagem consiste em construir uma ferramenta dividida nas seguintes etapas: análise léxica, na qual são identificados tokens no texto de entrada; análise sintática, para validação da sequência e criação da árvore de análise sintática; análise semântica, que permite identificar relações entre os termos encontrados no texto analisado e dar um significado dentro de um contexto para esses termos; geração de código alvo, neste caso, a geração das cadeias de aminoácidos e proteínas. Esta fase tem como entrada a estrutura gerada na fase de análise sintática e aprimorada na fase de análise semântica.

A construção de uma gramática para sequências de DNA ou RNA e Aminoácidos necessita de um estudo e entendimento do controle da expressão gênica. Esse estudo será realizado consultando livros e artigos científicos e definirá as bases para a construção da gramática e limitações que podem ser encontradas durante o desenvolvimento da ferramenta.

A sequência de RNA pode ser lida em ambas as direções (5' e 3' ou vice versa) e diferentes combinações podem ser utilizadas para formar um códon que irá produzir um aminoácido. Essa particularidade leva a necessidade de produzir seis versões de resultados como saída da ferramenta (chamada de fases de leitura), ou seja, considerando a leitura em ambos os sentidos e três possibilidades de intercalação dos nucleotídeos para geração dos aminoácidos.

A ferramenta apresentará como saída a sequência de aminoácidos identificados, considerando as seis versões e a sequência de proteínas traduzidas a partir dos aminoácidos.

A segunda etapa de desenvolvimento necessitará da realização de um estudo sobre técnicas de alinhamento e algoritmos necessários para o processo de comparação de duas ou mais sequências genômicas.

Por último, após o sequenciamento completo de plantas de arroz com características agronômicas relevantes (seguindo a metodologia de NGS - sequenciamento de nova-geração), a ferramenta desenvolvida será fundamental para a descoberta do ponto da mutação exata a qual foi capaz de interferir no conjunto da expressão gênica.

Devido ao suporte web que a ferramenta disponibilizará, durante o desenvolvimento do projeto serão utilizadas a linguagem de marcação HTML, a linguagem de folhas de estilo CSS, a linguagem de script PHP e também a linguagem Javascript. Além disso, para a definição da gramática será utilizado o software GALS.

Para que se possa registrar e manter um controle das mudanças feitas nos arquivos referentes ao projeto durante o decorrer do mesmo, será utilizada a plataforma Bitbucket para o controle de versões.

Referências

Aggarwal, P. K., & Mall, R. K. Climate change and rice yields in diverse agro-environments of India. II. Effect of uncertainties in scenarios and crop models on impact assessment. Climatic Change, 52, 331343. 2002.

Buchanan B.B., Gruissem W., Jones R.L. Biochemistry & Molecular Biology of Plants. 2000.

European Environmental Agency (EEA) Global and European temperature (CSI 012/CLIM 001), Assessment May 2011. Copenhagen (Disponível: <http://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature/global-and-european-temperature-assessment-4>: Acesso 10.02.2016). 2011.

IPCC Climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor MMB, Miller HL (eds) Working group 1 contribution to the fourth assessment report of the intergovernmental panel on climate change (IPCC). Chapters 3 (observations: surface and atmospheric climate change), 10 (global climate projections), 11 (regional climate projections). Cambridge University Press, Cambridge. 2007.

Li, T., Hasegawa, T., Yin, X. Y., Zhu, Y., Boote, K., Adam, M., Gaydon, D. (2015). Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biology, 21, 13281341. <https://doi.org/10.1111/gcb.2015.21.issue-3>.

MAPA. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Valor Bruto da Produção Agropecuária. Disponível em < <http://www.agricultura.gov.br/>>. Acesso em 26 de dezembro de 2017.

MAPA. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Projeções do Agronegócio Brasil 2014/15 a 2024/25. Projeções de Longo Prazo. AGE/MAPA. Brasília. 2015b, 133a.

Matthews, R. B., Kropff, M. J., & Bachelet, D. D. Modeling the impact of climate change on rice production in Asia. Wallingford, CT: CAB International. 1995.

Matthews, R. B., Kropff, M. J., Horie, T., & Bachelet, D. Simulating the impact of climate change on rice production in Asia and evaluating options for adaptation. Agricultural Systems, 54, 399425. 1997.

HOPCROFT, J. E.; ULLMAN, J. D.; MOTWANI, R. Introdução à Teoria de Autômatos, Linguagens e Computação. 2. ed. Rio de Janeiro, RJ, BR: Elsevier, 2002.

PIO, José L. de S.; OLIVEIRA, Camillo J. S.; ARAÚJO, Arnaldo de A.; OLIVEIRA, Newton J. de. Reconhecimento Sintático e Estrutural de Padrões. Atualizações em Informática. Rio de Janeiro, ed PUC-RIO, SBC, 2006.

Tao, F., Hayashi, Y., Zhang, Z., Sakamoto, T., & Yokozawa, M. Global warming, rice production, and water use in China: Developing a probabilistic assessment. Agricultural and Forest Meteorology, 148,94110. 2008.

VERLI, H. Bioinformática - da biologia à flexibilidade molecular. Sociedade Brasileira de Bioquímica e Biologia Molecular. Ed 1. São Paulo, 2014.

CATANI, A. et al. Ser Protagonista: biologia, 1º ano: ensino médio. 3. ed. São Paulo: Edições SM, 2016.

LOPES, S.; ROSSO, S. Bio: volume 2. 1. ed. São Paulo: Saraiva, 2010.

USDA 2010. World Agricultural Supply and Demand Estimates <http://www.usda.gov/oce/commodity/wasde/latest.pdf>. Acesso em 26 de dezembro de 2017.

Tran L S, Quach TN, Guttikonda SK, Aldrich DL, Kumar R, Neelakandan A, Valliyodan B, Nguyen H. Molecular characterization of stress-inducible GmNAC genes in soybean. Molecular Genetics Genomics. 281(6):647-64. 2009.

Financiamentos						
Entidade Financiadora		Natureza do Financiamento			Data Inicio	Data Fim
Sem Financiamento		Auxilio Financeiro			Membros do Projeto	
CPF	Nome	Categoria		CH Dedicada	Tipo de Participação	
067.381.266-97	GISELLE CAMARGO MENDES	DOCENTE		2	COORDENADOR(A) ADJUNTO(A)	
035.976.149-65	MARCELA LEITE	DOCENTE		2	COORDENADOR(A)	
2019						
Atividades		Ago	Set	Out	Nov	Dez
APROFUNDAR PESQUISA SOBRE PROCESSAMENTO DE GENOMAS E BIOINFORMÁTICA						
DESENVOLVER FERRAMENTA PARA PROCESSAMENTO DE SEQUÊNCIA GENOMA QUE IDENTIFIQUE PROTEÍNAS SINTETIZADAS POR UMA SEQUÊNCIA DADA DE AMINOÁCIDOS						
Histórico do Projeto						
Data	Situação			Usuário		
30/07/2019	CADASTRADO			MARCELA LEITE / marcela.leite		
30/07/2019	AGUARDANDO VALIDAÇÃO			MARCELA LEITE / marcela.leite		

Documento emitido por: MARCELA LEITE